

p-ISSN 1607-3274
e-ISSN 2313-688X



Радіоелектроніка Інформатика Управління

Radio Electronics Computer Science Control



2025/4



Міністерство освіти і науки України
Національний університет «Запорізька політехніка»

Радіоелектроніка, інформатика, управління

Науковий журнал

Виходить чотири рази на рік

№ 4(75) 2025

Заснований у 1998 році, видається з 1999 року.

Засновник і видавець – Національний університет «Запорізька політехніка».

ISSN 1607-3274 (друкований), ISSN 2313-688X (електронний).

Запоріжжя

НУ «Запорізька політехніка»

2025

Ministry of Education and Science of Ukraine
National University Zaporizhzhia Polytechnic

Radio Electronics, Computer Science, Control

The scientific journal

Published four times per year

№ 4(75) 2025

Founded in 1998, published since 1999.

Founder and publisher – National University Zaporizhzhia Polytechnic.

ISSN 1607-3274 (print), ISSN 2313-688X (on-line).

Zaporizhzhia

NU Zaporizhzhia Polytechnic

2025

Науковий журнал «Радіоелектроніка, інформатика, управління» (скорочена назва – РІУ) видається Національним університетом «Запорізька політехніка» (НУ «Запорізька політехніка») з 1999 р. періодичністю чотири номери на рік.

Ресстрація суб'єкта у сфері друкованих медіа: Рішення Національної ради України з питань телебачення і радіомовлення № 3040 від 07.11.2024 року. Ідентифікатор медіа: R30-05582.

ISSN 1607-3274 (друкований), ISSN 2313-688X (електронний).

Наказом Міністерства освіти і науки України № 409 від 17.03.2020 р. «Про затвердження рішень Атестаційної колегії Міністерства щодо діяльності спеціалізованих вчених рад від 06 березня 2020 року» **журнал включений до переліку наукових фахових видань України в категорії «А» (найвищий рівень)**, в яких можуть публікуватися результати дисертаційних робіт на здобуття наукових ступенів доктора наук і доктора філософії (кандидата наук).

Журнал включений до польського Переліку наукових журналів та рецензованих матеріалів міжнародних конференцій з присвоєною кількістю балів (додаток до оголошення Міністра науки та вищої освіти Республіки Польща від 31 липня 2019 р.: № 16981).

В журналі безкоштовно публікуються наукові статті англійською, російською та українською мовами.

Правила оформлення статей подано на сайті: <http://ric.zntu.edu.ua/information/authors>.

Журнал забезпечує **безкоштовний відкритий он-лайн доступ** до повнотекстових публікацій.

Журнал дозволяє авторам мати авторські права і зберігати права на видання без обмежень. Журнал дозволяє користувачам читати, завантажувати, копіювати, поширювати, друкувати, шукати або посилається на повні тексти своїх статей. Журнал дозволяє повторне використання його вмісту у відповідності Creative Commons ліцензії CC BY-SA..

Опублікованим статтям присвоюється унікальний ідентифікатор цифрового об'єкта DOI.

Журнал входить до наукометричної бази Web of Science.

Журнал реферується та індексується у провідних міжнародних та національних реферативних журналах і наукометричних базах даних, а також розміщується у цифрових архівах та бібліотеках з безкоштовним доступом у режимі on-line, повний перелік яких подано на сайті: <http://ric.zntu.edu.ua/about/editorialPolicies#custom-0>.

Тематика журналу: телекомунікації та радіоелектроніка, програми інженерія (включаючи теорію алгоритмів і програмування), комп'ютерні науки (математичне і комп'ютерне моделювання, оптимізація і дослідження операцій, управління в технічних системах, міжмашинна і людино-машинна взаємодія, штучний інтелект, включаючи системи, засновані на знаннях, і експертні системи, інтелектуальний аналіз даних, розпізнавання образів, штучні нейронні і нейро-нечіткі мережі, нечітку логіку, колективний інтелект і мультиагентні системи, гібридні системи), комп'ютерна інженерія (апаратне забезпечення обчислювальної техніки, комп'ютерні мережі), інформаційні системи та технології (структури та бази даних, системи, засновані на знаннях та експертні системи, обробка даних і сигналів).

Усі статті, пропонувані до публікації, одержують **об'єктивний розгляд**, що оцінюється за суттю без урахування раси, статі, віросповідання, етнічного походження, громадянства або політичної філософії автора(ів).

Усі статті проходять двоступінчасте закриті (анонімне для автора) **рецензування** штатними редакторами і незалежними рецензентами – провідними вченими за профілем журналу.

РЕДАКЦІЙНА КОЛЕГІЯ

Головний редактор – Субботін Сергій Олександрович – доктор технічних наук, професор, завідувач кафедри програмних засобів, Національний університет «Запорізька політехніка», Україна.

Заступник головного редактора – Піза Дмитро Макарович – доктор технічних наук, професор, директор інституту інформатики та радіоелектроніки, професор кафедри радіотехніки та телекомунікацій, Національний університет «Запорізька політехніка», Україна.

Члени редколегії:

Андрюлідакіс Іосіф – доктор філософії, голова департаменту телефонії Центру обслуговування мереж, Університет Яніни, Греція;

Бодяньський Євгеній Володимирович – доктор технічних наук, професор, професор кафедри штучного інтелекту, Харківський національний університет радіоелектроніки, Україна;

Веннекенс Юст – доктор філософії, доцент, доцент факультету інженерних технологій (кампус Де Наїр), Католицький університет Льовена, Бельгія;

Рекомендовано до видання Вченою радою НУ «Запорізька політехніка», протокол № 5 від 02.12.2025.

Журнал зверстаний редакційно-видавничим відділом НУ «Запорізька політехніка».

Веб-сайт журналу: <http://ric.zntu.edu.ua>.

Адреса редакції: Редакція журналу «РІУ», Національний університет «Запорізька політехніка», вул. Жуковського, 64, м. Запоріжжя, 69063, Україна.

Тел: (061) 769-82-96 – редакційно-видавничий відділ

E-mail: rvv@zntu.edu.ua

Вольф Карстен – доктор філософії, професор, професор кафедри технічної інформатики, Дортмундський університет прикладних наук та мистецтв, Німеччина;

Вуттке Ганс-Дітріх – доктор філософії, доцент, провідний науковий співробітник інституту технічної інформатики, Технічний університет Ільменау, Німеччина;

Горбань Олександр Миколайович – доктор фізико-математичних наук, професор, професор факультету математики, Університет Лестера, Велика Британія;

Городничий Дмитро Олегович – доктор філософії, кандидат технічних наук, доцент, провідний науковий співробітник Дирекції науки та інженерії, Канадська агенція прикордонної служби, Канада;

Дробахін Олег Олегович – доктор фізико-математичних наук, професор, перший проректор, Дніпровський національний університет імені Олеся Гончара, Україна;

Зайцева Олена Миколаївна – кандидат фізико-математичних наук, професор, професор кафедри інформатики, Жилінський університет в Жиліні, Словаччина;

Камеяма Мічитака – доктор наук, професор, професор факультету науки та інженерії, Університет Ішіномакі Сеншю, Японія;

Карташов Володимир Михайлович – доктор технічних наук, професор, завідувач кафедри медіаінженерії та інформаційних радіоелектронних систем, Харківський національний університет радіоелектроніки, Україна;

Леващенко Віталій Григорович – кандидат фізико-математичних наук, професор, завідувач кафедри інформатики, Жилінський університет в Жиліні, Словаччина;

Луенго Давид – доктор філософії, професор, завідувач кафедри теорії сигналів та комунікацій, Мадридський політехнічний університет, Іспанія;

Марковська-Качмар Урсула – доктор технічних наук, професор, професор кафедри обчислювального інтелекту, Вроцлавська політехніка, Польща;

Олійник Андрій Олександрович – доктор технічних наук, професор, професор кафедри програмних засобів, Національний університет «Запорізька політехніка», Україна;

Павліков Володимир Володимирович – доктор технічних наук, старший науковий співробітник, проректор з наукової роботи, Національний аерокосмічний університет ім. Н.Е. Жуковського «ХАІ», Україна;

Папшицький Марцін – доктор наук, професор, професор відділу інтелектуальних систем, Дослідний інститут систем Польської академії наук, м. Варшава, Польща;

Скруський Степан Юрійович – кандидат технічних наук, доцент, доцент кафедри комп'ютерних систем і мереж, Національний університет «Запорізька політехніка», Україна;

Табунчик Галина Володимирівна – кандидат технічних наук, професор, професор кафедри програмних засобів, Національний університет «Запорізька політехніка», Україна;

Тригано Томас – доктор філософії, старший викладач кафедри електричної та електронної інженерії, Інженерний коледж ім. С. Шамоу, м. Ашдод, Ізраїль;

Хенке Карстен – доктор технічних наук, професор, науковий співробітник факультету інформатики та автоматизації, Технічний університет Ільменау, Німеччина;

Шарпанських Олексій Альбертович – доктор філософії, доцент, доцент факультету аерокосмічної інженерії, Делфтський технічний університет, Нідерланди.

РЕДАКЦІЙНО-КОНСУЛЬТАТИВНА РАДА

Аррас Пітер – доктор філософії, доцент, доцент факультету інженерних технологій (кампус Де Наїр), Католицький університет Льовена, Бельгія;

Ліснянський Анатолій – кандидат фізико-математичних наук, головний науковий експерт, Ізраїльська електрична корпорація, Хайфа, Ізраїль;

Мадрицх Христіан – доктор філософії, професор факультету інженерії та інформаційних технологій, Університет прикладних наук Каринфії, Австрія;

Маркосян Мгер Вардкесович – доктор технічних наук, професор, директор Єреванського науково-дослідного інституту засобів зв'язу, професор кафедри телекомунікацій, Російсько-вірменський університет, м. Єреван, Вірменія;

Рубель Олег Володимирович – кандидат технічних наук, доцент факультету інженерії, Університет МакМастера, Гамільтон, Канада;

Тавхелідзе Автанділ – кандидат фізико-математичних наук, професор, професор школи бізнесу, технології та освіти, Державний університет ім. Ілії Чавчавадзе, Тбілісі, Грузія;

Урсутою Дору – доктор фізико-математичних наук, професор, професор кафедри електроніки та обчислювальної техніки, Трансильванський університет в Брашові, Румунія;

Шульц Пітер – доктор технічних наук, професор, професор факультету інженерії та комп'ютерних наук, Гамбургський університет прикладних наук (HAW Hamburg), Гамбург, Німеччина.

Факс: +38-061-764-46-62

© Національний університет «Запорізька політехніка, 2025

The scientific journal Radio Electronics, Computer Science, Control is published by the National University Zaporizhzhia Polytechnic NU Zaporizhzhia Polytechnic since 1999 with periodicity four numbers per year.

Registration of an entity in the field of print media: Decision of the National Council of Ukraine on Television and Radio Broadcasting No. 3040 of November 7, 2024. Media ID: R30-05582.

ISSN 1607-3274 (print), ISSN 2313-688X (on-line).

By the Order of the Ministry of Education and Science of Ukraine from 17.03.2020 № 409 "On approval of the decision of the Certifying Collegium of the Ministry on the activities of the specialized scientific councils dated 06 March 2020" **journal is included in the list of scientific specialized periodicals of Ukraine in category "A" (highest level)**, where the results of dissertations for Doctor of Science and Doctor of Philosophy may be published.

The journal is included to the Polish List of scientific journals and peer-reviewed materials from international conferences with assigned number of points (Annex to the announcement of the Minister of Science and Higher Education of Poland from July 31, 2019: Lp. 16981).

The journal publishes scientific articles in English, Russian, and Ukrainian free of charge.

The **article formatting rules** are presented on the site: <http://ric.zntu.edu.ua/information/authors>.

The journal provides policy of **on-line open (free of charge) access** for full-text publications. The journal allow the authors to hold the copyright without restrictions and to retain publishing rights without restrictions. The journal allow readers to read, download, copy, distribute, print, search, or link to the full texts of its articles. The journal allow reuse and remixing of its content, in accordance with Creative Commons license CC BY-SA.

Published articles have a unique digital object identifier (DOI).

The journal is included into Web of Science.

The journal is abstracted and indexed in leading international and national abstracting journals and scientometric databases, and also placed to the digital archives and libraries with a free on-line access, full list of which is presented at the site: <http://ric.zntu.edu.ua/about/editorialPolicies#custom-0>.

The journal scope: telecommunications and radio electronics, software engineering (including algorithm and programming theory), computer science (mathematical modeling and computer simulation, optimization and operations research, control in technical systems, machine-machine and man-machine interfacing, artificial intelligence, including data mining, pattern recognition, artificial neural and neuro-fuzzy networks, fuzzy logic, swarm intelligence and multiagent systems, hybrid systems), computer engineering (computer hardware, computer networks), information systems and technologies (data structures and bases, knowledge-based and expert systems, data and signal processing methods).

All articles proposed for publication receive an **objective review** that evaluates substantially without regard to race, sex, religion, ethnic origin, nationality, or political philosophy of the author(s).

All articles undergo a two-stage **blind peer review** by the editorial staff and independent reviewers – the leading scientists on the profile of the journal.

EDITORIAL BOARD

Editor-in-Chief – Sergey Subbotin – Dr. Sc., Professor, Head of Software Tools Department, National University Zaporizhzhia Polytechnic, Ukraine.

Deputy Editor-in-Chief – Dmytro Piza – Dr. Sc., Professor, Director of the Institute of Informatics and Radio Electronics, Professor of the Department of Radio Engineering and Telecommunications, National University Zaporizhzhia Polytechnic, Ukraine.

Members of the Editorial Board:

Iosif Androulidakis – PhD, Head of Telephony Department, Network Operation Center, University of Ioannina, Greece;

Evgeniy Bodyanskiy – Dr. Sc., Professor, Professor of the Department of Artificial Intelligence, Kharkiv National University of Radio Electronics, Ukraine;

Oleg Drobakhin – Dr. Sc., Professor, First Vice-Rector, Oles Honchar Dnipro National University, Ukraine;

Alexander Gorban – PhD, Professor, Professor of the Faculty of Mathematics, University of Leicester, United Kingdom;

Dmitry Gorodnichy – PhD, Associate Professor, Leading Research Fellow at the Directorate of Science and Engineering, Canada Border Services Agency, Ottawa, Canada;

Karsten Henke – Dr. Sc., Professor, Research Fellow, Faculty of Informatics and Automation, Technical University of Ilmenau, Germany;

Michitaka Kameyama – Dr. Sc., Professor, Professor of the Faculty of Science and Engineering, Ishinomaki Senshu University, Japan;

Volodymyr Kartashov – Dr. Sc., Professor, Head of the Department of Media Engineering and Information Radio Electronic Systems, Kharkiv National University of Radio Electronics, Ukraine;

Vitaly Levashenko – PhD, Professor, Head of Department of Informatics, University of Žilina, Slovakia;

David Luengo – PhD, Professor, Head of the Department of Signal Theory and Communication, Madrid Polytechnic University, Spain;

Ursula Markowska-Kaczmar – Dr. Sc., Professor, Professor of the Department of Computational Intelligence, Wrocław University of Technology, Poland;

Andrii Oliinych – Dr. Sc., Professor, Professor of the Department of Software Tools, National University Zaporizhzhia Polytechnic, Ukraine;

Marcin Paprzycki – Dr. Sc., Professor, Professor of the Department of Intelligent Systems, Systems Research Institute, Polish Academy of Sciences, Warsaw, Poland;

Volodymyr Pavlikov – Dr. Sc., Senior Researcher, Vice-Rector for Research, N. E. Zhukovsky National Aerospace University "KhAI", Ukraine;

Alexei Sharpanskykh – PhD, Associate Professor, Associate Professor of Aerospace Engineering Faculty, Delft University of Technology, Netherlands;

Stepan Skrupsky – PhD, Associate Professor, Associate Professor of the Department of Computer Systems and Networks, National University Zaporizhzhia Polytechnic, Ukraine;

Galyna Tabunshchik – PhD, Professor, Professor of the Department of Software Tools, National University Zaporizhzhia Polytechnic, Ukraine;

Thomas (Tom) Trigano – PhD, Senior Lecturer of the Department of Electrical and Electronic Engineering, Sami Shamoon College of Engineering, Ashdod, Israel;

Joost Vennekens – PhD, Associate Professor, Associate Professor, Faculty of Engineering (Campus de Nair), Katholieke Universiteit Leuven, Belgium;

Carsten Wolff – PhD, Professor, Professor of the Department of Technical Informatics, Dortmund University of Applied Sciences and Arts, Germany;

Heinz-Dietrich Wuttke – PhD, Associate Professor, Leading Researcher at the Institute of Technical Informatics, Technical University of Ilmenau, Germany;

Elena Zaitseva – PhD, Professor, Professor, Department of Informatics, University of Žilina, Slovakia.

EDITORIAL-ADVISORY COUNCIL

Peter Arras – PhD, Associate Professor, Associate Professor, Faculty of Engineering (Campus De Nair), Katholieke Universiteit Leuven, Belgium;

Anatoly Lisnianski – PhD, Chief Scientific Expert, Israel Electric Corporation Ltd., Haifa, Israel;

Christian Madritsch – PhD, Professor of the Faculty of Engineering and Information Technology, Carinthia University of Applied Sciences, Austria;

Mher Markosyan – Dr. Sc., Professor, Director of the Yerevan Research Institute of Communications, Professor of the Department of Telecommunications, Russian-Armenian University, Yerevan, Armenia;

Oleg Rubel – PhD, Associate Professor, Faculty of Engineering, McMaster University, Hamilton, Canada;

Peter Schulz – Dr. Sc., Professor, Professor, Faculty of Engineering and Computer Science, Hamburg University of Applied Sciences (HAW Hamburg), Hamburg, Germany;

Avtandil Tavkhelidze – PhD, Professor, Professor of the School of Business, Technology and Education, Ilia State University, Tbilisi, Georgia;

Doru Ursuțiu – Dr. Sc., Professor, Professor, Department of Electronics and Computer Engineering, University of Transylvania at Brasov, Romania.

Recommended for publication by the Academic Council of NU Zaporizhzhia Polytechnic, protocol № 5 dated 02.12.2025.

The journal is imposed by the editorial-publishing department of NU Zaporizhzhia Polytechnic.

The journal web-site is <http://ric.zntu.edu.ua>.

The address of the editorial office: Editorial office of the journal Radio Electronics, Computer Science, Control, National University Zaporizhzhia Polytechnic, Zhukovskiy street, 64, Zaporizhzhia, 69063, Ukraine.

Tel.: +38-061-769-82-96 – the editorial-publishing department.

E-mail: rvv@zntu.edu.ua

Fax: +38-061-764-46-62

© National University Zaporizhzhia Polytechnic, 2025

ЗМІСТ

НЕКРОЛОГ: ПІШОВ З ЖИТТЯ ДМИТРО ПІЗА.....	6
РАДІОЕЛЕКТРОНІКА ТА ТЕЛЕКОМУНІКАЦІЇ.....	7
<i>Жула С. С., Церне Е. О., Жула О. В.</i> ТЕОРЕТИЧНІ ОСНОВИ ОПТИМІЗАЦІЇ СТРУКТУРИ БАГАТОАНТЕННИХ РАДІОПЕЛЕНГАТОРІВ ДЛЯ ВИЗНАЧЕННЯ ПОЛОЖЕННЯ ДЖЕРЕЛ СТОХАСТИЧНИХ СИГНАЛІВ.....	7
<i>Vakaliuk T. A., Andreiev O. V., Nikitchuk T. M., Korenivska O. L., Nikitchuk S. M.</i> USE OF THE PARAMETERS OF THE LAW OF DISTRIBUTION OF THE MEASUREMENT ERRORS OF THE PULSE OXIMETER TO SELECT THE SENSOR SETTINGS.....	22
МАТЕМАТИЧНЕ ТА КОМП'ЮТЕРНЕ МОДЕЛЮВАННЯ.....	31
<i>Babakov R. M., Barkalov A. A., Titarenko L. A.</i> PSEUDO-RANDOM ENCODING OF STATES IN THE ALGORITHM FOR ALGEBRAIC SYNTHESIS OF A FINITE STATE MACHINE.....	31
<i>Berezsky O. M., Berezkyi M. O., Zarichnyi M. M.</i> A METHOD FOR DETERMINING THE FUZZY DISCRETE FRÉCHET DISTANCE.....	41
<i>Slavik O.</i> ABOUT SPECIAL CASES OF LAGRANGIAN INTERSTRIPATION OF APPROXIMATIONS OF FUNCTIONS OF TWO VARIABLES.....	50
<i>Turbal Y. V., Moroziuk A. Y.</i> OPTIMIZATION OF PERMANENT DECOMPOSITION PROCEDURES USING PARALLELIZATION ALGORITHM.....	59
НЕЙРОІНФОРМАТИКА ТА ІНТЕЛЕКТУАЛЬНІ СИСТЕМИ.....	65
<i>Bedratyuk L. P.</i> LOGIC-ONTOLOGICAL RECONSTRUCTION OF SCIENTIFIC DISCOURSE AND ITS IMPLEMENTATION IN AN AI-BASED REVIEWING SYSTEM.....	65
<i>Grygor O. O., Fedorov E. E., Leshchenko M. M., Rudakov K. S., Sakhno T. A.</i> METAHEURISTIC FRAMEWORKS FOR PARAMETER ESTIMATION IN APPROXIMATION MODELS.....	80
<i>Komleva N. O., Liubchenko V. V.</i> FORMALIZED METHODOLOGY FOR COMPATIBILITY AND ADAPTATION OF REQUIREMENTS IN INTELLIGENT DIAGNOSTIC SYSTEMS.....	92
<i>Kuz M. V., Lazarovych I. M., Kozlenko M. I., Pikuliak M. V., Kvasniuk A. D.</i> RESEARCH ON A HYBRID LSTM-CNN-ATTENTION MODEL FOR TEXT-BASED WEB CONTENT CLASSIFICATION.....	105
<i>Liashchynskyi P. B.</i> CLASSIFICATION OF HISTOLOGICAL IMAGES BASED ON CONVOLUTIONAL NEURAL NETWORKS.....	116
<i>Muhammad Junaid Iqbal, Mudasir Mahmood, Muhammad Farhan, Muhammad Sajjad Hussain</i> ALZHEIMER'S DISEASE PREDICTION BY USING DEEP STACKED ENSEMBLE MODEL ENHANCED WITH SQUEEZE-AND-EXCITATION ATTENTION MECHANISM.....	129
<i>Pukach A. I., Teslyuk V. M.</i> THE "PRISM AND RAYS METHOD" FOR RESEARCHING SUBJECTIVIZATION OF PERCEPTION OF HUMAN-MACHINE INTERACTION OBJECTS.....	139
<i>Tryhuba A. M., Ratushny R. T., Koval L. S., Ratushnyi A. R.</i> OPTIMIZED MODEL FOR PREDICTING THE AVAILABILITY OF OBJECTS BASED ON DEEP LEARNING AND GEOSPATIAL FEATURES.....	154
ПРОГРЕСИВНІ ІНФОРМАЦІЙНІ ТЕХНОЛОГІЇ.....	173
<i>Barkovska O. Y., Kovalenko A. A., Diachenko V. O., Bukharova L. D., Korobko V. Y.</i> PHOTOGRAMMETRIC MOTION CAPTURE SUBSYSTEM FOR CHANGE OF BODY POSITION ANALYSIS IN THE FRONTAL AND SAGITTAL PLANES.....	173
<i>Berdnyk M. G., Starodubskyi I. P.</i> USE OF GENETIC ALGORITHMS IN ADAPTIVE COMPILERS FOR CROSS-PLATFORM OPTIMIZATION.....	185
<i>Pastushenko M. S., Pastushenko O. M., Faizulaiev T. A.</i> ANALYSIS OF PROCEDURES FOR VOICE SIGNAL NORMALIZATION AND SEGMENTATION IN INFORMATION SYSTEMS.....	194
<i>Popov M., Zaitsev O., Stefantsev S.</i> AN ANALYTICAL APPROACH TO MULTI-CRITERIA CHOOSING TECHNOLOGICAL SCHEME FOR INFORMATION PROCESSING.....	202
<i>Suhoniak I., Marchuk G., Oleksiuk O.</i> MODIFY SHADERS AND RENDER TEXTURES ON CURVED SURFACES.....	209
<i>Хаханов В. І., Обрізан В. І., Шевченко О. Ю., Хаханова Г. В., Хаханов І. В., Демченко О. І., Воронов А. О.</i> МЕХАНІЗМИ СХЕМОТЕХНІЧНОГО МОДЕЛЮВАННЯ НА ОСНОВІ ВЕКТОРНОЇ ЛОГІКИ.....	219
УПРАВЛІННЯ У ТЕХНІЧНИХ СИСТЕМАХ.....	233
<i>Boyko N. I., Mokryk Y. L.</i> DEEP REINFORCEMENT LEARNING OPTIMIZATION METHODS FOR TRAFFIC LIGHTS AT SIGNALIZED INTERSECTIONS.....	233

CONTENTS

OBITUARY: DMYTRO PIZA PASSED AWAY.....	6
RADIO ELECTRONICS AND TELECOMMUNICATIONS.....	7
<i>Zhyla S. S., Tserne E. O., Zhyla O. V.</i>	
THEORETICAL FOUNDATIONS OF THE STRUCTURE OF MULTI-ANTENNA RADIO DIRECTION FINDERS OPTIMISATION FOR DETERMINING THE STOCHASTIC SIGNAL SOURCES POSITION.....	7
<i>Vakaliuk T. A., Andreiev O. V., Nikitchuk T. M., Korenivska O. L., Nikitchuk S. M.</i>	
USE OF THE PARAMETERS OF THE LAW OF DISTRIBUTION OF THE MEASUREMENT ERRORS OF THE PULSE OXIMETER TO SELECT THE SENSOR SETTINGS.....	22
MATHEMATICAL AND COMPUTER MODELING.....	31
<i>Babakov R. M., Barkalov A. A., Titarenko L. A.</i>	
PSEUDO-RANDOM ENCODING OF STATES IN THE ALGORITHM FOR ALGEBRAIC SYNTHESIS OF A FINITE STATE MACHINE.....	31
<i>Berezsky O. M., Berezkyi M. O., Zarichnyi M. M.</i>	
A METHOD FOR DETERMINING THE FUZZY DISCRETE FRÉCHET DISTANCE.....	41
<i>Slavik O.</i>	
ABOUT SPECIAL CASES OF LAGRANGIAN INTERSTRIPATION OF APPROXIMATIONS OF FUNCTIONS OF TWO VARIABLES.....	50
<i>Turbal Y. V., Moroziuk A. Y.</i>	
OPTIMIZATION OF PERMANENT DECOMPOSITION PROCEDURES USING PARALLELIZATION ALGORITHM..	59
NEUROINFORMATICS AND INTELLIGENT SYSTEMS.....	65
<i>Bedratyuk L. P.</i>	
LOGIC-ONTOLOGICAL RECONSTRUCTION OF SCIENTIFIC DISCOURSE AND ITS IMPLEMENTATION IN AN AI-BASED REVIEWING SYSTEM.....	65
<i>Grygor O. O., Fedorov E. E., Leshchenko M. M., Rudakov K. S., Sakhno T. A.</i>	
METAHEURISTIC FRAMEWORKS FOR PARAMETER ESTIMATION IN APPROXIMATION MODELS.....	80
<i>Komleva N. O., Liubchenko V. V.</i>	
FORMALIZED METHODOLOGY FOR COMPATIBILITY AND ADAPTATION OF REQUIREMENTS IN INTELLIGENT DIAGNOSTIC SYSTEMS.....	92
<i>Kuz M. V., Lazarovych I. M., Kozlenko M. I., Pikuliak M. V., Kvasniuk A. D.</i>	
RESEARCH ON A HYBRID LSTM-CNN-ATTENTION MODEL FOR TEXT-BASED WEB CONTENT CLASSIFICATION.....	105
<i>Liashchynskyi P. B.</i>	
CLASSIFICATION OF HISTOLOGICAL IMAGES BASED ON CONVOLUTIONAL NEURAL NETWORKS.....	116
<i>Muhammad Junaid Iqbal, Mudasir Mahmood, Muhammad Farhan, Muhammad Sajjad Hussain</i>	
ALZHEIMER'S DISEASE PREDICTION BY USING DEEP STACKED ENSEMBLE MODEL ENHANCED WITH SQUEEZE-AND-EXCITATION ATTENTION MECHANISM.....	129
<i>Pukach A. I., Teslyuk V. M.</i>	
THE "PRISM AND RAYS METHOD" FOR RESEARCHING SUBJECTIVIZATION OF PERCEPTION OF HUMAN-MACHINE INTERACTION OBJECTS.....	139
<i>Tryhuba A. M., Ratushny R. T., Koval L. S., Ratushnyi A. R.</i>	
OPTIMIZED MODEL FOR PREDICTING THE AVAILABILITY OF OBJECTS BASED ON DEEP LEARNING AND GEOSPATIAL FEATURES.....	154
PROGRESSIVE INFORMATION TECHNOLOGIES.....	173
<i>Barkovska O. Y., Kovalenko A. A., Diachenko V. O., Bukharova L. D., Korobko V. Y.</i>	
PHOTOGRAMMETRIC MOTION CAPTURE SUBSYSTEM FOR CHANGE OF BODY POSITION ANALYSIS IN THE FRONTAL AND SAGITTAL PLANES.....	173
<i>Berdnyk M. G., Starodubskyi I. P.</i>	
USE OF GENETIC ALGORITHMS IN ADAPTIVE COMPILERS FOR CROSS-PLATFORM OPTIMIZATION.....	185
<i>Pastushenko M. S., Pastushenko O. M., Faizulaiev T. A.</i>	
ANALYSIS OF PROCEDURES FOR VOICE SIGNAL NORMALIZATION AND SEGMENTATION IN INFORMATION SYSTEMS.....	194
<i>Popov M., Zaitsev O., Stefantsev S.</i>	
AN ANALYTICAL APPROACH TO MULTI-CRITERIA CHOOSING TECHNOLOGICAL SCHEME FOR INFORMATION PROCESSING.....	202
<i>Suhoniak I., Marchuk G., Oleksiuk O.</i>	
MODIFY SHADERS AND RENDER TEXTURES ON CURVED SURFACES.....	209
<i>Hahanov V. I., Obrizan V. I., Rakhlis D. Y., Khakhanova H. V., Hahanov I. V., Demchenko O. I., Voronov A. O.</i>	
MECHANISMS OF SCHEMATIC MODELING BASED ON VECTOR LOGIC.....	219
CONTROL IN TECHNICAL SYSTEMS.....	233
<i>Boyko N. I., Mokryk Y. L.</i>	
DEEP REINFORCEMENT LEARNING OPTIMIZATION METHODS FOR TRAFFIC LIGHTS AT SIGNALIZED INTERSECTIONS.....	233



З сумом повідомляємо, що на 80-му році пішов з життя Піза Дмитро Макарович (20 жовтня 1946, с. Будки Почаївського, нині Кременецького р-ну Тернопільської обл., Україна – 24 жовтня 2025 року, м. Запоріжжя, Україна), заступник головного редактора журналу «Радіоелектроніка, інформатика, управління», доктор технічних наук, професор, академік Української академії економічної кібернетики, член Української федерації інформатики.

По закінченню Запорізького машинобудівного інституту та отримання диплому за спеціальністю «Радіотехніка» у 1969 році працював на посадах інженера, провідного інженера, начальника сектору в Конструкторському бюро Науково-виробничого комплексу «Іскра». З 1985 року працював на посаді доцента кафедри «Радіотехніка» Запорізького машинобудівного інституту (нині – Національний університет «Запорізька політехніка»), у 1998–1999 – декан радіоприладобудівного факультету, з 1999 р. – проректор із питань перспектив розвитку університету, з 2001 р. за сумісництвом – директор інституту інформатики та радіоелектроніки, у 2021 р. перейшов на посаду професора кафедри радіотехніки і телекомунікацій.

Д. М. Піза як учений у галузі радіотехніки зробив вагомий особистий внесок у підвищення завадозахисності вітчизняних радіолокаційних станцій, на які є попит на закордонних ринках озброєнь. Був фундатором та головним редактором наукових журналів «Електротехніка і електроенергетика», «Радіоелектроніка, інформатика, управління». У 1998 року заснував наукову школу «Радіотехніка та інформаційні технології» та був її науковим керівником. Працював у складі науково-технічної ради Міністерства оборони України. У 1998 р. в Науковому центрі досліджень Генерального штабу Збройних сил України захистив докторську дисертацію за спеціальністю «АСУ та прогресивні інформаційні технології», отримав вчене звання професора. Входив до складу спеціалізованої вченої ради із захисту докторських дисертацій при Харківському національному університеті радіоелектроніки. Проводив експертизу дисертацій у Вищій атестаційній комісії України. Брав участь у численних науково-дослідних проєктах. Під керівництвом Д. М. Пізи захищено 10 кандидатських і 3 докторські дисертації. Він є автором понад 230 наукових праць, із них 5 монографій, понад 50 патентів, а також 4 навчальних посібників.

Відзначений медаллю «За трудову доблесть» та Почесною грамотою Кабінету Міністрів України, орденом «За заслуги перед Запорізьким краєм III ст.», відзнаками МОНУ «За наукові досягнення» та «Відмінник освіти України».

We are sad to inform that Dmytro Piza (October 20, 1946, Budky village, Pochaivskyi district, now Kremenets district, Ternopil region, Ukraine – October 24, 2025, Zaporizhzhia, Ukraine) has passed away at the age of 80. He was deputy editor-in-chief of the journal “Radio Electronics, Computer Science, Control”, Doctor of Technical Sciences, Professor, Academician of the Ukrainian Academy of Economic Cybernetics, member of the Ukrainian Federation of Informatics.

After graduating from the Zaporizhzhia Machine-Building Institute and receiving a diploma in “Radio Engineering” in 1969, he worked as an engineer, a leading engineer, and a head of the sector in the Design Bureau of the Iskra Scientific and Production Complex. Since 1985, he worked as an associate professor at the Department of Radio Engineering of the Zaporizhzhia Machine-Building Institute (now is National University “Zaporizhzhia Polytechnic”), in 1998–1999 he was the Dean of the Faculty of Radio Instrumentation, since 1999 the Vice-Rector for the Development Prospects of the University, since 2001 the Director of the Institute of Informatics and Radio Electronics, since 2021 the Professor of the Department of Radio Engineering and Telecommunications.

Dmytro Piza, as a scientist in the field of radioengineering, made a significant personal contribution to increasing the noise immunity of domestic radar stations, which are in demand on foreign arms markets. He was the founder and editor-in-chief of the scientific journals “Electrical Engineering and Power Engineering”, “Radio Electronics, Computer Science, Control”. In 1998, he founded the scientific school “Radio Engineering and Information Technologies” and served as its scientific supervisor. He worked as a member of the scientific and technical council of the Ministry of Defense of Ukraine. In 1998, at the Scientific Research Center of the General Staff of the Armed Forces of Ukraine, he defended his doctoral dissertation in the specialty “Automatic Control Systems and Advanced Information Technologies”, and then received the academic title of professor. He was a member of the specialized academic council for the defense of doctoral dissertations at the Kharkiv National University of Radio Electronics. He conducted the examination of dissertations at the Higher Attestation Commission of Ukraine. He participated in numerous scientific and research projects. Under the leadership of Dmytro Piza, 10 PhD and 3 doctoral dissertations were defended. He is the author of over 230 scientific papers, including 5 monographs, over 50 patents, and 4 textbooks.

Awarded the medal “For Labor Valor” and the Certificate of Honor of the Cabinet of Ministers of Ukraine, the Order “For Merit to the Zaporizhzhia Region, III degree”, the awards of the Ministry of Education and Science of Ukraine “For Scientific Achievements” and “Excellent Educationist of Ukraine”.

РАДИОЕЛЕКТРОНІКА ТА ТЕЛЕКОМУНІКАЦІЇ

RADIO ELECTRONICS AND TELECOMMUNICATIONS

УДК 621.396

ТЕОРЕТИЧНІ ОСНОВИ ОПТИМІЗАЦІЇ СТРУКТУРИ БАГАТОАНТЕННИХ РАДІОПЕЛЕНГАТОРІВ ДЛЯ ВИЗНАЧЕННЯ ПОЛОЖЕННЯ ДЖЕРЕЛ СТОХАСТИЧНИХ СИГНАЛІВ

Жила С. С. – д-р техн. наук, доцент, завідувач кафедри аерокосмічних радіоелектронних систем, Національний аерокосмічний університет «Харківський авіаційний інститут», Харків, Україна.

Церне Е. О. – д-р філософії, молодший науковий співробітник кафедри аерокосмічних радіоелектронних систем, Національний аерокосмічний університет «Харківський авіаційний інститут», Харків, Україна.

Жила О. В. – канд. фіз.-мат. наук, доцент кафедри вищої математики, Харківський національний університет радіоелектроніки, Харків, Україна.

АНОТАЦІЯ

Актуальність. Актуальність теми полягає у необхідності вдосконалення радіопеленгаторів для підвищення точності, стійкості до перешкод та адаптації до змінних умов експлуатації. Сучасні наукові досягнення вимагають розробки методів статистичного синтезу та аналізу обробки стохастичних сигналів у багатоантенних системах, що дозволить врахувати невідомість реальних умов. Важливим є розширення можливостей таких систем для застосування в радіолокації, радіонавігації, зв'язку та інших галузях. Це сприятиме створенню нових підходів для пеленгування джерел невідомих сигналів у складних сценаріях експлуатації.

Мета. Дослідження спрямоване на підвищенні точності вимірювань кутового положення джерел радіовипромінювання стохастичних сигналів.

Метод. Ґрунтується на статистичній теорії оптимізації радіотехнічних систем дистанційного зондування та радіолокації. Для джерел стохастичних сигналів побудовано моделі сигналів і шумів, сформульовано функціонал правдоподібності у спектральній області, враховуючи структуру обернених кореляційних матриць. Для визначення граничних похибок оцінювання кутового положення джерела радіовипромінювання використано нерівність Крамера-Рао.

Результати. Вперше теоретично обґрунтовано підхід до статистичної оптимізації структури багатоантенних радіосистем пеленгації джерел стохастичного випромінювання, що дозволяє врахувати просторову орієнтацію, геометрію антенної решітки та діаграми спрямованості. Побудовано оптимальний метод обробки рівнянь спостереження для оцінювання кутового положення джерел стохастичних сигналів. Запропоновано узагальнену структуру одноантенного пеленгатора, що містить узгоджений фільтр, декорелюючий фільтр та цифровий обчислювач. Доведено, що використання декореляційної обробки дозволяє підвищити точність оцінювання за рахунок збільшення кількості незалежних відліків сигналу. Одержано аналітичні вирази для оцінювання та граничних похибок, які враховують ширину спектру та параметри діаграми спрямованості.

Висновки. У роботі представлено новітні теоретичні основи синтезу радіопеленгаторів довільної конфігурації, які враховують різноманітність форм діаграм спрямованості, просторове розташування та орієнтацію пеленгаторів. Розроблені моделі сигналів і шумів із застосуванням критерію максимуму функціоналу правдоподібності вперше дозволяють вирішувати оптимізаційні задачі синтезу з урахуванням фізичного змісту кореляційних матриць. Отримані результати підтверджено вирішенням задачі вимірювання кутового положення джерела випромінювання, що доводить ефективність запропонованих підходів.

КЛЮЧОВІ СЛОВА: багатоантенні пеленгатори, статистична оптимізація, оптимальне оброблення сигналів, імітаційне моделювання.

НОМЕНКЛАТУРА

$s_k(t)$ – випромінений k -м джерелом сигнал,
 $k = \overline{1, K}$;
 $s_{r,i}(t, \vec{r}', \theta_{s,k})$ – прийняте i -ю антеною просторово-часове поле;
 t – час;
 $\vec{r}'$ – просторові координати поверхні антени;

$\theta_{s,k}$ – кути візування кожного з k -го джерела,
 $i = \overline{1, N}$;
 $\dot{G}_i(\theta - \theta_{0i})$ – діаграма спрямованості i -ї антени, що орієнтована своїм максимумом у напрямку θ_{0i} ;
 $\dot{S}(2\pi f)$ – комплексна стохастична спектральна щільність амплітуди випромінюваного сигналу;

$\dot{P}(2\pi f)$ – невідповідна комплексна частотна характеристика передавача стохастичного сигналу;

$\langle \cdot \rangle$ – символ статистичного усереднення;

K_{0i} – коефіцієнт підсилення i -го приймального каналу;

$\delta(\theta - \theta_s)$ – дельта-функція, що визначає просторове положення точкового джерела радіовипромінювання у напрямку θ_s ;

$\psi_i(f, \theta_s)$ – фазовий зсув прийнятих сигналів в кожній з антен відносно певного фазового центру антенної решітки;

$\Delta r_i(\theta_s)$ – відстань, що проходять електромагнітні хвилі від фазового центру антени до кожного її елемента;

$n_i(t)$ – внутрішні шуми в приймальних трактах;

$\underline{W}(t_1, t_2, \theta_s)$ – обернена матриця обернених кореляційних функцій;

$\lambda, \hat{\lambda}$ – параметри, що підлягають оцінюванню;

$u_i(t)$ – випадкові коливання;

$k(\lambda)$ – нормуючий коефіцієнт;

$\underline{R}_u(t_1 - t_2, \theta_s)$ – матриця кореляційних функцій;

$\delta(t_1 - t_2)$ – дельта-функція;

$\underline{I}$ – одинична матриця;

j – уявна одиниця;

c – швидкість розповсюдження електромагнітних хвиль;

T – час спостереження;

$F\{\cdot\}$ – оператор перетворення Фур'є;

$F^{-1}\{\cdot\}$ – оператор оберненого перетворення Фур'є;

$(\cdot)^+$ – знак ермітового спряження;

$\lambda_{\text{іст}}$ – істинне значення параметру, що підлягає оцінюванню;

$\sigma_{\hat{\lambda}}^2$ – гранична дисперсія оцінки $\hat{\lambda}$;

q – змінна, яка описує співвідношення сигнал-завада в прийнятих коливаннях;

ΔF – результуюча ширина спектру прийнятих коливань після узгодженої фільтрації і декореляції у додатній та від'ємній області частот;

$\hat{(\cdot)}$ – знак оцінки величини;

$\hat{P}_{s,T}$ – оцінка потужності корисного сигналу на інтервалі $(0, T)$.

ВСТУП

Актуальність. Радіопеленгатори і радіопеленгаційні комплекси знайшли широкого застосування в радіолокації [1–3], радіометрії [4, 5], радіонавігації [6–8], системах зв'язку [9] та радіокерування [10]. Осно-

вна задача окремого радіопеленгатора полягає в вимірюванні кутового положення джерела радіовипромінювання, а угруповання пеленгаторів – вимірювання просторового положення такого джерела або групи джерел. Існуючі теоретичні і практичні результати проєктування системи визначення кутового положення [11, 12] вагомі та створюють надійну основу для подальшого вдосконалення структури радіопеленгаторів, що передбачає покращення їхньої точності, збільшення робочого діапазону, підвищення стійкості до перешкод і забезпечення більшої гнучкості в адаптації до змінних умов експлуатації.

Варто відзначити, що в роботах [15–17] авторами вже досліджено особливості оптимальної обробки функціонально-детермінованих сигналів у багатоантенних радіопеленгаторах. Однак припущення про повну відомість моделі сигналу обмежує реальні сфери застосування отриманих результатів, адже в практичних умовах часто постає завдання пеленгування джерел стохастичних, апріорно невідомих сигналів, для яких оптимальний підхід до обробки та визначення параметрів прийнятих сигналів потребує додаткових досліджень та розробок.

У зв'язку з зазначеним стає актуальним статистичний синтез і аналіз методів обробки стохастичних сигналів у багатоантенних радіопеленгаторах, що дозволить розробити методологічні основи для адаптації систем пеленгування до реальних умов невизначеності, підвищити точність і надійність визначення параметрів сигналів, а також розширити можливості застосування таких систем у широкому діапазоні сучасних та перспективних сценаріїв експлуатації.

Об'єктом дослідження є процес статистичної оптимізації методів обробки стохастичних сигналів в багатоантенних радіопеленгаторах.

Предмет дослідження – основи теорії статистичного синтезу та аналізу методів обробки стохастичних сигналів в багатоантенних радіопеленгаторах.

Мета дослідження полягає в підвищенні точності вимірювань кутового положення джерел радіовипромінювання стохастичних сигналів.

1 ПОСТАНОВКА ЗАДАЧІ

Для розвитку теоретичних основ оптимізації структури багатоантенних радіопеленгаторів для випадку визначення положення джерел стохастичних сигналів опишемо вхідні і вихідні параметри.

Дано: у якості вихідних параметрів виступає просторове положення двох антен з координатами $(0, 0)$ і $(10, 0)$, що приймають електромагнітні випромінювання з частотною характеристикою

$$\dot{P}(2\pi f) = \begin{cases} 1, & f \in (2, 3 \text{ ГГц}, 2, 5 \text{ ГГц}); \\ 0, & f \notin (2, 3 \text{ ГГц}, 2, 5 \text{ ГГц}), \end{cases}$$

математичне сподівання прийнятих сигналів дорівнює нулю, амплітуда сигналів дорівнює одиниці, діаграма спрямованості кожної антени описується функцією

$$G_{1(2)}(\theta - \theta_{01(2)}) = \text{sinc}(\pi\theta D),$$

де розмір антени $D = 0,1 \text{ м}$.

Знайти: напрямок θ_s на точкове джерело радіовипромінювання.

2 ОГЛЯД ЛІТЕРАТУРИ

Питанням статистичного синтезу і аналізу методів обробки сигналів в радіопеленгаторах присвячено не багато робіт за останній час. Основні результати за цим напрямком отримані при вирішенні задач виявлення сигналу [18–20] при скануванні обраної ділянки простору діаграмою спрямованості. Представлені підходи розвивають статистичну теорію оптимізації радіопеленгаторів, дозволяють побудувати оптимальні приймачі, проте дозволяють лише вирішити питання чи існує сигнал на вході в приймач чи ні. Результати розробок за даним підходом не дозволяють оцінити точність пеленгації за кутовими координатами і її залежність від параметрів приймача. Також не можливо визначити діапазон кутів однозначних вимірювань, адже вирішується найпростіша задача перевищення порогу.

Більш ефективним підходом до побудови оптимальних структур радіопеленгаторів є статистичний синтез методів обробки сигналів з відомою формою та параметрами в присутності гаусівських шумів [21–24]. Результати отримані за таким підходом дозволяють отримати структуру пеленгатора та оцінити його потенційну точність роботи. В той самий такі системи не можуть виконувати основні свої функції, коли форма сигналу невідома або є шумовою. Зацікавленість в таких системах постійно зростає, адже спектр частот заповнюється новими джерелами радіовипромінювання і загальна кількість джерел інтерференції збільшується. Деякі автори намагаються технічно вирішити проблему пеленгації джерел шумових сигналів [25–28], але такі результати плідні лише для часткових випадків.

Значний вклад у розвиток та застосування методів пеленгації джерел випромінювання функціонально детермінованих сигналів і шумів постановників активних завад зробили автори робіт [29–31]. Наукова новизна отриманих результатів полягає в застосуванні вже відомих методів спектрального аналізу прийнятих сигналів антенними решітками і комбінації цих методів, а також статистичній обробці результатів вимірювань груп «надрозділяючих» методів. Даний підхід ефективний, проте не розкриває методів статистичного синтезу і аналізу наскрізної обробки сигналів, починаючи з реєстрації поля кожною антеною і закінчуючи потенційними дисперсіями оцінювання кутового положення джерела шумових сигналів. Основний вклад авторів зроблено в аналізі кореляційних © Жила С. С., Церне Е. О., Жила О. В., 2025
DOI 10.15588/1607-3274-2025-4-1

матриць рівнянь спостереження на виході фазованих антенних решіток та залежності якості оцінювання від точності пошуку обернених матриць. Менше уваги автори приділяють питанням розвитку математичного апарату синтезу нових методів обробки сигналів при багатоканальному прийомі.

Найбільш близьким по методології статистичного синтезу і аналізу методів обробки стохастичних сигналів в багатоантенних радіопеленгаторах є підхід, що представлено в роботі [32]. Автори пропонують виконувати статистичну оптимізацію методів в рамках критерію максимуму функції правдоподібності і записують функціонал правдоподібності в спектральній області для сигналів з стохастичним спектром. Тобто форма сигналу не конкретизується, а визначається лише діапазон частот та спектральна щільність потужності шумового сигналу. Даний підхід є плідним та дозволяє синтезувати найкращі методи та структури вимірювачів при параметричній невизначеності процесів, що підлягають обробці. Отримані результати та методологія може бути покращена за рахунок врахування кореляційних властивостей сигналів і шумів, визначаючи функцію правдоподібності з урахуванням прямої та оберненої кореляційної матриці процесів, що спостерігаються у приймачах.

3 МАТЕРІАЛИ ТА МЕТОДИ

Розглянемо основні складові теорії статистичного синтезу та аналізу методів обробки стохастичних сигналів в багатоантенних радіопеленгаторах.

Для початку задаємо моделі корисних сигналів та шумів і їх статистичні характеристики. Радіопеленгатори є пасивними пристроями, що приймають сигнали $s_k(t)$ від джерел радіовипромінювання, кутові положення яких необхідно визначити. Приймаючи ці сигнали рознесеними у просторі антенами можна стверджувати, що обробці підлягають просторово-часові поля $s_{r,i}(t, \vec{r}', \theta_{s,k})$. Використані наступні позначення: $s_k(t)$ – випромінений k -м джерелом сигнал, $k = \overline{1, K}$, $s_{r,i}(t, \vec{r}', \theta_{s,k})$ – прийняте i -ю антеною просторово-часове поле, t – час, $\vec{r}'$ – просторові координати поверхні антени, $\theta_{s,k}$ – кути візування кожного з k -го джерела, $i = \overline{1, N}$. Для спрощення задачі будемо вважати наступне: обробка просторового поля в межах кожної антени вже виконана, кожна антена характеризується діаграмою спрямованості $G_i(\theta - \theta_{0i})$, приймаємо стохастичні сигнали від одного джерела радіовипромінювання багатьма антенами, $k = 1$, джерело випромінювання є точковим. Модель стохастичного сигналу випромінювача не має визначеної функції, тому представимо її у вигляді перетворення Фур'є від спектру шуму

$$s(t) = \int_{-\infty}^{\infty} \dot{\Pi}(2\pi f) \dot{S}(2\pi f) e^{j2\pi f t} df. \quad (1)$$

Будемо вважати, що формування стохастичного сигналу полягає в обмеженні у спектрі білого гаусівського шуму функцією $\dot{\Pi}(2\pi f)$.

Характеристики $\dot{S}(2\pi f)$ наступні $\langle \dot{S}(2\pi f) \rangle = 0$,
 $\langle \dot{S}(2\pi f_1) \dot{S}^*(2\pi f_1) \rangle = 0, 5N_{0s} \delta(f_1 - f_2)$.

Використовуючи модель розсіяних електромагнітних хвиль [33], можна стверджувати, що сигнал на виході кожної з приймальних антен має вигляд

$$s_{r,i}(t, \theta_s) = K_{0,i} \int_{\Theta} G_i(\theta - \theta_{0,i}) \delta(\theta - \theta_s) \times \int_{-\infty}^{\infty} \dot{\Pi}(2\pi f) \dot{S}(2\pi f) e^{j2\pi f t - j\psi_i(f, \theta_s)} df d\theta. \quad (2)$$

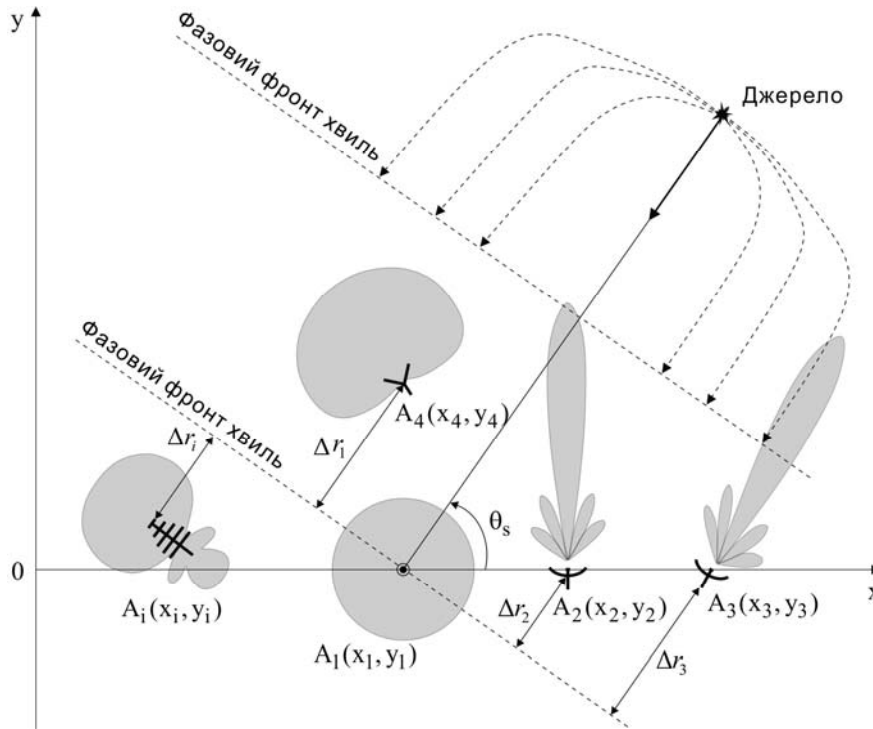


Рисунок 1 – Геометрія вимірювання кута візування θ_s в антенній решітці з довільним розміщенням різноманітних антен

Взаємна кореляційна функція сигналів на виходах i -го і j -го приймачів має вигляд

$$R_{s,ij}(t_1 - t_2, \theta_s) = \langle s_{r,i}(t_1, \theta_s) s_{r,j}(t_2, \theta_s) \rangle = K_{0,i} K_{0,j} G_i(\theta_s - \theta_{0,i}) G_j(\theta_s - \theta_{0,j}) \times \int_{-\infty}^{\infty} |\dot{\Pi}(2\pi f)|^2 e^{j2\pi f(t_1 - t_2) - j2\pi f[\Delta r_i(\theta_s) - \Delta r_j(\theta_s)]} df. \quad (3) \quad \text{де}$$

Кореляційна функція корисного сигналу в кожному окремому i -му приймачі дорівнює

На рис. 1 показана геометрія прийому антенами радіопеленгатора електромагнітного поля від джерела радіовипромінювання, що відповідає визначеним моделям сигналів. Вважається, що вимірювання кута візування точкового джерела стохастичного сигналу виконується в дальній зоні, тобто фронт випромінених електромагнітних хвиль є плоским в області їх реєстрації і промені від джерела до кожної антени є паралельними. Антени представлені в довільній формі, мають різну орієнтацію, просторове положення та діаграму спрямованості. З наведеної геометрії при відомих координатах антен, частоті і куті θ_s нескладно обчислити відстані затримки або випередження фронту хвилі $\Delta r_i(\theta_s)/c$ в кожній антені відносно фазового центру решітки.

$$R_{s,ii}(t_1 - t_2, \theta_s) = \langle s_{r,i}(t_1, \theta_s) s_{r,i}(t_2, \theta_s) \rangle = K_{0,i}^2 G_i^2(\theta_s - \theta_{0,i}) 0, 5N_{0s} R_h(t_1 - t_2), \quad (4)$$

$$R_h(t_1 - t_2) = \int_{-\infty}^{\infty} |\dot{\Pi}(2\pi f)|^2 e^{j2\pi f(t_1 - t_2)} df \quad (5)$$

– автокореляційна функція імпульсної характеристики вихідного тракту передавача стохастичного сигналу.

Математичне сподівання прийнятих сигналів визначимо наступним чином:

$$m_{s,i}(t) = \left\langle s_{r,i}(t, \theta_s) \right\rangle = K_{0,i} \int_{\Theta} G_i(\theta - \theta_{0,i}) \delta(\theta - \theta_s) \times \\ \times \int_{-\infty}^{\infty} \dot{P}(2\pi f) \left\langle \dot{S}(2\pi f) \right\rangle e^{j2\pi f t - j\psi_i(f, \theta_s)} df d\theta = 0. \quad (6)$$

В кожному приймачі до зареєстрованих корисних сигналів додаються внутрішні шуми $n_i(t)$, які будемо апроксимувати білими гаусівськими шумами з наступними статистичними характеристиками:

$$m_{n,i}(t) = \langle n_i(t) \rangle = 0 \quad (7)$$

– математичне сподівання внутрішнього шуму,

$$R_{n,ij}(t_1 - t_2) = \langle n_i(t_1) n_j(t_2) \rangle = 0 \quad (8)$$

– взаємна кореляційна функція внутрішніх шумів в різних каналах,

$$R_{n,i}(t_1 - t_2) = \langle n_i(t_1) n_i(t_2) \rangle = 0, 5 N_{0ni} \delta(t_1 - t_2) \quad (9)$$

– автокореляційна функція шуму в кожному i -му каналі. Для спрощення розрахунків будемо вважати, що спектральні щільності потужності $N_{0ni} / 2$ в кожному каналі є однаковою і дорівнює $N_{0ni} / 2$.

Коливання, що будуть підлягати подальшій оптимальній обробці, будемо називати рівняннями спостереження. Система рівнянь спостереження для представленої геометрії та умов вимірювання має наступний вигляд:

$$u_i(t) = s_{r,i}(t, \theta_s) + n_i(t). \quad (10)$$

Кореляційна функція спостережень (10) на виходах i -го та j -го каналів дорівнює

$$\begin{cases} R_{u,ij}(t_1 - t_2) = R_{s,ij}(t_1 - t_2, \theta_s), & \text{для } i \neq j, \\ R_{u,ii}(t_1 - t_2) = R_{s,ii}(t_1 - t_2, \theta_s) + \frac{N_{0ni}}{2} \delta(t_1 - t_2), & \text{для } i = j. \end{cases} \quad (11)$$

Математичні сподівання прийнятих сигналів в кожному вимірювачі дорівнюватимуть нулю, адже

$$m_{u,i}(t) = \langle u_i(t) \rangle = \\ = \langle s_{r,i}(t, \theta_s) \rangle + \langle 2 s_{r,i}(t, \theta_s) n_i(t) \rangle + \langle n_i(t) \rangle = 0. \quad (12)$$

Визначимо критерій максимуму функціоналу правдоподібності для вирішення оптимізаційних задач. Оптимізацію методів обробки рівнянь спостереження (10) будемо виконувати методом максимуму функціоналу правдоподібності. Даний метод полягає в визначенні оптимальної оцінки параметру λ за результатами пошуку максимуму функціоналу правдоподібності $P(\bar{u}(t) | \lambda)$, де $\bar{u}(t) = \|u_1(t), u_2(t), \dots, u_i(t)\|$. Функціонал $P(\bar{u}(t) | \lambda)$ – це умовний функціонал щільнос-

ті ймовірності реєстрації випадкових коливань $u_i(t)$ при фіксованому значенні λ . Для наведеної геометрії та задачі пеленгації джерел випромінювання стохастичних сигналів випадковими коливаннями є рівняння спостереження в кожній з антен (10), а параметром, що підлягає оцінюванню є кут візування $\lambda = \theta_s$.

Приклади функціоналів правдоподібності для задач дистанційного зондування землі та радіолокації наведені в роботах [34, 35]. Існує багато досліджень з використанням критерію функціоналу правдоподібності, проте кожна задача потребує своїх уточнень та обмежень. Використовуючи методіку конструювання широкого класу функціоналів правдоподібності для активних та пасивних радіовимірювачів [34, 35], запишемо $P(\bar{u}(t) | \lambda)$ для визначених умов вимірювання наступним чином:

$$P(\bar{u}(t) | \lambda = \theta_s) = k(\theta_s) \exp \left\{ -\frac{1}{2} \int_0^T \int_0^T [\bar{u}^T(t_1) - m_u(t_1)] \times \right. \\ \left. \times \underline{W}(t_1, t_2, \theta_s) [\bar{u}(t_2) - m_u(t_2)] dt_1 dt_2 \right\}, \quad (13)$$

де $k(\lambda)$ – нормуючий коефіцієнт, що немає фізичного змісту, проте має фізичний сенс його похідна. Виникає цей коефіцієнт під час граничного переходу від математичного представлення щільності ймовірності в дискретному вигляді до безперервного.

Функція $\underline{W}(t_1, t_2, \theta_s)$ знаходиться з інтегрального рівняння

$$\int_0^T \underline{R}_u(t_1, t_2, \theta_s) \underline{W}_u(t_2, t_3, \theta_s) dt_2 = \underline{I} \delta(t_1 - t_3). \quad (14)$$

Максимум функціонала (13) знайдемо шляхом його диференціювання і прирівнювання до нуля

$$\frac{dP(\bar{u}(t) | \theta_s)}{d\theta_s} = 0. \quad (15)$$

Частіше при вирішенні таких задач диференціюють не функціонал правдоподібності, а його логарифм, адже логарифм є монотонною функцією, що не змінює положення максимуму функціоналу $P(\bar{u}(t) | \theta_s)$. В такому випадку підставляючи (12) в (13) і знаходячи похідну від логарифму $P(\bar{u}(t) | \theta_s)$ в (15)

$$\frac{d \ln P(\bar{u}(t) | \theta_s)}{d\theta_s} = 0, \quad (16)$$

отримаємо рівняння правдоподібності

$$\begin{aligned} \text{spur} \int_0^T \int_0^T \frac{dR_u(t_1, t_2, \theta_s)}{d\theta_s} W_u(t_1, t_2, \theta_s) dt_1 dt_2 = \\ = - \int_0^T \int_0^T \bar{u}^T(t_1) \frac{dW_u(t_1, t_2, \theta_s)}{d\theta_s} \bar{u}(t_2) dt_1 dt_2. \end{aligned} \quad (17)$$

Подальша конкретизація кореляційної функції та функції, що їй обернена дозволить отримати оптимальний метод обробки $u_i(t)$ для оцінки кутового положення джерела випромінювання стохастичних сигналів.

Знайти аналітичний вираз $W_u(t_1, t_2, \theta_s)$ з (14) достатньо складно. Для спрощення розрахунків доцільно вираз (14) і рівняння правдоподібності (17) представити в спектральній області. Виконавши всі розрахунки в області частота, на кінцевому етапі можливо представити обробку у часі. Розглянемо більш детально такий підхід.

Вважаючи, що рівняння спостереження стаціонарні на інтервалі $(0, T)$ і час спостереження значно перевищує час кореляції, перейдемо в (14) до різнице-вих аргументів $\tau = t_1 - t_2$ і нескінченних меж інтегрування

$$\int_{-\infty}^{\infty} R_u(t_1 - t_2, \theta_s) W_u(t_2 - t_3, \theta_s) dt_2 = I \delta(t_1 - t_3). \quad (18)$$

Перетворення Фур'є від правої частини дорівнює одиничній матриці, а над лівою частиною виконаємо наступні дії

$$\begin{aligned} \int_0^T \int_{-\infty}^{\infty} \underline{G}_R(f, \theta_s) \exp(j2\pi f(t_1 - t_2)) df \left] W_u(t_2 - t_3, \theta_s) dt_2 = \right. \\ = \int_{-\infty}^{\infty} \underline{G}_R(f, \theta_s) \underline{G}_W(f, \theta_s) \exp(j2\pi f(t_1 - t_3)) df = \\ = F^{-1} \{ \underline{G}_R(f, \theta_s) \underline{G}_W(f, \theta_s) \}, \end{aligned} \quad (19)$$

де

$$\underline{G}_R(f, \theta_s) = F \{ R_u(t_1 - t_2, \theta_s) \} \quad (20)$$

– матриця спектральних щільностей потужності прийнятих коливань,

$$\underline{G}_W(f, \theta_s) = \underline{G}_R^{-1}(f, \theta_s) = F \{ W_u(t_1 - t_2, \theta_s) \} \quad (21)$$

– обернена матриця до $\underline{G}_R(f, \theta_s)$.

З отриманих розрахунків випливає, що перетворення Фур'є від правої та лівої частини рівняння (14) буде вираз

$$\underline{G}_R(f, \theta_s) \underline{G}_W(f, \theta_s) = I. \quad (22)$$

Перетворення Фур'є від рівняння правдоподібності (17) має наступний вигляд

$$\begin{aligned} T \text{ spur} \int_{-\infty}^{\infty} \frac{d\underline{G}_R(f, \theta_s)}{d\theta_s} \underline{G}_R^{-1}(f, \theta_s) df = \\ = - \int_{-\infty}^{\infty} \bar{U}_T^+(j2\pi f) \frac{d\underline{G}_R^{-1}(f, \theta_s)}{d\theta_s} \bar{U}_T(j2\pi f) df, \end{aligned} \quad (23)$$

де

$$\bar{U}_T(j2\pi f) = \int_0^T \bar{u}(t) e^{-j2\pi f t} dt = \int_{-\infty}^{\infty} \bar{u}_T(t) e^{-j2\pi f t} dt \quad (24)$$

– спектр укороченої інтервалом спостереження T реалізації векторного процесу $\bar{u}(t)$.

Таким чином у виразах (19)–(24) наведені вирази для вирішення оптимізаційної задачі в спектральній області і отримання методу обробки спостережуваних спектрів $\bar{U}_T(j2\pi f)$.

Граничні похибки оцінювання кутового положення. Теорія статистичного синтезу та аналізу методів обробки сигналів визначає граничні похибки визначення оцінок $\hat{\lambda}$, як нижню границю нерівності Крамера-Рао

$$\left\langle (\hat{\lambda} - \lambda)^2 | \lambda \right\rangle_{\lambda=\lambda_{\text{icT}}} \geq - \left\langle \frac{d^2 \ln P(\bar{u}(t) | \lambda)}{d\lambda^2} \right\rangle_{\lambda=\lambda_{\text{icT}}}^{-1}. \quad (25)$$

В такому випадку, дисперсія похибки оцінювання кута візування $\lambda = \theta_s$ матиме наступний вираз

$$\sigma_{\theta_s}^2 = - \left\langle \frac{1}{\frac{d^2 \ln P(\bar{u}(t) | \theta_s)}{d\theta_s^2}} \right\rangle_{\theta_s=\theta_{s \text{ icT}}}. \quad (26)$$

Аналіз аналітичних виразів дисперсії $\sigma_{\theta_s}^2$ має значну актуальність на етапі планування експериментів, зокрема для визначення оптимальної геометрії антенної решітки та просторової орієнтації діаграм спрямованості кожної з її елементів. Під оптимальністю розуміється досягнення мінімальних похибок оцінювання кута θ_s .

4 ЕКСПЕРИМЕНТИ

Для підтвердження ефективності запропонованих теоретичних основ синтезу оптимальних методів вимірювання кутового положення джерела стохастич-

них сигналів розглянемо елементарний випадок одноантенного нескануючого радіопеленгатора [36].

Постановка оптимізаційної задачі. За критерієм максимуму функціоналу правдоподібності необхідно на інтервалі часу $t \in [0, T]$ оптимальним чином оцінити кутове положення джерела радіовипромінювання θ_s за результатами прийому в одноантенному радіопеленгаторі корисного сигналу $s_r(t, \theta_s)$ на фоні внутрішніх шумів $n(t)$.

Вирішення оптимізаційної задачі. Для визначених умов проведення вимірювання рівняння правдоподібності набуде наступного вигляду

$$u(t) = s_r(t, \theta_s) + n(t), \quad (27)$$

де

$$s_r(t, \theta_s) = K_0 \int_{\Theta} G(\theta - \theta_0) \delta(\theta - \theta_s) \times \int_{-\infty}^{\infty} \dot{\Gamma}(2\pi f) \dot{S}(2\pi f) e^{j2\pi f t} df d\theta. \quad (28)$$

На відміну від загальної моделі сигналів (2) в отриманому виразі (28) відсутня інформація про фазовий зсув, що пропорційний кутовому положенню джерела радіовипромінювання. Тобто в одноантенному радіопеленгаторі вся інформація по кутове положення джерела радіовипромінювання міститься в амплітуді прийнятих коливань.

Кореляційна функція рівняння спостереження дорівнює

$$R_u(t_1 - t_2, \theta_s) = K_0^2 G^2(\theta_s - \theta_0) \frac{N_{0s}}{2} R_h(t_1 - t_2) + \frac{N_{0n}}{2} \delta(t_1 - t_2). \quad (29)$$

Розрахувавши перетворення Фур'є від (29), отримаємо спектральну щільність потужності прийнятих коливань

$$G_R(f, \theta_s) = K_0^2 G^2(\theta_s - \theta_0) \frac{N_{0s}}{2} |\dot{\Gamma}(2\pi f)|^2 + \frac{N_{0n}}{2}. \quad (30)$$

Похідну від спектральної щільності потужності, що входить в (23), розрахуємо наступним чином

$$\frac{dG_R(f, \theta_s)}{d\theta_s} = K_0^2 \frac{dG^2(\theta_s - \theta_0)}{d\theta_s} \frac{N_{0s}}{2} |\dot{\Gamma}(2\pi f)|^2. \quad (31)$$

Вираз для матриці $\underline{G}_W(f, \theta_s)$ у випадку оптимізації структури одноантенного пеленгатора стає функцією і дорівнює $G_W(f, \theta_s) = G_R^{-1}(f, \theta_s)$, а похідна від цієї функції матиме вигляд

$$\frac{dG_R^{-1}(f, \theta_s)}{d\theta_s} = -\frac{1}{G_R^{-2}(f, \theta_s)} \frac{dG_R(f, \theta_s)}{d\theta_s}. \quad (32)$$

Підставляючи (30) і (31) в (23), вирішимо рівняння правдоподібності для випадку реєстрації сигналів в одній антенні

$$\int_{-\infty}^{\infty} \frac{dG_R(f, \theta_s)}{d\theta_s} G_R^{-1}(f, \theta_s) df = \frac{1}{T} \int_{-\infty}^{\infty} |\dot{U}_T(j2\pi f)|^2 \frac{1}{G_R^{-2}(f, \theta_s)} \frac{dG_R(f, \theta_s)}{d\theta_s} df, \quad (33)$$

в наступному вигляді

$$\int_{-\infty}^{\infty} \frac{|\dot{\Gamma}(2\pi f)|^2}{\left[K_0^2 G^2(\theta_s - \theta_0) \frac{N_{0s}}{2} |\dot{\Gamma}(2\pi f)|^2 + \frac{N_{0n}}{2} \right]} df = \frac{1}{T} \int_{-\infty}^{\infty} |\dot{U}_T(j2\pi f)|^2 \times \frac{|\dot{\Gamma}(2\pi f)|^2}{\left[K_0^2 G^2(\theta_s - \theta_0) \frac{N_{0s}}{2} |\dot{\Gamma}(2\pi f)|^2 + \frac{N_{0n}}{2} \right]^2} df. \quad (34)$$

Для визначення θ_s помножимо і поділимо в лівій частині рівняння (34) підінтегральний вираз на знаменник, винесемо за дужки $K_0^2 G^2(\theta_s - \theta_0) \frac{N_{0s}}{2}$ і розв'яжемо отримане рівняння відносно $G^2(\theta_s - \theta_0)$

$$G^2(\theta_s - \theta_0) \int_{-\infty}^{\infty} \frac{\left[|\dot{\Gamma}(2\pi f)|^2 + \frac{1}{q} \right] |\dot{\Gamma}(2\pi f)|^2}{\left[|\dot{\Gamma}(2\pi f)|^2 + \frac{1}{q} \right]^2} df = \frac{1}{K_0^2 \frac{N_{0s}}{2}} \frac{1}{T} \int_{-\infty}^{\infty} |\dot{U}_T(j2\pi f)|^2 \frac{|\dot{\Gamma}(2\pi f)|^2}{\left[|\dot{\Gamma}(2\pi f)|^2 + \frac{1}{q} \right]^2} df, \quad (35)$$

де $q = (K_0^2 G^2(\theta_s - \theta_0) N_{0s}) / N_{0n}$ – співвідношення сигнал-завада в прийнятих коливаннях $u(t)$.

Права частина (35) представляє собою обробку прийнятих коливань в спектральній області. Спочатку з $u(t)$ необхідно сформувати періодограму Фур'є $|\dot{U}_T(j2\pi f)|^2$. Далі періодограму необхідно пропустити через узгоджений фільтр $|\dot{\Gamma}(2\pi f)|^2$ і декорелюючий фільтр $1/\left[|\dot{\Gamma}(2\pi f)|^2 + \frac{1}{q} \right]^2$. Узгоджений фільтр має повністю повторювати частотну характеристику передавача стохастичного сигналу для оптимальної фільтрації внутрішніх шумів приймача. Декорелюю-

чий фільтр розширює результуючий спектр частот прийнятих коливань і дозволяє збільшити кількість некорельованих відліків при оцінюванні кута θ_s . Ступінь декореляції залежить від співвідношення q , чим більше рівень сигналу над шумами тим краще виділяються придушені компоненти інформативного спектру. Після фільтрації періодограми в зазначених фільтрах необхідно виконати її усереднення за частотою і нормування на T для оцінки потужності прийнятих коливань.

В лівій частині (35), в чисельнику дробу під інтегралом також є величина $1/q$, якою можна знехтувати, адже на практиці доцільно обирати $q > 100$. В такому випадку маємо

$$\int_{-\infty}^{\infty} |\dot{P}(2\pi f)|^2 \frac{|\dot{P}(2\pi f)|^2}{\left[|\dot{P}(2\pi f)|^2 + \frac{1}{q}\right]^2} df = 2\Delta F. \quad (36)$$

З урахуванням введених операцій отримуємо

$$G^2(\theta_s - \theta_0) 2\Delta F = \frac{1}{K_0^2 N_{0s} / 2} \hat{P}_{s,T}. \quad (37)$$

Якщо вважати, що шум з прийнятих коливань відфільтрований повністю, або його вплив є незначним, то

$$\hat{P}_{s,T} \approx K_0^2 G^2(\theta_s - \theta_0) N_{0s} \Delta F. \quad (38)$$

Поділивши ліву і праву частину (37) на $2\Delta F$, отримуємо оцінку квадрату діаграми спрямованості в напрямку на джерело випромінювання

$$\begin{aligned} \hat{G}^2(\theta_s - \theta_0) &= \frac{1}{K_0^2 N_{0s} \Delta F} \hat{P}_{s,T} = \\ &= \frac{1}{K_0^2 N_{0s} \Delta F T} \int_{-\infty}^{\infty} |\dot{U}_{TW}(j2\pi f)|^2 df, \end{aligned} \quad (39)$$

де

$$\dot{U}_{TW}(j2\pi f) = \dot{U}_T(j2\pi f) \left[\frac{\dot{P}(2\pi f)}{|\dot{P}(2\pi f)|^2 + q^{-1}} \right]. \quad (40)$$

Згідно з теоремою Парсеваля-Лапласа

$$\int_{-\infty}^{\infty} |\dot{U}_{TW}(j2\pi f)|^2 df = \int_0^T u_{TW}^2(t) dt, \quad (41)$$

де

$$u_{TW}(t) = \int_{-\infty}^{\infty} \dot{U}_{TW}(j2\pi f) e^{j2\pi f t} df \quad (42)$$

– декорельоване у часі рівняння спостереження. В такому випадку оптимальний метод оцінки кутового положення джерела стохастичного сигналу в часовій області матиме вигляд

$$\hat{G}^2(\theta_s - \theta_0) = \frac{1}{K_0^2 N_{0s} \Delta F} \hat{P}_{s,T} = \frac{1}{K_0^2 N_{0s} \Delta F} \frac{1}{T} \int_0^T u_{TW}^2(t) dt. \quad (43)$$

Отримана оцінка в (39) і (43) не є оцінкою кута θ_s , так як цей кут є аргументом квадрату діаграми спрямованості або деякої функції

$L[\theta_s] = G^2(\theta_s - \theta_0)$. Бажано, щоб функція $L[\theta_s]$ була лінійною, тоді оцінка θ_s матиме вигляд

$$\hat{\theta}_s = L^{-1} \left[\frac{1}{K_0^2 N_{0s} \Delta F} \hat{P}_{s,T} \right] + \theta_0. \quad (44)$$

Необхідно зауважити, що кут θ_0 має бути відомим, адже це опорний напрямок від якого визначається кутове положення джерела радіовипромінювання. Звичайно, що більшість реальних діаграм спрямованості є нелінійними функціями, такими як гаусівська функція, $\text{sinc}[\theta_s - \theta_0]$ тощо. В цьому випадку оцінка $\hat{\theta}_s$ є неоднозначною і необхідно додатково використовувати обмеження по зміні положення антени або обирати лише лінійну ділянку діаграми спрямованості для вимірювання.

Також згідно з розвинутою теорією можливо визначити аналітичний вираз для граничної дисперсії оцінки кутового положення джерела стохастичного сигналу. Запишемо другу похідну від логарифму функціоналу правдоподібності у виразі (26) наступним чином

$$\begin{aligned} \frac{d^2 \ln P(\vec{u}(t) | \theta_s)}{d\theta_s^2} = \\ = -\frac{1}{2} \int_0^T \int_0^T \frac{d^2 R_u(t_1, t_2, \theta_s)}{d\theta_s^2} W_u(t_1, t_2, \theta_s) dt_1 dt_2 - \\ -\frac{1}{2} \int_0^T \int_0^T \frac{dR_u(t_1, t_2, \theta_s)}{d\theta_s} \frac{dW_u(t_1, t_2, \theta_s)}{d\theta_s} dt_1 dt_2 - \\ -\frac{1}{2} \int_0^T \int_0^T \frac{d^2 W_u(t_1, t_2, \theta_s)}{d\theta_s^2} u(t_1) u(t_2) dt_1 dt_2. \end{aligned} \quad (45)$$

Підставляючи (45) в (26) отримуємо дисперсію

$$\sigma_{\theta_s}^2 = \left[-\frac{1}{2} \int_0^T \int_0^T \frac{dR_u(t_1, t_2, \theta_s)}{d\theta_s} \frac{dW_u(t_1, t_2, \theta_s)}{d\theta_s} dt_1 dt_2 \right]^{-1} \Bigg|_{\theta_s = \theta_{s \text{ icm}}} \quad (46)$$

Для конкретизації (46) перепишемо отриманий вираз у спектральній області. Для цього будемо вважати, що рівняння спостереження стаціонарні і можливо перейти у виразах до різницевого аргументу $\tau = t_1 - t_2$

$$\sigma_{\theta_s}^2 = \left[-\frac{T}{2} \int_{-T}^T \left(1 - \frac{|\tau|}{T} \right) \frac{dR_u(\tau, \theta_s)}{d\theta_s} \frac{dW_u(\tau, \theta_s)}{d\theta_s} d\tau \right]^{-1} \Bigg|_{\theta_s = \theta_{s \text{ icm}}} \quad (47)$$

Зазвичай при використанні стохастичних зондуючих сигналів час кореляції τ є значно меншим за час спостереження, тому можна вважати $|\tau| \ll T$ і $(1 - |\tau|/T) \approx 1$. З урахуванням приведених припущень перепишемо вираз (47)

$$\sigma_{\theta_s}^2 = \left[-\frac{T}{2} \int_{-T}^T \frac{dR_u(\tau, \theta_s)}{d\theta_s} \frac{dW_u(\tau, \theta_s)}{d\theta_s} d\tau \right]^{-1} \Bigg|_{\theta_s = \theta_{s \text{ icm}}} \quad (48)$$

або в спектральній області

$$\sigma_{\theta_s}^2 = \left[\frac{T}{2} \int_{-\infty}^{\infty} \frac{dG_R(-f, \theta_s)}{d\theta_s} \frac{dG_R(f, \theta_s)}{d\theta_s} \frac{1}{G_R^2(f, \theta_s)} df \right]^{-1} \Bigg|_{\theta_s = \theta_{s \text{ icm}}} \quad (49)$$

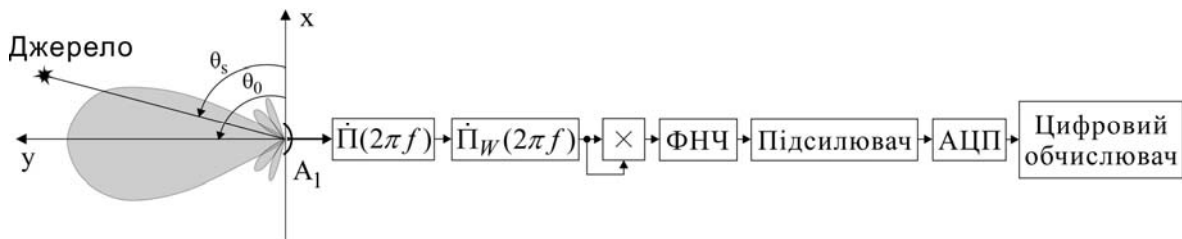


Рисунок 2 – Структурна схема оптимального одноантенного радіопеленгатору джерела стохастичних сигналів

Робота схеми на рис. 2 полягає в наступному. Антена, що орієнтована в напрямку θ_0 приймає електромагнітні коливання від джерела стохастичних сигналів з напрямку θ_s . Сигнали на виході антени послідовно обробляються в узгодженому фільтрі з частотною характеристикою $\dot{G}(2\pi f)$ і в декорелюючому

Підставляючи (31) в (49), конкретизуємо вираз для граничних похибок оцінок кута

$$\sigma_{\theta_s}^2 = \frac{G^4(\theta_s - \theta_0)}{T \Delta F \left(\frac{dG^2(\theta_s - \theta_0)}{d\theta_s} \right)^2} \Bigg|_{\theta_s = \theta_{s \text{ icm}}} \quad (50)$$

З отриманого виразу (50) випливає, що гранична дисперсія похибки оцінювання кута візування джерела стохастичного сигналу обернено пропорційна часу спостереження, ширині спектру частот прийнятих сигналів після фільтрації в узгодженому фільтрі та декорелюючому фільтрі, а також нормованій крутизні квадрату діаграми спрямованості в точці істинного напрямку на джерело випромінювання. Залежність граничних похибок від нормованої крутизни $G^2(\theta_s - \theta_0)$ впливає з необхідності вирішувати обернену задачу визначення θ_s з рівняння

$$L[\theta_s] = \hat{G}^2(\theta_s - \theta_0) \quad .$$

5 РЕЗУЛЬТАТИ

З аналізу методу оцінки $\hat{\theta}_s$ (44) слідує практичний результат – структура одноантенного пеленгатора джерела стохастичних сигналів, що показана на рис. 2.

фільтрі з $\dot{G}_W(2\pi f) = 1 / \left[\left| \dot{G}(2\pi f) \right|^2 + q^{-1} \right]$. Після фільтрації сигнал надходить на квадратичний детектор і усереднюється у фільтрі нижніх частот.

Результат оцінки потужності прийнятих коливань нормується в підсилювачі до необхідного рівня, що може бути оцифрованим в аналого-цифровому перетворювачі, і далі перетворюється в цифровий вигляд

та оброблюється в цифровому обчислювачі. В цифровому обчислювачі реалізуються операції нормування отриманих коливань на величини $K_0^2 N_{0s} \Delta F$ та вирішується обернена задача виділення θ_s з $L[\cdot]$.

Окрім реалізації різних математичних операцій розрахунку θ_s за отриманими оцінками потужності можливо також накладати певні умови на вимірювання або виконувати сканування діаграмою спрямованості всього діапазону кутів. Так, можливо визначити умову, що оцінка θ_s при зміні θ_0 матиме оптимальне

значення $\hat{\theta}_s = \theta_0$ при досягненні правою стороною найбільшого значення, адже ліва сторона для реальних діаграм спрямованості досягає максимуму при $\theta_s - \theta_0 = 0$.

Відповідно до структурної схеми була розроблена імітаційна модель в програмному забезпеченні Simulink від компанії The MathWorks, що зображена на рис. 3. До складу моделі входять блоки Формування сигналу передавача, Вплив внутрішніх шумів, Кла-

сичний одноканальний пеленгатор, Модифікований одноканальний пеленгатор, Оцінка точності класичної обробки, Оцінка точності модифікованої обробки, Індикатор результату вимірювань, Індикатор оцінки точності.

Результат оцінки точності вимірювання кутового положення джерела радіовипромінювання показано на рис. 4 і рис. 5. На рис. 4 показана зміна кутового положення джерела стохастичного сигналу, що задається слайдером та результат його оцінювання. Рис. 5 ілюструє зміну середньоквадратичного відхилення оцінок від істинного значення кута для класичного та модифікованого методів. Для моделювання були вибрані наступні параметри: діаграма спрямованості гаусівська з шириною 120° , $N_{0s} = 1 \text{ Вт/Гц}$, $\Delta F = 1 \text{ МГц}$, $T = 1 \text{ с}$, $q = 30 \text{ дБ}$, $K_0 = 1$.

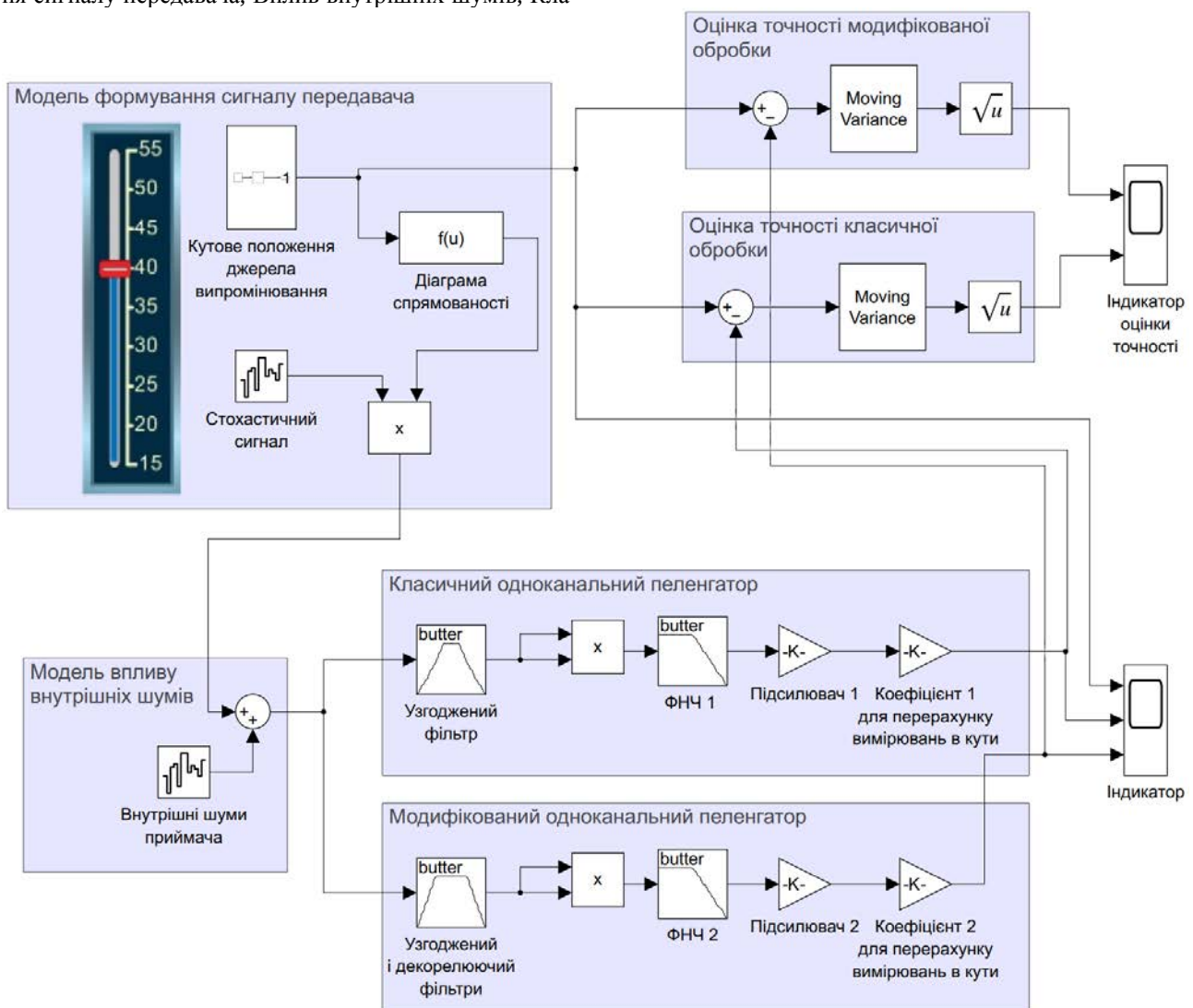


Рисунок 3 – Імітаційна модель дослідження точності роботи класичного та модифікованого пеленгаторів



Рисунок 4 – Емпіричні результати вимірювань: чорна лінія – істинне значення кута, червона лінія – оцінка класичним методом, зелена лінія – оцінка модифікованим методом

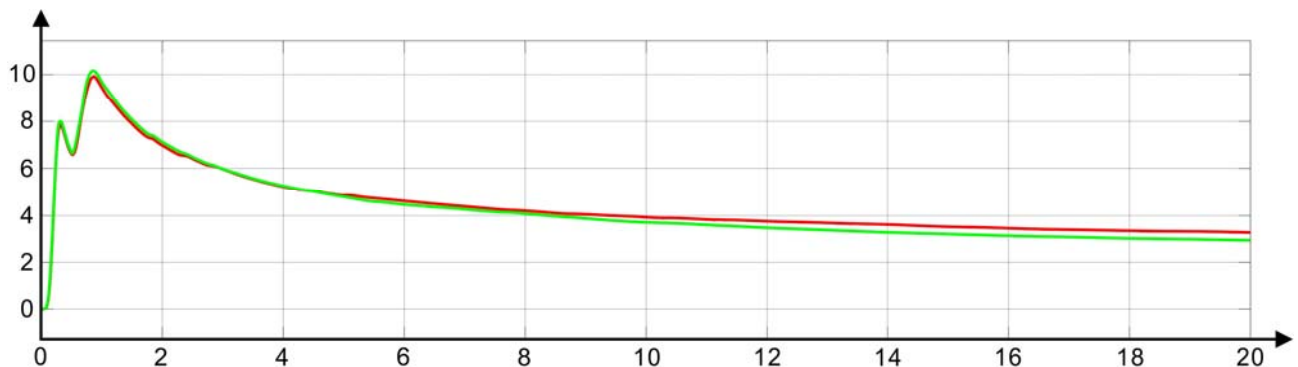


Рисунок 5 – Середньоквадратичне відхилення оцінок: червона лінія – класичний метод, зелена лінія – модифікований метод

З отриманих графіків слідє, що точність вимірювання кута для визначених параметрів з використанням класичного методу пеленгації складає $3,5^\circ$. Модифікований метод показує підвищення точності на 10% за рахунок оптимальної операції декореляції прийнятих коливань.

6 ОБГОВОРЕННЯ

Отримані результати у вигляді структури радіопеленгатора і імітаційної моделі дозволили підтвердити працездатність методів, що отримані з використанням запропонованих теоретичних основ оптимізації радіосистем вимірювання кутового положення джерел випромінювання стохастичних сигналів. В той самий час виграш в точності був отриманий при співвідношенні сигнал-завада 30 дБ, що є досить великим для сучасних радіовимірювальних систем. Для обговорення залежності результатів вимірювань від рівня шумів в розроблені моделі були проведенні численні моделювання при різних q . Результати наведені в табл. 1

Таблиця 1 – Залежності точності радіопеленгації від співвідношення сигнал-шум

Співвідношення сигнал-шум	Похибка оцінювання кута класичним методом	Похибка оцінки кута модифікованим методом	Виграш в точності
30 дБ	$3,257^\circ$	$2,923^\circ$	10%
23 дБ	3,569	$3,143^\circ$	12%
20 дБ	3,764	$3,304^\circ$	12,2%
10 дБ	6,116	5,455	10,8%
9 дБ	6,644	5,952	10,4%
6 дБ	8,968	8,146	9,2%
4 дБ	11,37	10,42	8,3%
3 дБ	12,8	11,74	8,3%

З аналізу отриманих в табл.1 похибок випливає, що зі зменшенням співвідношення q точність визначення кута погіршується. При цьому до 20 дБ погіршення складає $0,05^\circ$ на 1 дБ, а після 20 дБ втрата точності значно погіршується і складає одиниці градусів на 1 дБ. Виграш запропонованого методу до 9 дБ складає від 10% до 12%, а після цього значення співвідношення сигнал-завада виграш починає постійно зменшуватися. Розрахункове потенціальне значення σ_{θ_s} за формулою (50) складає $1,581^\circ$. З цього можна зробити висновки, що для отримання близької до потенціальної точності q має бути не менше 20 дБ, а для отримання виграшу за новим модифікованим методом – не менше 9 дБ.

ВИСНОВКИ

Наукова новизна. У роботі представлено теоретичні основи синтезу радіопеленгаторів довільної конфігурації, які враховують різницю у формах діаграм спрямованості, довільну орієнтацію та просторове положення окремих пеленгаторів. Наукова новизна роботи полягає у розробці моделей сигналів, шумів та їх статистичних характеристик, а також у використанні критерію максимуму функціоналу правдоподібності для вирішення оптимізаційних задач синтезу радіопеленгаторів. Вперше показано, що представлення рівняння правдоподібності у спектральній області дозволяє визначити фізичний зміст обернених кореляційних матриць, що використовується в оптимальній обробці сигналів.

Практична значимість дослідження полягає у формулюванні методики наскрізного синтезу нових методів, що дозволяють визначити гранично досяжні похибки вимірювань. Запропоновані підходи можуть бути використані для побудови радіопеленгаторів із підвищеною точністю та адаптованістю до різних умов роботи. Ефективність розроблених теоретичних основ підтверджено на прикладі вирішення задачі вимірювання кутового положення джерела радіовипромінювання за допомогою одноантенного нескануючого радіопеленгатора.

Для майбутніх досліджень. У перспективі результати роботи можуть бути використані для подальшого розвитку теорії синтезу радіопеленгаторів у багатоканальних системах, дослідження їх ефективності в умовах високого рівня шумів, а також для реалізації в багатопозиційних радіолокаційних системах. Крім того, запропоновані підходи можуть бути розширені для пеленгації джерел із змінними характеристиками випромінювання.

ПОДЯКИ

Робота виконана за фінансової підтримки Міністерства освіти і науки України. Номери державної реєстрації проєктів: 0123U102002 і 0123U102000.

ЛІТЕРАТУРА

1. Monopulse direction finder for harmonic radar system / [I. Hertl, M. Strýček, V. Poláček, V. Brno] // Proceedings of the 18th International Conference on Microwaves, Radar and Wireless Communications. – Vilnius, Lithuania, 2010. – P. 1–3.
2. A new method of multi-frequency active aperture synthesis for imaging of SAR blind zone under aerospace vehicle / [V. Pavlikov, V. Volosyuk, S. Zhyla et al.] // Proceedings of the 14th International Conference The Experience of Designing and Application of CAD Systems in Microelectronics (CADSM). – 2017. – P. 118–120. DOI: 10.1109/CADSM.2017.7916099.
3. Zhang C. Direction Finding in MIMO Radar With Unknown Mutual Coupling / C. Zhang, H. Huang, B. Liao // IEEE Access. – 2017. – Vol. 5. – P. 4439–4447. DOI: 10.1109/ACCESS.2017.2684465.
4. Pavlikov V. V. Structural optimization of Dicke-type radiometer / V. V. Pavlikov, S. S. Zhyla, O. V. Odokienko // Proceedings of the 2nd International Young Scientists Forum on Applied Physics and Engineering (YSF). – 2016. – P. 171–174. DOI: 10.1109/YSF.2016.7753829.
5. Radiometer with signal energies ratio / [V. V. Pavlikov, S. S. Zhyla, A. V. Odokienko, M. O. Antonov] // Proceedings of the IEEE Radar Methods and Systems Workshop (RMSW). – 2016. – P. 99–102. DOI: 10.1109/RMSW.2016.7778560.
6. Cwalina K. Wideband Radio Direction Finder Implemented in Software Defined Radio Technology / K. Cwalina, P. Rajchowski, J. Sadowski // Applied Mechanics and Materials. – 2016. – Vol. 817. – P. 348. DOI: 10.4028/www.scientific.net/amm.817.348.
7. Ziolkowski C. Radio Bearing of Sources with Directional Antennas in Urban Environment / C. Ziolkowski, J. M. Kelter // International Journal of Microwave and Wireless Technologies. – 2018. – Vol. 10, № 7. – P. 759–771. DOI: 10.1017/S1759078718000648.
8. Gulko V. L. Polarization Amplitude-Phase Direction Finding Methods in Two-Canal UHF Radio Beacon Navigation Systems / V. L. Gulko, A. A. Mescheryakov // arXiv. – 2020. DOI: 10.48550/arXiv.2012.12728.
9. Direction Finding // Wireless Communications Systems. – John Wiley & Sons, 2020. – P. 301–323. DOI: 10.1002/9781119419204.ch8.
10. Cook H. A. Radio Direction-Finding Techniques for an Unmanned Aerial Vehicle / H. A. Cook, M. T. E. Kahn, V. Balyan // Micro-Electronics and Telecommunication Engineering. Lecture Notes in Networks and Systems. – Springer, Singapore, 2020. – Vol. 106. – P. 1. DOI: 10.1007/978-981-15-2329-8_1.
11. Active Aperture Synthesis Radar for High Spatial Resolution Imaging / [V. Pavlikov, V. Volosyuk, S. Zhyla et al.] // Proceedings of the 9th International Conference on Ultrawideband and Ultrashort Impulse Signals (UWBUSIS). – 2018. – P. 252–255. DOI: 10.1109/UWBUSIS.2018.8520021.
12. UWB Active Aperture Synthesis Radar: Operating Principle and Development of the Radar Block Diagram / [V. Pavlikov, V. Volosyuk, S. Zhyla et al.] // Proceedings of the IEEE Microwaves, Radar and Remote Sensing Symposium (MRRS). – 2017. – P. 27–30. DOI: 10.1109/MRRS.2017.8075018.
13. Novak A. Radio direction finding in air traffic services / A. Novak // Promet-Traffic-Traffico. – 2005. – Vol. 17, № 5. – P. 273–276.
14. Ilčev D. S. New Aspects of Progress in the Modernization of the Maritime Radio Direction Finders (RDF) / D. S. Ilčev // Transactions on Maritime Science. – 2021. – Vol. 10, № 01. – P. 68–83. DOI: 10.31861/sisiot2023.2.02006.
15. Pavlikov V. Optimal signal processing for radiometric imaging with multi-antenna & multi-band passive radars / V. Pavlikov, S. Zhyla, O. V. Odokienko // Proceedings of the International Conference on Antenna Theory and Techniques (ICATT). – 2015. – P. 7–12. DOI: 10.1109/ICATT.2015.7136821.
16. Statistical Synthesis and Analysis of Functionally Deterministic Signal Processing Techniques for Multi-Antenna

- Direction Finder Operation / [S. Zhyla, E. Tserne, Y. Volkov et al.] // *Computation*. – 2024. – Vol. 12, № 9. – P. 170. DOI: 10.3390/computation12090170.
17. Volosyuk V. Statistical Theory of Optimal Functionally Deterministic Signals Processing in Multichannel Aerospace Imaging Radar Systems / V. Volosyuk, S. Zhyla // *Computation*. – 2022. – Vol. 10, № 12. – P. 213. DOI: 10.3390/computation10120213.
18. Chen Z. Cooperative Energy Detection Algorithm Based on Background Noise and Direction Finding Error / [Z. Chen, Y. Zhang] // *AEU – International Journal of Electronics and Communications*. – 2018. – Vol. 95. – P. 326–341. DOI: 10.1016/j.aeue.2018.08.029.
19. Boulogeorgos A. A. Energy Detection Spectrum Sensing Under RF Imperfections / A. A. Boulogeorgos, N. D. Chatzidiamantis, G. K. Karagiannidis // *IEEE Transactions on Communications*. – 2016. – Vol. 64, № 7. – P. 2754–2766. DOI: 10.1109/TCOMM.2016.2561294.
20. Spatial-Temporal Opportunity Detection for Spectrum-Heterogeneous Cognitive Radio Networks: Two-Dimensional Sensing / [Q. Wu, G. Ding, J. Wang, Y. D. Yao] // *IEEE Transactions on Wireless Communications*. – 2013. – Vol. 12, № 2. – P. 516–526. DOI: 10.1109/TWC.2012.122212.111638.
21. Selection of Number of Compensatory Channels and Placement of Receivers with Non-Identical Frequency Response at Phased-Array Radar under Gaussian Noise Jamming / [V. Riabukha, A. Semeniaka, Y. Katiushyn et al.] // *Radioelectronics and Communications Systems*. – 2022. – Vol. 65, № 1. – P. 1–10. DOI: 10.3103/S0735272722010010.
22. Riabukha V. P. Analysis of Protection Efficiency of Phased-Array Radars Against the Gaussian Noise Jamming / V. P. Riabukha, A. V. Semeniaka, Y. A. Katiushyn // *Telecommunications and Radio Engineering*. – 2021. – Vol. 80, № 2. – P. 1. DOI: 10.1615/TelecomRadEng.2021036556.
23. Holland R. L. The Statistical Theory of Radio Direction Finding / R. L. Holland // *Technical Memorandum AccNr ADA104156*. – US Defense Technical Information Center. – 1981.
24. Statistical Estimates of the Accuracy of Mono-Pulse Direction Finding of Sources of Radio Emission by Dual-Channel Phase Devices / [V. Voitko, A. Ilnyckij, O. Iliashov et al.] // *Visnyk NTUU KPI Serii A – Radiotekhnika Radioaparotobuduvannia*. – 2020. – № 81. – P. 30–37. DOI: 10.20535/RADAP.2020.81.30-37.
25. Fabirovskyy S. Research and Development of a High-Precision Direction Finder for Radio Emission Sources Detecting / S. Fabirovskyy, V. Storozh, A. Solomko // *Information and Communication Technologies, Electronic Engineering*. – 2023. – Vol. 3, № 10. – P. 158–169. DOI: 10.23939/ictee2023.02.158.
26. Impact of Mutual Coupling on Wideband Adcock Direction Finders / [S. Henault, Y. M. Antar, S. Rajan et al.] // *Proceedings of the Canadian Conference on Electrical and Computer Engineering (CCECE)*. – 2008. – P. 001327–001332. DOI: 10.1109/CCECE.2008.4564755.
27. Development of Radio Direction Finder Using 6 Log Periodic Dipole Array Antennas / [K. N. Purwanto, A. Yahya, N. H. H. Khamis et al.] // *Proceedings of the 5th International Conference on Information Technology, Computer, and Electrical Engineering (ICITACEE)*. – 2018. – P. 157–160. DOI: 10.1109/ICITACEE.2018.8576965.
28. Polikarovskiykh O. Implementing the Search Algorithm of the Correlation Interferometer Direction Finder Through the GNU Radio Software Platform / O. Polikarovskiykh, I. Hula // *Security of Infocommunication Systems and Internet of Things*. – 2023. – Vol. 1, № 2. – P. 02006. DOI: 10.31861/sisiot2023.2.02006.
29. Combining Target Detection Against the Background of Jamming Signals and Jamming Signal DOA Estimation / [D. I. Lekhovyt'skiy, D. V. Atamanskiy, V. P. Riabukha, et al.] // *Proceedings of the International Conference on Antenna Theory and Techniques (ICATT)*. – 2015. – P. 1–5. DOI: 10.1109/ICATT.2015.7136777.
30. Atamanskiy D. V. Noise Emissions Sources Direction-Finding in the Process of Their Background Air Threats Detection in Radars with Phased Antenna Array / D. V. Atamanskiy // *Radioelectronics and Communications Systems*. – 2017. – Vol. 60, № 7. – P. 303–311. DOI: 10.3103/S0735272717070032.
31. Spectral Estimation Methods for a Joint System of the Non-Noise-Like Targets Detection and the Noise Radiating Sources Localization / [D. V. Atamanskiy, V. P. Riabukha, V. M. Kartashov, et al.] // *Radio Electronics, Computer Science, Control*. – 2022. – № 1. – P. 7. DOI: 10.15588/1607-3274-2022-1-1.
32. Synthesis of an Optimal Algorithm for Processing Random Signals during Phase Direction Finding / [A. O. Zhukov, I. N. Valyaev, V. P. Kovalenko et al.] // *Journal of Physics: Conference Series*. – 2020. – Vol. 1679, № 3. – P. 032021. DOI: 10.1088/1742-6596/1679/3/032021.
33. Volosyuk V. Phenomenological Description of the Electromagnetic Field and Coherent Images in Radio Engineering and Optical Systems / V. Volosyuk, V. Pavlikov, S. Zhyla // *Proceedings of the International Conference on Mathematical Methods in Electromagnetic Theory (MMET)*. – 2018. – P. 302–305. DOI: 10.1109/MMET.2018.8460321.
34. Volosyuk V. K. Statistical Theory of Radio Engineering Systems for Remote Sensing and Radar / V. K. Volosyuk, V. F. Kravchenko // *Fizmatlit*. – Moscow, Russia. – 1988.
35. Modern Methods for Optimal Spatio-Temporal Signal Processing in Active, Passive, and Combined Active-Passive Radio-Engineering Systems / [V. K. Volosyuk, Y. V. Gulyaev, V. F. Kravchenko, et al.] // *Journal of Communications Technology and Electronics*. – 2014. – Vol. 59, № 2. – P. 97–118. DOI: 10.1134/S1064226914020090.
36. Single Antenna Power Measurements Based Direction Finding / [J. P. Lie, T. Blu, C. M. S. See] // *IEEE Transactions on Signal Processing*. – 2010. – Vol. 58, № 11. – P. 5682–5692. DOI: 10.1109/TSP.2010.2065227.

Received 03.06.2025.
Accepted 20.10.2025.

THEORETICAL FOUNDATIONS OF THE STRUCTURE OF MULTI-ANTENNA RADIO DIRECTION FINDERS OPTIMISATION FOR DETERMINING THE STOCHASTIC SIGNAL SOURCES POSITION

Zhyla S. S. – Dr. Sc., Associate Professor, Head of the Department of Aerospace Radio-Electronic Systems, National Aerospace University “Kharkiv Aviation Institute”, Kharkiv, Ukraine.

Tserne E. O. – PhD, Junior Researcher of the Department of Aerospace Radio-Electronic Systems, National Aerospace University “Kharkiv Aviation Institute”, Kharkiv, Ukraine.

Zhyla O. V. – Candidate of Physical and Mathematical Sciences, Associate Professor at the Department of Higher Mathematics, Kharkiv National University of Radio Electronics, Kharkiv, Ukraine.

ABSTRACT

Context. The relevance of the topic lies in the need to improve radio direction finders to increase accuracy, resistance to interference and adaptation to changing operating conditions. Modern scientific achievements require the development of methods for statistical synthesis and analysis of stochastic signal processing in multi-antenna systems, which will allow to take into account the uncertainty of real conditions. It is important to expand the capabilities of such systems for use in radar, radio navigation, communications and other industries. This will facilitate the creation of new approaches for direction finding of unknown signal sources in complex operating scenarios.

Objective. The study is based on improving the measurements of the angular position accuracy of radio sources of stochastic signals.

Method. The approach is based on the statistical theory of optimization of radio remote sensing and radar systems. Signal and noise models are constructed for stochastic signal sources, and the likelihood functional in the spectral domain is formulated, taking into account the structure of inverse correlation matrices. The Cramer-Rao inequality is used to determine the limiting errors of estimation of the angular position of the radio source.

Results. For the first time the approach to statistical optimization of the structure of multi-antenna radio systems for direction finding of stochastic radiation sources is theoretically justified, allowing to take into account the spatial orientation, antenna array geometry and radiation pattern. An optimal method of processing the observation equations for estimating the angular position of stochastic signal sources is constructed. A generalized structure of a single-antenna direction finder containing a matched filter, a decoherence filter and a digital calculator is proposed. It is proved that the use of decorrelating processing allows to increase the estimation accuracy by increasing the number of independent signal samples. Analytical expressions for estimation and limiting errors, which take into account the spectrum width and directional pattern parameters, are obtained.

Conclusions. This paper presents the latest theoretical foundations for the synthesis of radio direction finders of arbitrary configuration, which take into account the variety of radiation pattern shapes, spatial location and orientation of direction finders. The developed models of signals and noise using the maximum likelihood function criterion for the first time allow solving optimisation problems of synthesis with consideration the physical content consideration of correlation matrices. The obtained results are confirmed by solving the problem of measuring the radiation source angular position, which proves the proposed approaches effectiveness.

KEYWORDS: multi-antenna direction finders, statistical optimization, optimal signal processing, simulation modeling.

REFERENCES

1. Hertl I., Strýček M., Poláček V., Brno V. Monopulse direction finder for harmonic radar system, *Proceedings of the 18th International Conference on Microwaves, Radar and Wireless Communications*. Vilnius, Lithuania, 2010, pp. 1–3.
2. Pavlikov V., Volosyuk V., Zhyla S., Van H. N., Van K. N. A new method of multi-frequency active aperture synthesis for imaging of SAR blind zone under aerospace vehicle, *Proceedings of the 14th International Conference The Experience of Designing and Application of CAD Systems in Microelectronics (CADSM)*, 2017, pp. 118–120. DOI: 10.1109/CADSM.2017.7916099.
3. Zhang C., Huang H., Liao B. Direction Finding in MIMO Radar With Unknown Mutual Coupling, *IEEE Access*, 2017, Vol. 5, pp. 4439–4447. DOI: 10.1109/ACCESS.2017.2684465.
4. Pavlikov V. V., Zhyla S. S., Odokienko O. V. Structural optimization of Dicke-type radiometer, *Proceedings of the 2nd International Young Scientists Forum on Applied Physics and Engineering (YSF)*, 2016, pp. 171–174. DOI: 10.1109/YSF.2016.7753829.
5. Pavlikov V. V., Zhyla S. S., Odokienko A. V., Antonov M. O. Radiometer with signal energies ratio, *Proceedings of the IEEE Radar Methods and Systems Workshop (RMSW)*, 2016, pp. 99–102. DOI: 10.1109/RMSW.2016.7778560.
6. Cwalina K., Rajchowski P., Sadowski J. Wideband Radio Direction Finder Implemented in Software Defined Radio Technology, *Applied Mechanics and Materials*, 2016, Vol. 817, P. 348. DOI: 10.4028/www.scientific.net/amm.817.348.
7. Ziolkowski C., Kelner J. M. Radio Bearing of Sources with Directional Antennas in Urban Environment, *International Journal of Microwave and Wireless Technologies*, 2018, Vol. 10, № 7, pp. 759–771. DOI: 10.1017/S1759078718000648.
8. Gulko V. L., Mescheryakov A. A. Polarization Amplitude-Phase Direction Finding Methods in Two-Canal UHF Radio Beacon Navigation Systems, *arXiv*, 2020. DOI: 10.48550/arXiv.2012.12728.
9. Direction Finding, *Wireless Communications Systems*. John Wiley & Sons, 2020, pp. 301–323. DOI: 10.1002/9781119419204.ch8.
10. Cook H. A., Kahn M. T. E., Balyan V. Radio Direction-Finding Techniques for an Unmanned Aerial Vehicle, *Micro-Electronics and Telecommunication Engineering. Lecture Notes in Networks and Systems*. Springer, Singapore, 2020, Vol. 106, P. 1. DOI: 10.1007/978-981-15-2329-8_1.

11. Pavlikov V., Volosyuk V., Zhyla S., Van H. N., Van K. N. Active Aperture Synthesis Radar for High Spatial Resolution Imaging, *Proceedings of the 9th International Conference on Ultrawideband and Ultrashort Impulse Signals (UWBUSIS)*, 2018, pp. 252–255. DOI: 10.1109/UWBUSIS.2018.8520021.
12. Pavlikov V., Volosyuk V., Zhyla S., Van H. N., Van K. N. UWB Active Aperture Synthesis Radar: Operating Principle and Development of the Radar Block Diagram, *Proceedings of the IEEE Microwaves, Radar and Remote Sensing Symposium (MRRS)*, 2017, pp. 27–30. DOI: 10.1109/MRRS.2017.8075018.
13. Novak A. Radio direction finding in air traffic services, *Promet-Traffic-Traffico*, 2005, Vol. 17, № 5, pp. 273–276.
14. Ilčev D. S. New Aspects of Progress in the Modernization of the Maritime Radio Direction Finders (RDF), *Transactions on Maritime Science*, 2021, Vol. 10, № 01, pp. 68–83. DOI: 10.31861/sisiot2023.2.02006.
15. Pavlikov V., Zhyla S., Odokienko O. V. Optimal signal processing for radiometric imaging with multi-antenna & multi-band passive radars, *Proceedings of the International Conference on Antenna Theory and Techniques (ICATT)*, 2015, pp. 7–12. DOI: 10.1109/ICATT.2015.7136821.
16. Zhyla S., Tserne E., Volkov Y., et al. Statistical Synthesis and Analysis of Functionally Deterministic Signal Processing Techniques for Multi-Antenna Direction Finder Operation, *Computation*, 2024, Vol. 12, № 9, P. 170. DOI: 10.3390/computation12090170.
17. Volosyuk V., Zhyla S. Statistical Theory of Optimal Functionally Deterministic Signals Processing in Multichannel Aerospace Imaging Radar Systems, *Computation*, 2022, Vol. 10, № 12, P. 213. DOI: 10.3390/computation10120213.
18. Chen Z., Zhang Y. Cooperative Energy Detection Algorithm Based on Background Noise and Direction Finding Error, *AEU – International Journal of Electronics and Communications*, 2018, Vol. 95, pp. 326–341. DOI: 10.1016/j.aeue.2018.08.029.
19. Boulageorgos A. A., Chatzidiamentis N. D., Karagiannidis G. K. Energy Detection Spectrum Sensing Under RF Imperfections, *IEEE Transactions on Communications*, 2016, Vol. 64, № 7, pp. 2754–2766. DOI: 10.1109/TCOMM.2016.2561294.
20. Wu Q., Ding G., Wang J., Yao Y. D. Spatial-Temporal Opportunity Detection for Spectrum-Heterogeneous Cognitive Radio Networks: Two-Dimensional Sensing, *IEEE Transactions on Wireless Communications*, 2013, Vol. 12, № 2, pp. 516–526. DOI: 10.1109/TWC.2012.122212.111638.
21. Riabukha V., Semeniaka A., Katiushyn Y. et al. Selection of Number of Compensatory Channels and Placement of Receivers with Non-Identical Frequency Response at Phased-Array Radar under Gaussian Noise Jamming, *Radioelectronics and Communications Systems*, 2022, Vol. 65, № 1, pp. 1–10. DOI: 10.3103/S0735272722010010.
22. Riabukha V. P., Semeniaka A. V., Katiushyn Y. A. Analysis of Protection Efficiency of Phased-Array Radars Against the Gaussian Noise Jamming, *Telecommunications and Radio Engineering*, 2021, Vol. 80, № 2, P. 1. DOI: 10.1615/TelecomRadEng.2021036556.
23. Holland R. L. The Statistical Theory of Radio Direction Finding, *Technical Memorandum AccNr ADA104156*. US Defense Technical Information Center, 1981.
24. Voitko V., Ilnycki A. J., Iliashov O., Steiskal A., Marchenko A. Statistical Estimates of the Accuracy of Mono-Pulse Direction Finding of Sources of Radio Emission by Dual-Channel Phase Devices, *Visnyk NTUU KPI Seriia – Radiotekhnika Radioaparotobuduvannia*, 2020, № 81, pp. 30–37. DOI: 10.20535/RADAP.2020.81.30-37.
25. Fabirovskyy S., Storozh V., Solomko A. Research and Development of a High-Precision Direction Finder for Radio Emission Sources Detecting, *Information and Communication Technologies, Electronic Engineering*, 2023, Vol. 3, № 10, pp. 158–169. DOI: 10.23939/ictee2023.02.158.
26. Henault S., Antar Y. M., Rajan S., Inkol R., Wang S. Impact of Mutual Coupling on Wideband Adcock Direction Finders, *Proceedings of the Canadian Conference on Electrical and Computer Engineering (CCECE)*, 2008, pp. 001327–001332. DOI: 10.1109/CCECE.2008.4564755.
27. Purwanto K. N., Yahya A., Khamis N. H. H. et al. Development of Radio Direction Finder Using 6 Log Periodic Dipole Array Antennas, *Proceedings of the 5th International Conference on Information Technology, Computer, and Electrical Engineering (ICITACEE)*, 2018, pp. 157–160. DOI: 10.1109/ICITACEE.2018.8576965.
28. Polikarovskiykh O., Hula I. Implementing the Search Algorithm of the Correlation Interferometer Direction Finder Through the GNU Radio Software Platform, *Security of Information Systems and Internet of Things*, 2023, Vol. 1, № 2, P. 02006. DOI: 10.31861/sisiot2023.2.02006.
29. Lekhovyskiy D. I., Atamanskiy D. V., Riabukha V. P. et al. Combining Target Detection Against the Background of Jamming Signals and Jamming Signal DOA Estimation, *Proceedings of the International Conference on Antenna Theory and Techniques (ICATT)*, 2015, pp. 1–5. DOI: 10.1109/ICATT.2015.7136777.
30. Atamanskiy D. V. Noise Emissions Sources Direction-Finding in the Process of Their Background Air Threats Detection in Radars with Phased Antenna Array, *Radioelectronics and Communications Systems*, 2017, Vol. 60, № 7, pp. 303–311. DOI: 10.3103/S0735272717070032.
31. Atamanskiy D. V., Riabukha V. P., Kartashov V. M. et al. Spectral Estimation Methods for a Joint System of the Non-Noise-Like Targets Detection and the Noise Radiating Sources Localization, *Radio Electronics, Computer Science, Control*, 2022, № 1, P. 7. DOI: 10.15588/1607-3274-2022-1-1.
32. Zhukov A. O., Valyaev I. N., Kovalenko V. P. et al. Synthesis of an Optimal Algorithm for Processing Random Signals during Phase Direction Finding, *Journal of Physics: Conference Series*, 2020, Vol. 1679, № 3, P. 032021. DOI: 10.1088/1742-6596/1679/3/032021.
33. Volosyuk V., Pavlikov V., Zhyla S. Phenomenological Description of the Electromagnetic Field and Coherent Images in Radio Engineering and Optical Systems, *Proceedings of the International Conference on Mathematical Methods in Electromagnetic Theory (MMET)*, 2018, pp. 302–305. DOI: 10.1109/MMET.2018.8460321.
34. Volosyuk V. K., Kravchenko V. F. Statistical Theory of Radio Engineering Systems for Remote Sensing and Radar, *Fizmatlit*. Moscow, Russia, 1988.
35. Volosyuk V. K., Gulyaev Y. V., Kravchenko V. F., et al. Modern Methods for Optimal Spatio-Temporal Signal Processing in Active, Passive, and Combined Active-Passive Radio-Engineering Systems, *Journal of Communications Technology and Electronics*, 2014, Vol. 59, № 2, pp. 97–118. DOI: 10.1134/S1064226914020090.
36. Lie J. P., Blu T., See C. M. S. Single Antenna Power Measurements Based Direction Finding, *IEEE Transactions on Signal Processing*, 2010, Vol. 58, № 11, pp. 5682–5692. DOI: 10.1109/TSP.2010.2065227.

USE OF THE PARAMETERS OF THE LAW OF DISTRIBUTION OF THE MEASUREMENT ERRORS OF THE PULSE OXIMETER TO SELECT THE SENSOR SETTINGS

Vakaliuk T. A. – Doctor of Pedagogical Sciences, Professor, Head of the Department of Software Engineering, Zhytomyr Polytechnic State University, Zhytomyr, Ukraine.

Andreiev O. V. – PhD, Associate Professor, Associate Professor of the Department of Computer Technologies in Medicine and Telecommunications, Zhytomyr Polytechnic State University, Zhytomyr, Ukraine.

Nikitchuk T. M. – PhD, Associate Professor, Dean of the Faculty of Information and Computer Technologies Zhytomyr Polytechnic State University, Zhytomyr, Ukraine.

Korenivska O. L. – PhD, Associate Professor, Associate Professor of the Department of Computer Technologies in Medicine and Telecommunications, Zhytomyr Polytechnic State University, Zhytomyr, Ukraine.

Nikitchuk S. M. – Senior Lecturer, Department of National Security, Public Management and Administration, Zhytomyr Polytechnic State University, Zhytomyr, Ukraine.

ABSTRACT

Context is due to the need to develop a methodology for optimising the parameters of pulse oximeter sensor settings based on the practical determination of the accuracy of heart rate and blood oxygen saturation measurements, which, unlike existing ones, do not require analysis of the amplitude of the LED current.

A pulse oximeter is one of the sensors that monitor the patient's vital signs, heart rate and blood oxygen saturation in particular. These indicators are determined based on the analysis of the values of the variable and constant components of the current of the red and infrared LEDs of the pulse oximeter sensor. Therefore, the accuracy of determining vital signs depends on the correct choice of the brightness and duration of the LEDs' radiation. It is possible to select the current and duration of the LEDs' radiation, as well as the ADC parameters of the sensor using software. In this case, the final conclusion regarding the correctness of the selected sensor settings is made based on the practical determination of the accuracy of heart rate and blood oxygen saturation measurements.

The object is to develop a methodology for assessing the correctness of the pulse oximeter sensor settings based on the analysis of the stationarity of errors in heart rate and blood oxygen saturation measurements.

Method. An experimental study of the accuracy of heart rate and blood oxygen saturation measurements by statistical analysis of measurement errors of the developed pulse oximeter model.

Results. The practical application of the proposed methodology for determining the optimal parameters of the pulse oximeter sensor settings was tested using the example of heart rate measurements.

Conclusion. A methodology has been developed to assess the correctness of the choice of sensor setting parameters based on analysing the stationarity of errors in measuring heart rate and oxygen saturation in the patient's blood. With the help of the developed methodology, the optimal settings parameters of the MAX30102 sensor of the pulse oximeter developed based on the ESP32 board were selected, which ensures the minimum error in measuring heart rate and blood oxygen saturation.

KEYWORDS: sensor, Internet of Things, measurement errors, pulse oximeter.

ABBREVIATIONS

ADC is an analog to digital converter;

API is an application-programming interface;

CSV is a comma-separated value;

IoT is an internet of things;

HR is a heart rate;

SpO2 is an oxygen saturation in the blood.

NOMENCLATURE

$\varepsilon(t)$ is a random process of the measurement errors;

$b(t_i)$ is a random process of the estimate errors;

σ_ε^2 is a dispersion of the measurement errors;

m_ε is a mathematical expectation of measurement errors;

m_b is a mathematical expectation of the full sample of estimation errors;

m_1 is a mathematical expectation of the first part of the estimation errors sample;

m_2 is a mathematical expectation of the second part of the estimation errors sample;

S_b^2 is a dispersion of the full sample of estimation errors;

S_1^2 is a dispersion of the first part of the estimation errors sample;

S_2^2 is a dispersion of the second part of the estimation errors sample;

t_b is a Student's t -distribution statistic;

F is a F -distribution statistic;

k_1 is a freedom's degree of the first part of the estimation errors sample;

k_2 is a freedom's degree of the second part of the estimation errors sample;

$t(\alpha, n-1)$ is a quantile of the Student's t -distribution;
 $F(\alpha, k_1, k_2)$ is a quantile of the F -distribution
 n is a measurements number;
 l is a subsample size;
 $y^*(t_i)$ is a discrete measurements;
 H_0 is a statistical hypothesis;
 Δ is a confidence interval;
 α is a level of significance.

INTRODUCTION

Maintaining sufficient oxygen levels in the blood is vital for human health. Hypoxia is a condition where the oxygen concentration in the blood falls below the normal level, and it can be a direct threat to life. Low blood oxygen concentration primarily affects the brain, heart, and kidneys. Blood oxygen saturation (SpO₂) is measured by determining arterial blood's gas structure and acid-base state. However, these studies require blood sampling from an artery, qualified personnel, and specialised equipment and are typically conducted in a medical facility. Instead, using a non-invasive method, a pulse oximeter determines the heart rate and the amount of oxygen in the blood through the skin of the finger or ear. The pulse oximeter functions by emitting light and two distinct wavelengths, generally in the red and infrared spectrums, to the skin capillaries and measures how much light is reflected, which, in turn, depends on the structure of arterial blood and the amount of oxidised haemoglobin [1]. The results of the pulse oximeter measurement show the oxygen saturation in the blood, or SpO₂ level, as a percentage. The standard value for blood saturation in a healthy person is over 96%. A 3–4% decrease indicates a serious illness [2].

Pulse oximeters are common devices in healthcare settings, and with the advancement of IoT technologies, they have become essential tools for remote patient monitoring. The accuracy of these devices is crucial for correct diagnosis and treatment decisions, especially in critical care scenarios. This accuracy depends primarily on the correct calibration and settings of the sensor parameters.

The object of research is to remotely monitor the results of blood saturation and heart rate measurements of patients.

The subject of research is the impact of software-defined parameters of the pulse oximeter sensor on the accuracy of measuring both heart rate and blood oxygen saturation.

The purpose of the work is the development of a methodology for assessing the correctness of the selection of the settings parameters of the MAX30102 sensor based on the analysis of the stationarity of measurement errors of heart rate and oxygen saturation in the patient's blood.

1 PROBLEM STATEMENT

Various pulse oximeters are now available, including those using IoT technologies to monitor patients' heart

rate and blood oxygen saturation remotely. The microcontroller uses built-in libraries to read and process digital readings from the sensor's ADC output and to improve measurement accuracy; it is possible to programmatically change the current and duration of LED emission, as well as control the ADC parameters. The sensor settings are optimised to improve measurement accuracy by selecting the amplitude of the LED currents. The conclusion about the correctness of the selected settings parameters is made based on the practical determination of the accuracy of heart rate measurements and blood oxygen saturation. Therefore, it is advisable to develop a methodology for assessing the correctness of the choice of sensor setting parameters based on analysing the stationarity of errors in measuring heart rate and oxygen saturation in the patient's blood.

In the practical use of pulse oximeter sensors, it is necessary to determine the current amplitude of each LED, which will correspond to the optimal setting. The user can then analyse the correctness of the selected settings by determining the accuracy of HR and SpO₂ (%) measurements using the SpO₂ algorithm [3].

Suppose that the measurement of the parameter $x(t)$ is described by a random process

$$y(t) = x(t) + \varepsilon(t).$$

In this case, the random process $\varepsilon(t)$ satisfies the conditions of stationarity with $m_\varepsilon = 0$ and dispersion σ_ε^2 , which is a priori unknown.

At the same time, the dispersion σ_ε^2 of HR and SpO₂ (%) measurements depends on the brightness and duration of the LEDs, as well as the ADC parameters of the pulse oximeter sensor. Let us assume that at discrete points in time $t_i, i = 1, \dots, n$ a sample of measurements is obtained $y^*(t_i)$ for different values of the sensor settings parameters. It is necessary to develop a method for optimizing the sensor settings parameters based on the obtained sample of measurements $y^*(t_i)$, which is based on the stationarity of HR and SpO₂ (%) measurement errors and does not require analysis of the current amplitude of the sensor LEDs.

2 REVIEW OF THE LITERATURE

The pulse oximeter is one of the sensors that monitor vital health indicators; it occupies a place at the level of 'things' in the overall architecture of the healthcare system based on the use of cloud IoT technologies [4,5,6]. Most pulse oximeter models use MAX30100, MAX30101 and MAX30102 (MAX3010x) sensors [7, 8]. The development of a portable pulse oximeter that uses the MAX30102 sensor, connected to the Arduino platform via the I2C interface, where measurements are processed and the results are displayed on an external TFT screen, was considered in [9]. Modern mobile technologies and the Internet of Things (IoT) allow for remote monitoring

of patients' blood saturation measurements and displaying the results on remote services. This makes it possible to monitor the blood saturation of patients with mild and average disease severity at home, significantly reducing the demand for hospital beds. Secondly, doctors will not have direct contact with patients [10]. For example, in [10,11], it is proposed to create an IoT pulse oximeter based on the ESP32 platform. This platform includes a built-in WiFi module, enabling the transmission of measurement data from the MAX30100 sensor via the Internet to the ThingSpeak cloud service and displaying them in the graphical form dependencies for a sufficiently long time interval. A personal doctor periodically monitors these data by accessing the ThingSpeak server using an API key and channel number identifier with visualisation on a smartphone, tablet or personal computer. A similar approach is described in [12] but is based on the ESP8266 microcontroller. The construction and development of an IoT-based pulse oximeter capable of wirelessly transmitting the obtained readings to the Blynk mobile application is discussed in [13]. It is recommended that an Arduino-based microcontroller be used for signal processing and that the ESP-01 WiFi module be used to ensure remote data transmission.

IoT devices remotely monitoring patients' blood saturation can be used at home. However, the works do not indicate how the patient will be informed about the need for measurements in conditions when the monitoring device is not constantly on the patient's finger. Therefore, in [14], a variant of building an IoT oximeter with feedback control using the Telegram messenger based on the ESP32 DEVKIT V1 board and the MAX30102 sensor was proposed. This board has a built-in WiFi module and an LED that can be switched on at the command of a pre-created Telegram bot when measurements are required. Much attention in the literature is paid to developing various variants of pulse oximeters, including those using IoT technologies, for remote monitoring of patients' heart rate and blood oxygen saturation. However, the impact of sensor parameters configured by software on the accuracy of measuring both heart rate and blood oxygen saturation remains a significant gap in the literature.

3 MATERIALS AND METHODS

Pulse oximeter sensors typically contain two LEDs that work in the red and infrared spectrums, a photodetector, an analog amplifier, and an analog-to-digital converter. The fundamental operation of these sensors relies on the differential absorption of light by oxygenated and deoxygenated hemoglobin at the two wavelengths.

To improve measurement accuracy, several parameters can be programmatically adjusted:

1. LED brightness (current amplitude).
2. LED pulse duration.
3. Sampling rate (ADC parameters).
4. Signal filtering and processing algorithms.

The optimization of these parameters involves finding a balance between measurement accuracy, power consumption, and signal-to-noise ratio. Higher LED bright-

ness generally provides better signal quality but increases power consumption [15]. Similarly, longer pulse durations and higher sampling rates improve signal quality at the cost of increased power consumption.

We propose a methodology based on statistical analysis of the stationarity of measurement errors to assess the correctness of the selected sensor settings. This approach provides an objective criterion for parameter optimization without requiring direct analysis of LED current amplitudes.

The MAX30102 sensor is also an integrated heart rate and blood oxygen saturation sensor. The acceptable values of the MAX3010x sensor parameters and general recommendations for their selection are listed in [3]. In particular, to improve the accuracy of blood saturation and heart rate measurements with the MAX3010x sensor, users should optimise the LED signal power, pulse duration, and sampling rate. It is noted that when measuring heart rate (HR), it is desirable to have the highest possible LED current, increasing the LEDs' brightness. However, to minimise the power supply consumption of the power supply, the current of the LEDs should be reduced. Regarding power consumption, the longer the pulse duration of the LEDs, the more power it consumes. Increasing the sampling rate also increases the power consumption of the power supply. The signal-to-noise ratio is also reduced by reducing the pulse duration and sampling rate. It is recommended that the user determines the amplitude of the current of each LED that will correspond to the optimal setting using the graphical interface of the MAX3010x set with data logging in CSV format. The user can then analyse the correctness of the selected settings. This analysis can be made based on the practical determination of the accuracy of HR and SpO2 (%) measurements using the SpO2 algorithm [3].

Therefore, to simplify the assessment of the correctness of the selected settings of the MAX30102 sensor, the following methodology is proposed based on the statistical analysis of the stationarity of HR and SpO2 measurement errors.

Suppose that at discrete points in time $t_i, i = 1, \dots, n$ sample of measurements $y^*(t_i)$ is obtained. If measurements $y^*(t_i)$ with a certain accuracy describe the real behavior of the parameter $x(t)$, so the statistical properties $\varepsilon(t)$ and $b(t_i) = b_i = y^*(t_i) - x(t_i)$ with increasing sample size will become increasingly. Then, the conclusion about the stationarity of the random process $\varepsilon(t)$, can be made on the basis of the analysis of the statistical properties of the process $b(t_i)$. If the dispersion of measurement errors σ_ε^2 is not known a priori, the simplest criteria for checking stationarity consists in sequentially testing the hypotheses $H_0: E\{b(t_i)\} = m_\varepsilon$ and $H_0: E\{b^2(t_i)\} = \text{const}$ [16]. For the normal law of distribution of measurement errors, under the actual hypothesis

$H_0 : E\{b(t_i)\} = m_b = m_\varepsilon$, statistic $t_b = \frac{m_b}{S_b} \sqrt{n-1}$ obeys

distribution of Student with $n-1$ degrees of freedom. If $t_b > t(\alpha, n-1)$, the hypothesis H_0 must be declined. The quantiles of $t(\alpha, n-1)$ for different values of n and significance level α are tabulated, and the values of m_b and S_b are calculated using the formulas:

$$m_1 = \frac{1}{n} \sum_{i=1}^n b_i; S_b^2 = \frac{1}{n-1} \sum_{i=1}^n (b_i - m_b)^2.$$

Testing the hypothesis $H_0 : E\{b(t_i)\} = m_\varepsilon$ can be achieved in the following sequence. The sample $b_i, i=1, \dots, n$ is divided into two approximately equal, non-overlapping groups, $b_1, \dots, b_l; b_{l+1}, \dots, b_n$. For each group, the following values are calculated

$$m_1 = \frac{1}{l} \sum_{i=1}^l b_i; m_2 = \frac{1}{n-l} \sum_{i=l+1}^n b_i;$$

$$S_1^2 = \frac{1}{l-1} \sum_{i=1}^l (b_i - m_1)^2;$$

$$S_2^2 = \frac{1}{n-l-1} \sum_{i=l+1}^n (b_i - m_2)^2.$$

Estimates the value of Fisher's statistic:

$$F = \max \left[\frac{S_1^2}{S_2^2}; \frac{S_2^2}{S_1^2} \right].$$

The degrees of freedom are determined:

$$k_1 = l-1, k_2 = n-l-1 \text{ with } S_1^2 > S_2^2;$$

$$k_1 = n-l-1, k_2 = l-1 \text{ with } S_1^2 < S_2^2.$$

For a defined k_1, k_2 and a fixed significance level α the critical value $F(\alpha, k_1, k_2)$ is determined. If $F < F(\alpha, k_1, k_2)$, the hypothesis $H_0 : E\{b^2(t_i)\} = \text{const}$ must be considered valid with confidence $1-\alpha$. Thus, a random process $\varepsilon(t)$ satisfies the conditions of stationarity with $m_\varepsilon = 0$ and dispersion σ_ε^2 , if both hypothesis $H_0 : E\{b(t_i)\} = m_\varepsilon$ and $H_0 : E\{b^2(t_i)\} = \text{const}$ are true.

The optimal settings parameters of the pulse oximeter sensor match the maximum accuracy of measurements of heart rate HR and blood oxygen saturation SpO2 [3]. Therefore, for a stationary random process of measurement errors, the optimal setting parameters can be considered those at which $m_b \rightarrow m_\varepsilon = 0$, a $S_b \rightarrow \min$. In this case, the optimal setting corresponds to the minimum value of the confidence interval

$$\Delta = m_b \pm t(\alpha, n-1) \frac{S_b}{\sqrt{n}}.$$

4 EXPERIMENTS

To make the research, we will use a previously developed layout of a pulse oximeter based on the ESP32 DEVKIT V1 board and the MAX30102 sensor, shown in Fig. 1 [14].

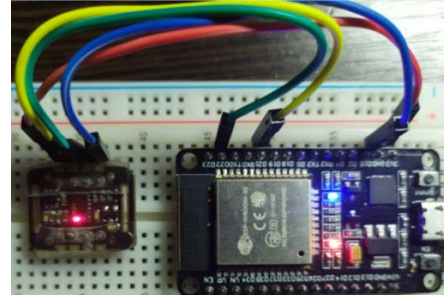


Figure 1 – Layout of IoT pulse oximeter on ESP32

The results of HR and SpO2 measurements were output to the serial port, and further statistical processing of the measurements and the construction of graphical dependencies were carried out in the Excel software environment.

In practice, the programming of pulse oximeter controllers, the typical examples of programme codes are used, which are embedded in the libraries of software systems. For instance, the Arduino IDE software environment offers the use of the program code “SparkFun MAX3010x Pulse and Proximity Sensor Library_Example_SPO2.ino”, which uses the library “spo2_algorithm.h” to determine HR and SpO2. The program code has the following options, which can be modified by the user:

```
ledBrightness = 60; //Options: 0=off to 255=50mA;
sampleAverage = 4; //Options: 1, 2, 4, 8, 16, 32;
ledMode = 2; //Options: 1-Red only, 2- (Red + IR);
sampleRate = 200; //Options: 50, 100, 200, 400, 800, 1000, 1600, 3200;
pulseWidth = 411; //Options: 69, 118, 215, 411;
adcRange = 4096; //Options: 2048, 4096, 8192.
```

The acceptable values of the pulse duration and frequency of the sensor sampling when measuring SpO2 and HR are shown in Fig. 2 [3]. As shown in Fig. 1, when simultaneously measuring HR and SpO2, the pulse duration emitted by the LEDs can vary from 69 to 411 μs , and the frequency of sampling from 50 to 400 Hz.

Allowed setting for SpO2 configuration

SAMPLES PER SECOND	PULSE WIDTH (μs)			
	69	118	215	411
50	o	o	o	o
100	o	o	o	o
200	o	o	o	o
400	o	o	o	o
800	o	o	o	o
1000	o	o		
1600	o			
3200				

Allowed setting for HR configuration

SAMPLES PER SECOND	PULSE WIDTH (μs)			
	69	118	215	411
50	o	o	o	o
100	o	o	o	o
200	o	o	o	o
400	o	o	o	o
800	o	o	o	o
1000	o	o	o	o
1600	o	o	o	o
3200	o			

Figure 2 – Valid settings for the MAX3010x

For instance, Figures 3 and 4 illustrate HR measurements for varying pulse durations at an LED brightness level of 60 and the amount of ADC quantisation levels of 8192 and 4096, respectively.

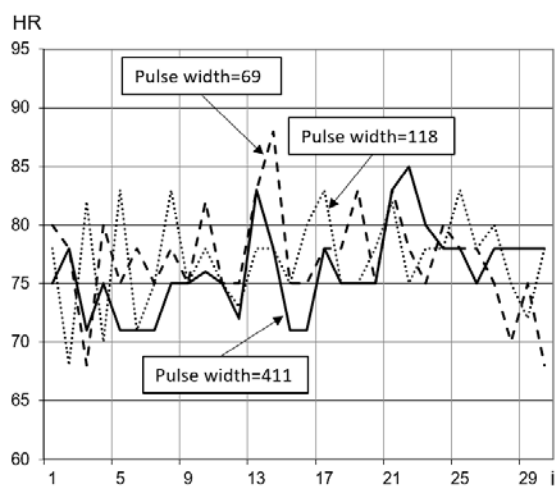


Figure 3 – Measurements HR for ledBrightness = 60, adcRange = 8192

As demonstrated in Figures 5 and 6, measurements of heart rate (HR) were taken at the same sensor settings, but for a level of LED brightness of 30.

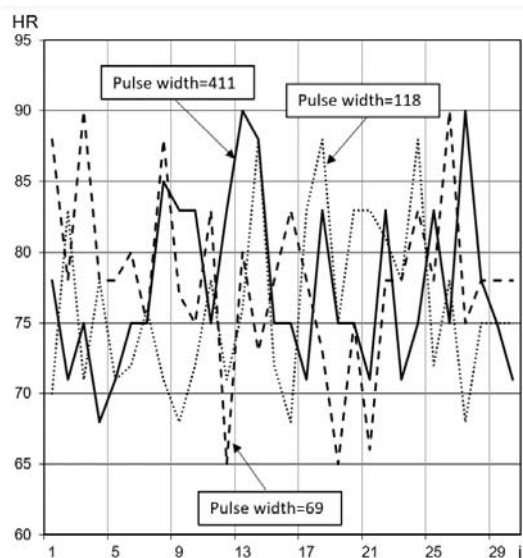


Figure 4 – Measurements HR for ledBrightness = 60, adcRange = 4096

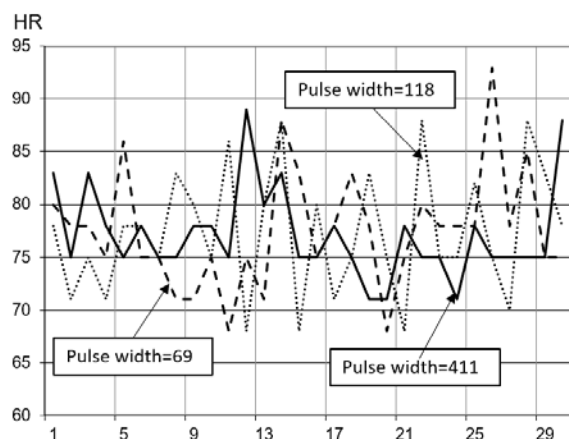


Figure 5 – Measurements HR for ledBrightness = 30, adcRange = 8192

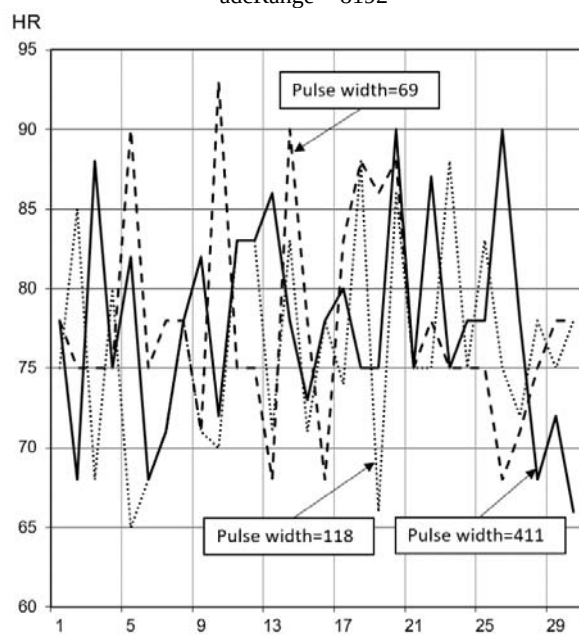


Figure 6 – Measurements HR for ledBrightness = 30, adcRange = 4096

Upon analysis of the presented data, it is challenging to draw any visual conclusions regarding which settings of the MAX30102 sensor ensure the minimum error in heart rate measurement.

5 RESULTS

The feasibility of the proposed methodology for determining the optimal parameters of the pulse oximeter sensor settings was tested using heart rate measurements. The stationarity of the random process of measurement errors was analysed using a sample of $n=30$ counts, which were obtained with a frequency of sampling of 200 Hz. The actual value of the heart rate during the measurements was 76 beats per minute.

As illustrated in Table 1, the values employed in the methodology for analysing the stationarity of a random process of measurement errors have been calculated for different parameters of the sensor settings when using an ADC with 8192 quantisation levels.

Table 1 – Checking the correctness of the sensor settings (adcRange = 8192)

Value	Pulse width (ledBrightness = 60)			Pulse width (ledBrightness = 30)		
	69	118	411	69	118	411
m_b	1.13	1.07	0.07	1.53	1.33	1.17
S_b	4.33	3.98	3.74	5.62	5.98	4.36
t_b	1.41	1.44	0.1	1.47	1.2	1.44
S_1^2	21.52	20.55	11.98	32.54	36.07	17.67
S_2^2	16.69	10.43	11.52	31.12	37.64	16.95
F	1.29	1.97	1.04	1.05	1.04	1.04
Δ	3.24	2.98	2.8	4.2	4.48	3.27

The quantile of the Student's distribution with $(n-1)$ degrees of freedom and significance level of $\alpha = 0.05$ is $t(\alpha, n-1) = 2.05$. The critical value of Fisher's statistic $F(\alpha, k_1, k_2)$ for $k_1 = k_2 = 14$ and the significance level $\alpha = 0.05$ is 2.48 [17]. As illustrated in Table 1, the calculated values of the Student's distribution statistic $t_b < 2.05$ and the Fisher's statistic $F < 2.48$ were obtained for all sensor settings. Thus, the random process of measurement errors meets the conditions of stationarity. The minimum value of m_b and S_b for the LED brightness level of 60, is achieved at a pulse duration of 411 μs . When the brightness level of the LEDs is reduced by two times, the minimum values of m_b and S_b correspond to a pulse duration of 411 μs . The confidence interval Δ also reaches its minimum value at a pulse duration of 411 μs for both LED brightness levels. Therefore, the value of the LED pulse duration of 411 μs can be considered optimal for two levels of LED brightness.

As illustrated in Table 2, the analysis of the stationarity of the random process of measurement errors for different sensor settings when using an ADC with 4096 quantisation levels.

Table 2 – Checking the correctness of the sensor settings (adcRange = 4096)

Value	Pulse width (ledBrightness = 60)			Pulse width (ledBrightness = 30)		
	69	118	411	69	118	411
m_b	2.07	0.23	1.53	1.83	0.27	1.67
S_b	6.27	6.06	6.08	6.69	6.45	6.57
t_b	1.78	0.21	1.36	1.48	0.22	1.37
S_1^2	40.92	29.55	43.52	51.07	43.74	40.24
S_2^2	38.35	39.71	31.78	41.26	37.78	49.09
F	1.07	1.34	1.37	1.24	1.16	1.22
Δ	4.69	4.53	4.55	5.01	4.83	4.92

As illustrated in Table 2, the calculated values of distributions statistics of the Student and the Fisher satisfy the conditions of stationarity of measurement errors of all sensor settings. The minimum values of m_b, S_b , та Δ are achieved at a LED pulse duration of 118 μs for both LED brightness levels. Consequently, the value of the LED pulse width of 118 μs can be regarded as optimal for two levels of LED brightness. At the same time, the values of the standard deviations and confidence intervals of the measurement errors for pulse durations of 118 and 411 μs are insignificantly different.

6 DISCUSSION

As a result, the developed methodology simplifies the verification of the correctness of the selected sensor settings and, unlike the existing ones, does not require the analysis of the LED current amplitude. Based on the statistical analysis of the stationarity of the HR measurement errors, the following conclusions can be made. When utilising an ADC with 8192 quantisation levels, the LED pulse duration value of 411 μs can be considered optimal for two LED brightness levels. When the number of quantisation levels of the ADC is reduced by two times, the optimum value for two LED brightness levels was found to be 118 μs . However, the calculations demonstrated that for a pulse duration of 411 μs , the accuracy of HR measurements decreases slightly. Reducing the quantisation levels of the ADC and the brightness of the LEDs leads to an increase in the standard deviation and confidence intervals of measurement errors. At a pulse duration of 69 μs , the maximum displacement of the average value of m_b is marked. The optimum sampling rate is 200 Hz. Reducing or increasing the sampling rate led to the appearance of incorrect heart rate measurements in the sample.

Meanwhile, the value of the sampling rate has almost no effect on the results of SpO2 measurements. Fig. 7 shows SpO2 measurements for different sampling rates with the number of ADC quantisation levels 4096.

As demonstrated in Figure 7, even for a minimum pulse duration of 69 μs and a brightness level of 30 LEDs, all the achieved SpO2 measurements have a value of at least 96%, representing the normal blood saturation of a healthy person. It is noteworthy that the minimum range of SpO2 measurements corresponds to the sampling frequency of 200 Hz.

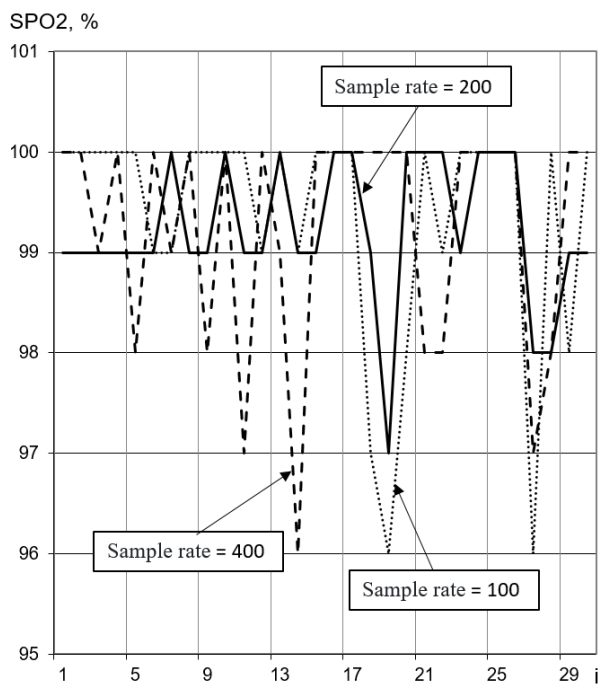


Figure 7 – SpO2 measurements for ledBrightness =30, pulseWidth =69 мкс, adcRange =4096

CONCLUSION

The scientific problem concerning the methodology for optimising the parameters of the pulse oximeter sensor settings, which has been further developed based on the practical determination of the accuracy of measurements of heart rate and blood oxygen saturation. Unlike the existing ones, this methodology does not require analysis of the current amplitude of the sensor LEDs.

Practical significance. The developed methodology has significantly simplified the selection of optimal settings for the pulse oximeter sensor, ensuring minimal error in heart rate measurement. Research has demonstrated that higher accuracy is provided at more enormous LED brightness when measuring heart rate. Conversely, reducing the quantisation levels of the ADC and the brightness of the LEDs has been demonstrated to increase measurement error.

The value of the ADC sampling rate has minimal effect on the results of blood oxygen saturation measurements, and it is advisable to select a sampling rate of 200 Hz for measuring heart rate.

Directions for further research. It should be noted that the duration of the pulse and the LED brightness affect the amount of energy consumed by the power supply, so it is worthwhile to do some additional research on the selected sensor settings when designing pulse oximeters with an autonomous power supply.

ACKNOWLEDGEMENTS

We thank the management of Zhytomyr Polytechnic State University for the opportunity to conduct scientific research.

REFERENCES

1. Nitzan M., Romem A., Koppel R. Pulse oximetry: fundamentals and technology update, *Medical Devices: Evidence and Research*, 2014, P. 231. DOI: 10.2147/mder.s47319
2. Putu Anna Andika I., Rahmawati Triana, Ridha Mak'ruf M. Pulse oximeter portable, *Journal of Electronics, Electromedical Engineering, and Medical Informatics*, 2019, Vol. 1, № 1, pp. 28–32. DOI: 10.35882/jeeemi.v1i1.6
3. Recommended configurations and operating profiles for MAX30101/MAX30102 EV Kits, *Maxim Integrated*. [Electronic resource]. Access mode: <https://www.analog.com/media/en/technical-documentation/user-guides/max3010x-ev-kits-recommended-configurations-and-operating-profiles.pdf>
4. Nikitchuk T. M., Nakonechnyi O. Yu., Vakulka M. Yu. [et al.] Mathematical model of the base unit of the biotechnical system as a type of edge devices, *Journal of Physics: Conference Series*, 2022, Vol. 2288, № 1, P. 012004. DOI: 10.1088/1742-6596/2288/1/012004
5. Khan S. N. Health Monitoring Using IOT, *International Journal for Research in Applied Science and Engineering Technology*, 2025, Vol. 13, № 2, pp. 285–303. DOI: 10.22214/ijraset.2025.66845
6. Adiputra R. R., Hadiyoso S., Hariyani Y. S. Internet of Things: Low Cost and Wearable SpO2 Device for Health Monitoring, *International Journal of Electrical and Computer Engineering*, 2018, Vol. 8, № 2, pp. 939. DOI: 10.11591/ijece.v8i2.pp939-945
7. Aruna S., Sumithra M. G., Suganthi D. Vijaya [et al.] Iot-based smart saline level and health monitoring system, *Journal of Nonlinear Analysis and Optimization*, 2024, Vol. 15, № 01, pp. 1741–1750. DOI: 10.36893/jnao.2024.v15i01.1741-1750
8. Sarkar S., Ghosh A., Chakraborty M., Mondal A. Design, hardware implementation of a domestic pulse oximeter using IoT for covid-19 patient, *International Journal of Microsystems and Iot*, 2024, Vol. 2(1), pp. 469–475. DOI: 10.5281/zenodo.10629635
9. O'Carroll O., MacCann R., O'Reilly A. et al. Remote monitoring of oxygen saturation in individuals with covid-19 pneumonia, *European Respiratory Journal*, 2020, Vol. 56, № 2, P. 2001492. DOI: 10.1183/13993003.01492-2020
10. Rao P. R., Saleem S. K., Parvatham P. K. et al. Covid-19 patient health management system using IoT, *Next Generation of Internet of Things. Proceedings of ICNGIoT 2021*, 2021, pp. 635–646. DOI: 10.1007/978-981-16-0666-3_52
11. Miron-Alexe V. IoT pulse oximetry status monitoring for home quarantined covid-19 patients, *Journal of Science and Arts*, 2020, Vol. 20, № 4, pp. 1029–1036. DOI: 10.46939/j.sci.arts-20.4-c04
12. Hema L., Priya R. M., Indumathi R. Design and development of IoT based pulse oximeter, *International Journal of Pure and Applied Mathematics*, 2018, Vol. 119, № 16, pp. 1863–1868. URL: <https://acadpubl.eu/hub/2018-119-16/1/183.pdf>
13. Sidhartha G. IoT Based Pulse Oximeter, *International Journal for Research in Applied Science and Engineering Technology*, 2021, Vol. 9, № 8, pp. 2946–2951. DOI: 10.22214/ijraset.2021.37782
14. Vakaliuk T. A., Andreiev O. V., Dubyna O. F. et al. Use of wireless technologies in IoT projects, *Journal of Edge Computing*, 2024, Vol. 3(2), pp. 147–167. DOI: 10.55056/jec.750
15. Contardi U. A., Morikawa M., Brunelli B., Thomaz D. V. MAX30102 Photometric BiosensorCoupled to ESP32-

- Webserver Capabilities for Continuous Point of Care Oxygen Saturation and Heart rate Monitoring, *Eng. Proc.*, 2022, Vol. 16(1), P. 9. DOI: 10.3390/IECB2022-11114
16. Andreev O. V., Mukomel O. M., Topolnicki P. P., Friz S. P. A technique of check of the developed mathematical models on the data of measurements, *Bulletin of Zhytomyr State Technological University. Series "Technical Sciences"*, 2017, Vol. 1(40), pp. 65–69. DOI: 10.26642/tn-2007-1(40)-65-69
17. Korn T. M., Korn G. A. *Mathematical Handbook for Scientists and Engineers: Definitions, Theorems, and Formulas for Reference and Review*. Dover Publications, 2013, 1152 p. (Dover Civil and Mechanical Engineering)
- Received 25.05.2025.
Accepted 14.08.2025.

УДК 616.71:004.891

ВИКОРИСТАННЯ ПАРАМЕТРІВ ЗАКОНУ РОЗПОДІЛУ ПОХИБОК ВИМІРІВ ПУЛЬСОКСИМЕТРУ ДЛЯ ВИБОРУ НАЛАШТУВАНЬ ДАТЧИКА

Вакалюк Т. А. – д-р пед. наук, професор, завідувач кафедри інженерії програмного забезпечення, Державний університет «Житомирська політехніка», Житомир, Україна.

Андрєєв О. В. – канд. техн. наук, доцент, доцент кафедри комп'ютерних технологій у медицині та телекомунікаціях, Державний університет «Житомирська політехніка», Житомир, Україна.

Нікітчук Т. М. – канд. техн. наук, доцент, декан факультету інформаційно-комп'ютерних технологій, Державний університет «Житомирська політехніка», Житомир, Україна.

Корєнівська О. Л. – канд. техн. наук, доцент, доцент кафедри комп'ютерних технологій у медицині та телекомунікаціях, Державний університет «Житомирська політехніка», Житомир, Україна.

Нікітчук С. М. – старший викладач кафедри національної безпеки, публічного управління та адміністрування, Державний університет «Житомирська політехніка», Житомир, Україна.

АНОТАЦІЯ

Актуальність обумовлена необхідністю розробки методики оптимізації параметрів налаштувань датчика пульсоксиметру на основі практичного визначення точності вимірів частоти серцевих скорочень і сатурації кисню в крові, яка на відміну від існуючих, не вимагає аналізу амплітуди струму світлодіодів.

Пульсоксиметр, як один із сенсорів, що здійснюють моніторинг життєво важливих показників здоров'я пацієнта, вимірює частоту серцевих скорочень і сатурацію кисню в крові. Ці показники визначаються на основі аналізу значень змінної та сталої складової струму червоного та інфрачервоного світлодіодів датчика пульсоксиметру. Тому від правильного вибору яскравості і тривалості випромінювання світлодіодів залежить точність визначення життєво важливих показників здоров'я. Існує можливість програмного вибору струму і тривалості випромінювання світлодіодів, а також параметрів АЦП. При цьому, остаточний висновок щодо правильності обраних параметрів налаштувань датчика робиться на основі практичного визначення точності вимірів частоти серцевих скорочень і сатурації кисню в крові.

Метою роботи є розробка методики оцінки правильності вибору параметрів налаштувань датчика пульсоксиметру, що базується на аналізі стаціонарності похибок вимірів частоти серцевих скорочень та сатурації кисню в крові.

Метод. Експериментальне дослідження точності вимірів частоти серцевих скорочень і сатурації кисню в крові шляхом статистичного аналізу похибок вимірювань створеного макету пальсоксиметра.

Результати. Проведена перевірка практичного застосування запропонованої методики для визначення оптимальних параметрів налаштувань датчика пульсоксиметру на прикладі вимірювань частоти серцевих скорочень.

Висновки. Розроблена методика оцінки правильності вибору параметрів налаштувань датчика, що базується на аналізі стаціонарності похибок вимірів частоти серцевих скорочень та сатурації кисню в крові пацієнта. За допомогою розробленої методики обрані оптимальні параметри налаштувань макету пульсоксиметру на базі плати ESP32 та датчика MAX30102, які забезпечують мінімальну похибку виміру частоти серцевих скорочень та сатурації кисню в крові.

КЛЮЧОВІ СЛОВА: датчик, інтернет речей, похибки вимірів, пульсоксиметр.

ЛІТЕРАТУРА

1. Nitzan M. Pulse oximetry: fundamentals and technology update / M. Nitzan, A. Romem, R. Koppel // *Medical Devices: Evidence and Research*. – 2014. – P. 231–239. DOI: <https://doi.org/10.2147/mder.s47319>.
2. I Putu Anna Andika. Pulse oximeter portable / I Putu Anna Andika, Triana Rahmawati, M. Ridha Mak'ruf // *Journal of Electronics, Electromedical Engineering, and Medical Informatics*. – 2019. – Vol. 1, № 1. – P. 28–32. DOI: <https://doi.org/10.35882/jeeemi.v1i1.6>.
3. Recommended configurations and operating profiles for MAX30101/MAX30102 EV Kits // Maxim Integrated. – [Electronic resource]. – Access mode: <https://www.analog.com/media/en/technical-documentation/user-guides/max3010x-ev-kits-recommended-configurations-and-operating-profiles.pdf>.
4. Mathematical model of the base unit of the biotechnical system as a type of edge devices / [T. M. Nikitchuk, O. Yu. Nakonechnyi, M. Yu. Vakulka et al.] // *Journal of Physics: Conference Series*. – 2022. – Vol. 2288, № 1. – P. 012004. DOI: <https://doi.org/10.1088/1742-6596/2288/1/012004>.
5. Khan S. N. Health Monitoring Using IOT / S. N. Khan // *International Journal for Research in Applied Science and Engineering Technology*. – 2025. – Vol. 13, № 2. – P. 285–303. DOI: <https://doi.org/10.22214/ijraset.2025.66845>.
6. Adiputra R. R. Internet of Things: Low Cost and Wearable SpO2 Device for Health Monitoring / R. R. Adiputra, S. Hadiyoso, Y. S. Hariyani // *International Journal of Electrical and Computer Engineering*. – 2018. – Vol. 8, № 2. – P. 939–945. DOI: <https://doi.org/10.11591/ijece.v8i2.pp939-945>.
7. Iot-based smart saline level and health monitoring system / [S. Aruna, M. G. Sumithra, D. Vijaya Suganthi et al.] // *Journal of Nonlinear Analysis and Optimization*. – 2024. – Vol. 15, № 01. – P. 1741–1750. DOI: <https://doi.org/10.36893/jnao.2024.v15i01.1741-1750>.
8. Design, hardware implementation of a domestic pulse oximeter using IoT for covid – 19 patient / [S. Sarkar, A. Ghosh, M. Chakraborty, A. Mondal] // *International Journal of Microsystems and Iot*. – 2024. – Vol. 2, № 1. – P. 469–475. DOI: <https://doi.org/10.5281/zenodo.10629635>.
9. Remote monitoring of oxygen saturation in individuals with covid-19 pneumonia / [O. O'Carroll, R. MacCann, A. O'Reilly et al.] // *European Respiratory Journal*. – 2020. – Vol. 56, № 2. – P. 2001492. DOI: <https://doi.org/10.1183/13993003.01492-2020>.
10. Covid-19 patient health management system using IoT / [P. R. Rao, S. K. Saleem, P. K. Parvatham et al.] // *Next Generation of Internet of Things. Proceedings of ICNGIoT 2021*. – 2021. – P. 635–646. DOI: https://doi.org/10.1007/978-981-16-0666-3_52.
11. Miron-Alexe V. IoT pulse oximetry status monitoring for home quarantined covid-19 patients / V. Miron-Alexe // *Journal of Science and Arts*. – 2020. – Vol. 20, № 4. – P. 1029–1036. DOI: <https://doi.org/10.46939/j.sci.arts-20.4-c04>.
12. Hema L. Design and development of IoT based pulse oximeter / L. Hema, R. M. Priya, R. Indumathi // *International Journal of Pure and Applied Mathematics*. – 2018. – Vol. 119, № 16. – P. 1863–1868. DOI: <https://acadpubl.eu/hub/2018-119-16/1/183.pdf>.
13. Sidhartha G. IoT Based Pulse Oximeter / G. Sidhartha // *International Journal for Research in Applied Science and Engineering Technology*. – 2021. – Vol. 9, № 8. – P. 2946–2951. DOI: <https://doi.org/10.22214/ijraset.2021.37782>.
14. Use of wireless technologies in IoT projects / T. A. Vakaliuk, O. V. Andreiev, O. F. Dubyna [et al.] // *Journal of Edge Computing*. – 2024. – Vol. 3, № 2. – P. 147–167. DOI: <https://doi.org/10.55056/jec.750>.
15. MAX30102 Photometric Biosensor Coupled to ESP32-Webserver Capabilities for Continuous Point of Care Oxygen Saturation and Heart Rate Monitoring / [U. A. Contardi, M. Morikawa, B. Brunelli, D. V. Thomaz] // *Engineering Proceedings*. – 2022. – Vol. 16, № 1. – P. 9. DOI: <https://doi.org/10.3390/IECB2022-11114>.
16. Методика перевірки адекватності розроблених математичних моделей за даними проведених вимірювань / [О. В. Андрєєв, О. М. Мукомел, П. П. Топольницький, С. П. Фриз] // *Вісник ЖДТУ. Серія «Технічні науки»*. – 2017. – Vol. 1(40). – P. 65–69. DOI: 10.26642/tn-2007-1(40)-65-69
17. Korn T. M. *Mathematical Handbook for Scientists and Engineers: Definitions, Theorems, and Formulas for Reference and Review* / T. M. Korn, G. A. Korn. – Dover Publications, 2013. – 1152 p. – (Dover Civil and Mechanical Engineering).

МАТЕМАТИЧНЕ ТА КОМП'ЮТЕРНЕ МОДЕЛЮВАННЯ

MATHEMATICAL AND COMPUTER MODELING

UDC 004.94 : 004.2

PSEUDO-RANDOM ENCODING OF STATES IN THE ALGORITHM FOR ALGEBRAIC SYNTHESIS OF A FINITE STATE MACHINE

Babakov R. M. – Dr. Sc., Associate Professor, Professor of the Information Technologies Department, Vasyl' Stus Donetsk National University, Vinnytsia, Ukraine.

Barkalov A. A. – Dr. Sc., Professor, Professor of the Institute of Computer Science and Electronics, University of Zielona Gora, Zielona Gora, Poland.

Titarenko L. A. – Dr. Sc., Professor, Professor of the Institute of Computer Science and Electronics, University of Zielona Gora, Zielona Gora, Poland.

ABSTRACT

Context. The problem of algebraic synthesis of finite state machine with datapath of transitions is considered. The circuit of this state machine may require less hardware expenses and have a lower cost compared to circuits of other classes of digital control units. The object of research is the process of finding complete and partial solutions of the problem of algebraic synthesis of finite state machine using specialized algorithms. One of such algorithms is the previously known algorithm of complete sequential enumeration of state coding variants with a fixed set of transition operations. In the vast majority of cases, complete sequential enumeration is performed too long, which makes its practical application in the process of synthesizing of finite state machines with operational transformation of state codes impossible. This paper proposes a new approach, which consists in replacing complete sequential enumeration of state coding variants with pseudo-random coding. This allows you to increase the number of state codes that change in each iteration of the algorithm and can contribute to a faster search for satisfactory solutions to the algebraic synthesis problem.

Objective. Development and research of an algorithm for finding solutions to the algebraic synthesis problem of a finite state machine with datapath of transitions based on pseudo-random selection of state codes.

Method. The research is based on the structure of finite state machine with datapath of transitions. The synthesis of the finite state machine circuit involves a mandatory stage of algebraic synthesis, the result of which is the combination of a certain way of states encoding with the assignment of arithmetic-logical operations to state machine transitions. Such combination is called the solution to the algebraic synthesis problem. In the general case, there are many solutions for a given finite state machine, each of which can be either complete (when operations are mapped to all transitions) or partial (when part of transitions cannot be implemented using any of the given operations). The more transitions are implemented by given operations, the less hardware expenses will be required to implement the state machine circuit and the better solution found. The search for the best solution requires consideration of a large number of possible variants of states encoding. The paper includes a modification of a previously known algorithm, which consists in replacing the complete sequential enumeration of variants of states encoding with pseudo-random code generation. Both algorithms were implemented in the form of software using the Python language and tested on the example of a finite state machine that implements an abstract control algorithm. In the course of the experiments, it was investigated which of the algorithms would find the best solution to algebraic synthesis problem in a fixed time. The experiments were repeated for different sets of transition operations. The purpose of the experiments was to evaluate which of state code assignment strategies is more effective: sequential enumeration of state codes or their pseudo-random generation.

Results. Using the example of an abstract control algorithm, it is demonstrated that in general, pseudo-random assignment of state codes allows finding better solutions to the algebraic synthesis problem in the same time than sequential enumeration of state codes. Factors such as computer speed or the method of pseudo-random generation of state codes do not have a significant impact on the results of the experiments. The advantage of pseudo-random generation of state codes is preserved when using different sets of transition operations.

Conclusions. The basis of the algebraic synthesis of finite state machine with datapath of transitions is an algorithm for finding solutions to the algebraic synthesis problem. The article proposes an algorithm for finding such solutions based on pseudo-random encoding of finite state machine states. The software implementation of this algorithm has proven that such approach is generally better than sequential enumeration for state encoding variants, since it allows finding better solutions (solutions with fewer operationally unimplemented transitions) in the same time. The pseudo-random assignment of state codes can be the basis of future algorithms for the algebraic synthesis of finite state machines.

KEYWORDS: finite state machine, datapath of transitions, arithmetic and logical operations, algebraic synthesis algorithm, sequential enumeration, pseudo-random generation of state codes.

ABBREVIATIONS

FSM is a finite state machine;
CU is a control unit
DT is a datapath of transitions;
GSA is a graph-scheme of algorithm;
TO is a transition operation.

NOMENCLATURE

A is a sets of FSM states;
 X is a sets of logical conditions;
 Y is a sets of microoperations;
 M is a number of FSM states;
 L is a number of logical conditions analyzed by FSM;
 P is a number of microoperations formed by FSM;
 R is a bit capacity of state code;
 O is a set of transitions operations;
 O_i is an element of set O ;
 I is a number of TO;
 B is a number of uncovered FSM transitions;
 $B_{\min}$ is a minimal number of uncovered transitions;
 a^t is a current state of FSM;
 $K(a^t)$ is a code of current state;
 a^{t+1} is a state if transition;
 $K(a^{t+1})$ is a code of state of transition;
 G is a graph-scheme of implemented control algorithm.

INTRODUCTION

Any digital system includes a CU, which coordinates the joint operation of all system blocks [1–3]. Structurally, CU can be implemented in the form of a FSM, where a given control algorithm is implemented in a circuit way [4–6]. The FSM circuit allows for FSM transitions that depend on several input signals (the so-called multidirectional transitions) to be performed in one cycle of operation, while other types of CUs spend several cycles on this. Due to this, FSM has the highest speed among all types of CUs and is used in critical application systems, where speed is a decisive factor. At the same time, the FSM circuit is characterized by the highest hardware expenses compared to other types of CUs. This makes the scientific problem of reducing hardware expenses in the FSM circuit relevant [4, 7]. Reducing hardware expenses allows improving such characteristics of the circuit as cost, dimensions, reliability, etc. [8–10].

In [11], an approach to constructing an FSM circuit is proposed, which involves the transformation of state codes during FSM transitions using a certain set of arithmetic-logical operations. This approach, known as operational transformation of state codes, leads to FSM structure with datapath of transitions, in which some of transitions are implemented using a set of operations, and some are implemented in a canonical way using a system of Boolean equations [12, 13]. Each arithmetic-logical operation (called in this structure as transition operation) is implemented in the datapath by a separate combinational circuit with fixed hardware expenses. The

structure of FSM with DT allows you to use any TO to implement any number of FSM transitions (provided that the states are appropriately coded). This allows not to increase (or to increase to a lesser extent) the hardware expenses while increasing the number of transitions implemented by given TOs compared to other FSM structures.

One of stages of synthesis of FSM with DT is so-called algebraic synthesis, which involves a special coding of states, which is coordinated in a certain way with formation of set of transition operations [12]. The set of state codes together with a given set of OPs is called a formal solution to the algebraic synthesis problem. For a given FSM with a fixed set of OPs, there are as many formal solutions as there are different variants of states encoding. If we consider a separate solution, we can determine that part of FSM transitions can be implemented using a given set of OPs (is covered by the given OPs), and the other part cannot be implemented by any of the given OPs (is not covered by the OPs) and must be implemented in a canonical way using a system of Boolean equations. Among the formal solutions found, the best is the one that covers a larger number of FSM transitions. Formal solutions to the algebraic synthesis problem, in which all transitions are covered, are called as complete solutions, otherwise they are called as partial solutions [12, 13].

Currently, there are no known algorithms, rules, or approaches that allow algebraic synthesis to be performed in an optimal way in a short time. In [14], an algorithm for algebraic synthesis is proposed, based on a complete sequential enumeration for state coding options for a fixed set of transition operations. The application of this algorithm to FSM of even low complexity requires a lot of time, which does not allow finding the best possible formal solution. Therefore, the algorithm provides for a time limit in its operation, during which it searches for a formal solution with a minimal number of uncovered FSM transitions. The result of the algorithm can be either a complete or a partial solution to the algebraic synthesis problem, but only within the range of solutions searched for a certain time. Usually, the longer the complete enumeration algorithm runs, the better solutions to the algebraic synthesis problem will be found.

This article is devoted to the development and research of the algorithm for finding formal solutions to the problem of algebraic synthesis that uses other method of states encoding than sequential enumeration of variants.

The object of the study is the process of finding complete and partial solutions to the problem of algebraic synthesis of an FSM using specialized state coding algorithms.

The subject of the study is algorithms for finding complete and partial solutions to the problem of algebraic synthesis of FSM with DT based on a complete sequential enumeration or pseudo-random generation of FSM state codes.

The purpose of the work is to develop and study an algorithm for finding solutions to the problem of algebraic synthesis of finite state machine with datapath of transitions based on a pseudo-random generation of state codes.

1 PROBLEM STATEMENT

Let a finite state machine with datapath of transitions be given in the form of GSA [10, 15, 16]. According to GSA, a set of states $A = \{a_0, \dots, a_{M-1}\}$, a set of input signals $X = \{x_0, \dots, x_L\}$ and a set of microoperations $Y = \{y_1, \dots, y_P\}$ are formed. The set of transition operations $O = \{O_1, \dots, O_I\}$ is also given. Each element of $O_i \in O$ represents a certain arithmetic-logical operation, which implies the appropriate interpretation of the binary codes of the machine states. The GSA and the set O are the input data for the algebraic synthesis of the FSM with DT.

This article solves the problem of developing and studying algorithm for the algebraic synthesis of FSM with DT according to a given GSA, which uses pseudo-random generation of state codes.

2 REVIEW OF THE LITERATURE

The known methods of optimizing the logic circuit of a finite state machine usually lead to changes in its structure [3, 6, 10]. In this article, we consider the structure of an FSM with DT [11]. In it, the transition function is realized by datapath of transitions, which implements arithmetic and logic operations to convert state codes. The synthesis of FSM with DT based on a given GSA is considered in [12].

In [14], the algorithm for finding formal solutions to the problem of algebraic synthesis of FSM with DT according to a given GSA was proposed. In this paper, this algorithm will be denoted by A1. It assumes a fixed set of transition operations and uses a sequential enumeration of all possible state coding variants. For each considered variant of state coding, the algorithm determines the number of uncovered transitions. This number allows the algorithm to choose the coding variant in which the number of uncovered transitions is minimal among all considered variants.

Let a given GSA be denoted by M states of a Mealy or Moore finite state machine, for which $R = \lceil \log_2 M \rceil$ binary bits are sufficient to encode [3, 10]. In total, there are 2^R various codes that are usually interpreted in FSM with DT as binary vectors (in logical operations) or as scalar numbers without a sign (in arithmetic operations). The number of variants for encoding M states by 2^R codes is determined by the combinatorial formula for the number of permutations in accordance with expression (1).

$$N = \frac{(2^R)!}{(2^R - M)!}. \quad (1)$$

As an example, consider the GSA G shown in Fig. 1. The GSA is abstract and is denoted by $M = 18$ states $a_0 - a_{17}$ of the Moore state machine. Since the operational transformation of state codes concerns only the transition function of the FSM and is not related to the output function, the GSA in Fig. 1 does not contain microoperations (output signals). Instead, the operational nodes indicate the corresponding states of the Moore state machine. The initial and final nodes are marked with a same state a_0 , which indicates the cyclicity of the control algorithm [3]. To encode $M = 18$ states, $R = 5$ binary digits are sufficient. Five binary digits allow for 32 different codes that can be used to encode the states.

According to (1), when $M = 18$, $R = 5$, the number N of state coding variants is approximately equal to 3×10^{24} . The authors have implemented algorithm A1 in Python, which on i5–13500 processor allows analyzing about 200,000 variants in one second. Under such conditions, it would take 50×10^{10} years to go through all 3×10^{24} variants sequentially. Even a theoretical increase in the program's performance by a million times would require 500,000 years for a complete search, which is unpleasant for practical use. It should be understood that the GSA G given as an example is relatively small, while the graph-schemes of real control algorithms can have a number of states several times or dozens of times larger.

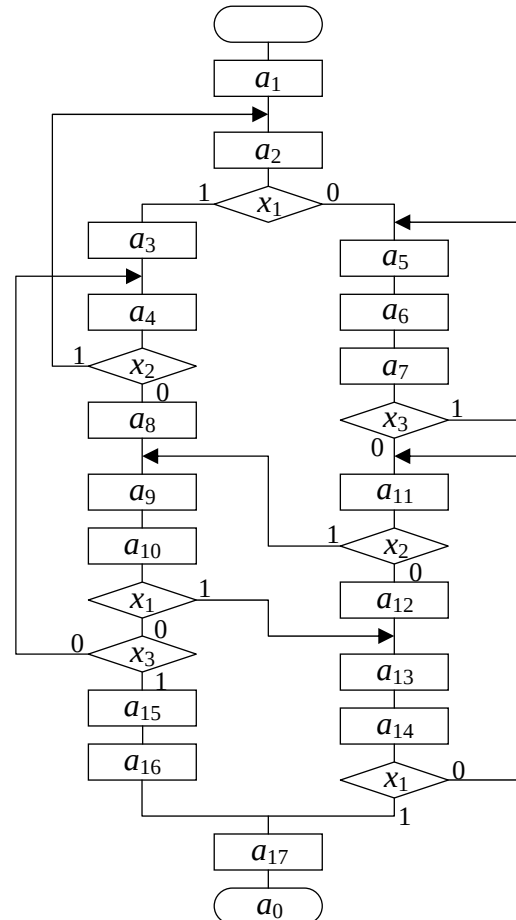


Figure 1 – GSA G

It should also be noted that the complete enumeration is performed for a given set of transition operations. Even if we do manage to perform complete enumeration of state encoding options, there is no guarantee that a complete solution to the algebraic synthesis problem will be found (i.e., a solution with all transitions covered). Indeed, the best way to encode states will be found, but it may turn out to be a partial solution with a certain number of uncovered transitions, and the resulting state machine circuit may not have a gain in hardware expenses.

In this case, we will be forced to choose another set of transition operations and re-execute complete enumeration for state coding variants. We will repeat this process until we select a set of TOs such that the found solution to the algebraic synthesis problem will allow us to obtain a state machine circuit with satisfactory characteristics.

Thus, it is obvious that it is impossible to use complete enumeration of state coding variants in practice. In this aspect, the A1 algorithm proposed in [14] provides for a time limit on execution. At the beginning of the work, a certain time t_A is set, upon reaching which the algorithm stops working and produces the best solution to the algebraic synthesis problem among the found solutions. This can be either a complete solution or a partial solution with a minimum number of uncovered transitions among all solutions considered during time t_A . It is quite possible that the best solution found in a limited time will allow you to design a state machine circuit with better (lower) hardware expenses. Thus, limiting the algorithm to the running time is currently the only way to obtain practically useful results.

3 MATERIALS AND METHODS

During complete enumeration of coding variants, the codes of some states change more often, the codes of other states change less often. For example, in the case of GSA G , states $a_0 - a_{17}$ initially acquire codes equal to their indices:

0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17

Next, a certain algorithm for permutation (rearrangement) of state codes works. In the author's software implementation, the *permutations* function from the *itertools* module of the Python language is used to permute the codes. Here are the first few generated sets of state codes:

0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17

0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 18

0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 19

...

0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 28

0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 29

0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 30

0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 31

0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 17, 16

0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 17, 18

0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 17, 19

© Babakov R. M., Barkalov A. A., Titarenko L. A., 2025
DOI 10.15588/1607-3274-2025-4-3

0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 17, 20
0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 17, 21
0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 17, 22
0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 17, 23
...

On the sets shown, the code of the last state a_{17} is continuously changing and the code of the penultimate state a_{16} is changing once. The other codes remain unchanged. If we take any relatively short period of time (a minute or an hour), we can assume that the state codes hardly change during this time (except for the codes of the last few states). This is especially true in the case of large-sized GSAs (with tens or hundreds of states). The algorithm repeatedly processes almost the same set of codes for some period of time, repeatedly performing almost the same work. As a result, during this time, the algorithm will produce mostly similar results. It's good if the algorithm immediately finds a fairly "efficient" solution (with zero or few uncovered transitions). But if the first solutions found have a large number of uncovered transitions, you should expect that in the near future the algorithm will find solutions with approximately the same number of uncovered transitions.

From the above, we can conclude that the algorithm of complete sequential enumeration of state coding variants is generally inefficient. Even if we modify the process of code permutation so that the codes of the first rather than the last states are changed most often, this will not lead to any fundamental changes in the results, since within any small period of time, the vast majority of state codes will still remain unchanged.

The authors put forward the following hypothesis: an increase in the number of states that change their codes during each iteration of the process of enumeration for state coding variants contributes to finding better solutions (solutions with fewer uncovered transitions) for a fixed algorithm running time. To test the hypothesis, this article proposes a modification of algorithm A1, in which instead of a complete sequential enumeration for state coding variants, a pseudo-random selection of them is used. The modified algorithm is shown in Fig. 2 and is denoted by A2 in this article. Let us consider its features.

1. Similar to algorithm A1, algorithm A2 runs continuously for the allotted time t_A , after which it terminates and returns a formal solution (a state encoding variant and a set of transition operations) that has the smallest number of uncovered transitions among all solutions viewed during the runtime.

2. The algorithm can terminate ahead of schedule if a complete solution to the algebraic synthesis problem is found. Here, it is conventionally assumed that all complete solutions (if there are several for a given GSA and a set of TOs) lead to the same hardware expenses for implementing the FSM circuit. Thus, finding the first complete solution makes it unnecessary to search for other complete solutions.

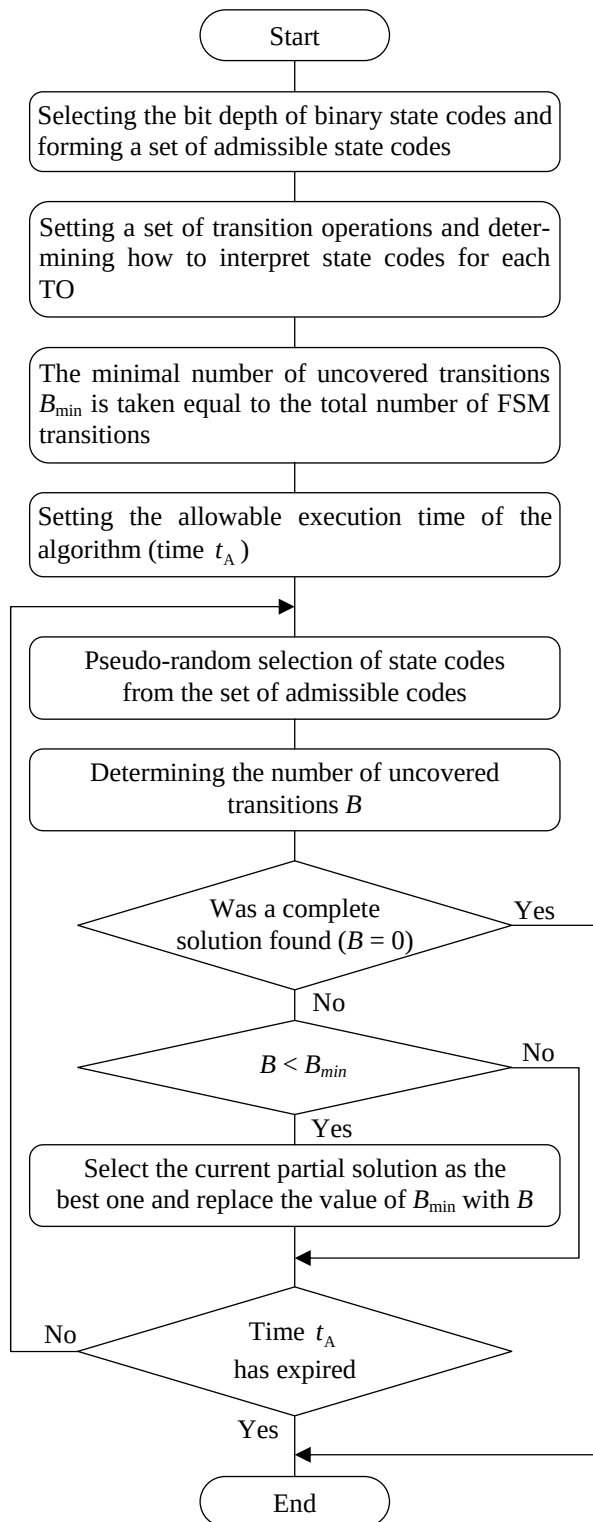


Figure 2 – Block diagram of the algorithm A2

3. At each iteration of the algorithm, the codes of all states are selected in a pseudo-random way among the set of admissible state codes. In Python, for example, the *shuffle* function of the standard *np.random* module can be used for this purpose, applied to the list of admissible state codes.

4. The search for the best partial solution is similar to algorithm A1. So, the number of uncovered transitions is determined by the matrix method proposed in [13].

5. Unlike algorithm A1, algorithm A2 allows for the re-generation of previously generated code sets. For example, if a sequence of codes was generated: 15, 5, 6, 25, 9, 0, 26, 28, 2, 18, 11, 3, 1, 7, 23, 22, 10, 12, nothing prevents this sequence from being generated repeatedly later. This fact reduces the performance of the algorithm to some extent, since the number of coding variants considered by algorithm A2 in time t_A will generally be less than the number considered under the same conditions by algorithm A1.

In practice, this can show itself as follows. Let's imagine a small GSA containing $M = 8$ states. To encode them, $R = 3$ binary digits are enough. With these values, expression (1) gives $N = 40320$ possible state encoding variants. Algorithm A1, capable of processing up to 200,000 coding variants per second, will process 40320 variants in about 0.2 seconds. During this time, it will find all complete solutions that exist for a given GSA and a set of TOs.

Algorithm A2 will work differently under the same conditions. Since the value of t_A is given, the algorithm will work for exactly that long, or until the first complete solution is found. If, for example, there is only one complete solution for a given GSA and a set of TOs, algorithm A2, when randomly selecting for state coding variants, may not generate the needed coding variant at all and fail to find a complete solution in the allotted time t_A . In this case, the user will be provided with the best of the partial solutions found, which may not be the absolutely best.

So, sometimes algorithm A2 can produce a worse result than algorithm A1. For practical applications, it is important to know which of these algorithms usually produces a better result on the same input data for the same running time. Since the answer to this question is not obvious, the authors conducted experimental studies of the comparative effectiveness of both algorithms. Let's consider these experiments.

4 EXPERIMENTS

The purpose of the experiments is to compare the efficiency of the algorithm A2 proposed in this paper and the prototype algorithm A1 considered in [14]. Since we are interested in finding formal solutions with fewer uncovered transitions, we will consider the algorithm that finds a solution with fewer uncovered transitions in the same time to be better. We will compare the algorithms using their software implementations in Python. The hardware used is a computer with i5–13500 processor running the Windows 11 operating system.

As you know, the search for solutions is influenced by a given GSA and a set of transition operations. When designing an FSM, its graph-scheme of algorithm is usually considered to be given and cannot be changed, while the set of transition operations remains variable. So,

we will conduct experiments using the example of a GSA G (Fig. 1), changing the set of transition operations in each experiment.

The experiment includes the following stages:

1. Setting the set of transition operations and operation time t_A .
2. Continuous execution of algorithm A1 for five minutes with fixing the best solution found (solution with minimal number of uncovered transitions).
3. Repeat step 2 for algorithm A2.
4. Printing the final results.

Let's consider the result of the experiment for the GSA G and five transition operations $O_1 - O_5$ given by expressions (2) – (6).

$$O_1: K(a^{t+1}) = K(a^t) + 3_{10}; \quad (2)$$

$$O_2: K(a^{t+1}) = K(a^t) \oplus 10011_2; \quad (3)$$

$$O_3: K(a^{t+1}) = K(a^t) \& 01111_2; \quad (4)$$

$$O_4: K(a^{t+1}) = K(a^t) \vee 01010_2; \quad (5)$$

$$O_5: K(a^{t+1}) = K(a^t) \div 4_{10}. \quad (6)$$

In this experiment, all the TOs are chosen at random. Each TO is a certain arithmetic or logic operation performed between two operands: code of current state of the FSM and a constant. In arithmetic operations (O_1, O_5), operands and result are interpreted as unsigned decimal numbers in range [0; 31]. In logical operations ($O_2 - O_4$), operands and result are interpreted as 5-bit binary vectors. In order to shorten the writing, we will denote these transition operations similarly to the notation in [14]: “+3”, “ $\oplus 19$ ”, “ $\& 15$ ”, “ $\vee 10$ ”, “ $\div 4$ ”.

Fig. 3 shows a screenshot of the program results. The results of algorithm A1 are located after the line “Algorithm A1”, the results of algorithm A2 are located after the line “Algorithm A2”. The result of each algorithm contains the following information:

- a list of used transition operations (written by the corresponding Python operators);
- the best solution found (state encoding variant);
- the number of uncovered transitions;
- the number of considered state encoding variants;
- the algorithm running time (300 seconds or 5 minutes).

Fig. 3 shows that in this case, algorithm A2 performed better than algorithm A1. It found a solution with 10 uncovered transitions, while algorithm A1 found a solution with 14 uncovered transitions. It also can be seen that algorithm A2 performs the state encoding variants more slowly. In 300 seconds, it managed to process about 54 million variants, while algorithm A1 managed to process about 62 million. This is probably due to the slower speed of the pseudo-random state code mixing function. Nevertheless, the slower speed did not prevent algorithm A2 from finding a better solution to the algebraic synthesis problem compared to algorithm A1.

The software implementation of the algorithms is designed so that slightly different results can be obtained at each run. For algorithm A1, the number of considered state encoding variants can differ, since in the conditions of a multitasking operating system, the processor access time allocated to the program will be slightly different each time. For algorithm A2, a completely different result can be obtained, since each time the program is run, the initial state of the pseudorandom number generator of the Python interpreter will be different. However, in terms of the ratio of the number of uncovered transitions, each new result of the program will be close to the one considered above.

```
Python 3.12.0 (tags/v3.12.0:0fb18b0, Oct 2 2023, 13:03:39) [MSC v.1935 64 bit (AMD64)] on win32
Type "help", "copyright", "credits" or "license()" for more information.
>>>
= RESTART: C:\Python312\A1_A2.py

Operation +3 added.
Operation ^19 added.
Operation &15 added.
Operation |10 added.
Operation /4 added.

Algorithm A1:

BEST SOLUTION FOUND:
Operations: +3 ^19 &15 |10 /4
a0=0 a1=1 a2=2 a3=3 a4=4 a5=5 a6=6 a7=7 a8=8 a9=9
a10=10 a13=13 a15=11 a11=26 a12=29 a14=16 a17=19
a16=14
Uncovered transitions: 14

Total permutations of state codes: 61670957

Finished in 300 seconds.

Algorithm A2:

BEST SOLUTION FOUND:
Operations: +3 ^19 &15 |10 /4
a0=16 a1=26 a2=2 a3=17 a4=20 a5=0 a6=10 a7=19
a8=31 a9=22 a10=25 a11=30 a12=14 a13=6
a14=13 a15=13
Uncovered transitions: 10

Total permutations of state codes: 54176186

Finished in 300 seconds.
>>>
```

Figure 3 – Example of program results

5 RESULTS

The experiment considered above was repeated for 20 different sets of TOs. The number of uncovered transitions obtained for each set of OPs by algorithms A1 and A2 is given in Table 1. Column “N” contains the row numbers of the table; columns $O_1 - O_5$ contain used TOs (one row of the table corresponds to one set of TOs); columns A1, A2 contain the number of uncovered transitions obtained by the corresponding algorithms for each set of TOs.

Among the values given in columns A1 and A2 of each row of the table, the smaller value is considered the better. For clarity, the smaller values are highlighted in bold and marked with a dark background. As can be seen, algorithm A1 had an advantage over algorithm A2 only in five cases, that is, in five out of twenty considered sets of OPs. In most cases, algorithm A2 found solutions with fewer uncovered transitions in the same running time. This may indicate a higher overall efficiency of pseudo-random encoding of states compared to continuous sequential enumeration of encoding variants.

Table 1 – Comparison of algorithms A1 and A2 by the number of uncovered transitions for different sets of OPs

N	Transition operations					Number of uncovered transitions	
	O_1	O_2	O_3	O_4	O_5	A1	A2
1	+ 3	+ 12	$\oplus$ 19	$\&$ 10	$\vee$ 11	14	10
2	+ 2	+ 19	$\oplus$ 6	$\oplus$ 27	$\&$ 15	14	10
3	+ 5	+ 8	+ 17	$\oplus$ 23	$\vee$ 19	16	10
4	$\oplus$ 1	$\oplus$ 3	$\oplus$ 6	$\oplus$ 18	$\oplus$ 29	7	10
5	+ 5	+ 12	+ 17	+ 25	+ 30	15	8
6	+ 1	$\oplus$ 12	$\oplus$ 18	$\oplus$ 25	$\oplus$ 31	9	9
7	+ 2	+ 10	$\&$ 19	$\vee$ 22	$\vee$ 29	15	11
8	+ 3	$\oplus$ 9	$\oplus$ 17	$\times$ 2	$\div$ 2	8	10
9	+ 1	+ 2	+ 3	+ 4	+ 5	7	8
10	$\oplus$ 15	$\oplus$ 28	$\div$ 4	$\&$ 22	$\vee$ 26	15	10
11	+ 2	+ 9	+ 15	+ 18	- 7	14	9
12	+ 5	$\&$ 7	$\&$ 23	$\vee$ 13	$\vee$ 23	18	11
13	+ 14	+ 15	$\oplus$ 10	$\div$ 2	$\times$ 4	13	10
14	+ 1	+ 2	- 1	- 2	$\times$ 8	9	12
15	+ 10	+ 15	$\oplus$ 11	$\oplus$ 18	$\oplus$ 26	17	10
16	+ 17	- 9	$\times$ 4	$\&$ 20	$\vee$ 25	17	10
17	+ 1	- 1	$\times$ 2	$\times$ 4	$\div$ 2	6	11
18	+ 21	- 6	$\oplus$ 14	$\vee$ 10	$\vee$ 21	17	10
19	- 2	- 5	- 11	- 20	- 27	12	9
20	+ 2	- 5	$\&$ 12	$\vee$ 23	$\div$ 4	14	11

In the experiment, algorithm A1 performed a sequential enumeration for state encoding variants for each set of OPs, always starting from the initial variant (when the state codes correspond to their indices): 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17

This allowed us to analyze only the very beginning of the complete enumeration and did not allow us to “peek” into its other intervals.

In order to investigate the efficiency of sequential search of state encoding variants at different intervals of complete enumeration, a modification of algorithm A1 was performed, which consists in the following. The entire list of admissible state codes (in the case of GSA G, these are codes 0, 1, 2, ..., 30, 31) is shuffled in a pseudo-random manner once at the beginning of the algorithm. Then the algorithm generates sequential state encoding variants, starting from this pseudo-random set of codes. This approach allows us to investigate the random “lifetime” of a complete sequential enumeration of state encoding variants and to evaluate the efficiency of algorithm A1 regardless of the moment of its work.

Let us denote the modified algorithm A1 by A1* and compare its efficiency with algorithm A2. To do this, we

will use the same sets of TOs and the same conditions as in the previous experiment. The results obtained are given in Table 2. As can be seen, in all cases considered, algorithm A2 showed better results compared to algorithm A1*. This is additional evidence of its effectiveness.

Table 2 – Comparison of algorithms A1* and A2

N	Transition operations					Number of uncovered transitions	
	O_1	O_2	O_3	O_4	O_5	A1*	A2
1	+ 3	+ 12	$\oplus$ 19	$\&$ 10	$\vee$ 11	15	10
2	+ 2	+ 19	$\oplus$ 6	$\oplus$ 27	$\&$ 15	14	11
3	+ 5	+ 8	+ 17	$\oplus$ 23	$\vee$ 19	14	9
4	$\oplus$ 1	$\oplus$ 3	$\oplus$ 6	$\oplus$ 18	$\oplus$ 29	15	9
5	+ 5	+ 12	+ 17	+ 25	+ 30	14	8
6	+ 1	$\oplus$ 12	$\oplus$ 18	$\oplus$ 25	$\oplus$ 31	16	10
7	+ 2	+ 10	$\&$ 19	$\vee$ 22	$\vee$ 29	17	10
8	+ 3	$\oplus$ 9	$\oplus$ 17	$\times$ 2	$\div$ 2	14	10
9	+ 1	+ 2	+ 3	+ 4	+ 5	14	10
10	$\oplus$ 15	$\oplus$ 28	$\div$ 4	$\&$ 22	$\vee$ 26	16	9
11	+ 2	+ 9	+ 15	+ 18	- 7	13	9
12	+ 5	$\&$ 7	$\&$ 23	$\vee$ 13	$\vee$ 23	17	12
13	+ 14	+ 15	$\oplus$ 10	$\div$ 2	$\times$ 4	14	9
14	+ 1	+ 2	- 1	- 2	$\times$ 8	17	13
15	+ 10	+ 15	$\oplus$ 11	$\oplus$ 18	$\oplus$ 26	15	8
16	+ 17	- 9	$\times$ 4	$\&$ 20	$\vee$ 25	15	10
17	+ 1	- 1	$\times$ 2	$\times$ 4	$\div$ 2	17	10
18	+ 21	- 6	$\oplus$ 14	$\vee$ 10	$\vee$ 21	14	10
19	- 2	- 5	- 11	- 20	- 27	14	9
20	+ 2	- 5	$\&$ 12	$\vee$ 23	$\div$ 4	14	9

6 DISCUSSION

The proposed algorithm for algebraic synthesis of FSM with DT (algorithm A2) differs from the prototype algorithm (algorithm A1) in the way of searching for variants of FSM states encoding. Due to the pseudo-random selection of state codes, the average number of states whose codes change in each subsequent iteration of searching increases. As a result, the new algorithm in most of the analyzed situations allowed finding better solutions to the algebraic synthesis problem compared to the prototype algorithm.

The higher efficiency of algorithm A2 was proven only experimentally and could not be predicted theoretically. The results obtained were influenced by factors such as the structure of the given GSA, the used sets of transition operations and the program running time. However, according to the authors, made restrictions made it possible to determine the general trend, which is the greater efficiency of algorithm A2 compared to algorithm A1 when searching for solutions with a smaller number of uncovered automaton transitions.

The numbers in last two columns of Table 1 and Table 2 should be considered only in the context of their mutual comparison. One should not look at the fact that no complete solution was found in the experiments conducted. This was not the goal, but if necessary, a complete solution could perhaps be found by selecting suitable TOs and increasing the algorithm runtime.

CONCLUSIONS

The article proposes a new algorithm for algebraic synthesis of finite state machine with datapath of transitions. Its feature is the enumeration of state coding variants by pseudo-random selection of codes. The goal of the algorithm is to find complete or partial solutions to the algebraic synthesis problem for a given graph-scheme of algorithm and a set of transition operations.

The scientific novelty of the article lies in the fact that for the first time an algorithm for searching for complete and partial solutions to the algebraic synthesis problem with pseudo-random generation of state codes has been developed, implemented in software, and studied. Experimental research of the test GSA had shown that the new algorithm is generally capable of finding more efficient solutions compared to the previously known prototype algorithm based on a complete sequential enumeration of state coding variants. This proves the usefulness of its development and the feasibility of using it as part of FSM synthesis algorithms.

The practical use of the obtained results is possible in the development of methods and algorithms for the synthesis of finite state machine with datapath of transitions within the framework of specialized CADs of digital control units.

Prospects for further research are to solve a range of scientific and practical problems related to optimize the running time of the proposed algorithm by parallelizing the complete enumeration algorithm on multi-core computing systems. Also, the authors see a separate direction in studying the effectiveness of the algorithm on large-sized GSAs, in particular on pseudo-randomly generated GSAs. As a result, this will allow a more accurate assessment of the effectiveness of FSM with operational transformation of state codes according to the criterion of hardware expenses in the device circuit by modeling the synthesis of FSM circuits using hardware description languages.

ACKNOWLEDGEMENTS

The work was conducted in the direction of the state budget scientific research project of Vasyl' Stus Donetsk National University "Methods, algorithms and tools of computer-aided design of control units of computing systems" (state registration number 0122U200085).

REFERENCES

1. Bailliul J., Samad T. Encyclopedia of Systems and Control. Springer, London, UK, 2015, 1554 p. DOI: <https://doi.org/10.1007/978-1-4471-5058-9>
2. Czerwinski R., Kania D. Finite state machines logic synthesis for complex programmable logic devices. Berlin, Springer, 2013, 172 p. DOI: <https://doi.org/10.1007/978-3-642-36166-1>
3. Baranov S. Logic and System Design of Digital Systems. Tallin, TUTPress, 2008, 267 p.
4. Micheli G. Synthesis and Optimization of Digital Circuits. McGraw-Hill, Cambridge, MA, USA, 1994, 579 p.
5. Minns P., Elliot I. FSM-Based Digital Design Using Verilog HDL. JohnWiley and Sons, Hoboken, NJ, USA, 2008, 408 p. DOI: <https://doi.org/10.1002/9780470987629>
6. Skliarova I., Sklyarov V., Sudnitsion A. Design of FPGA-based circuits using hierarchical finite state machines. Tallinn, TUT Press, 2012, 240 p.
7. Klimovich A. S., Solov'ev V. V. Minimization of Mealy finite-state machines by internal states gluing, *Journal of Computer and Systems Science International*, 2012, Volume 51, pp. 244–255. DOI: <https://doi.org/10.1134/S1064230712010091>
8. Grout I. Digital Systems Design with FPGAs and CPLDs. Elsevier Science, Amsterdam, The Netherlands, 2008, 784 p. DOI: <https://doi.org/10.1016/B978-0-7506-8397-5.X0001-3>
9. Kubica M., Opara A., and Kania D. Technology Mapping for LUT-Based FPGA, *Lecture Notes in Electrical Engineering*, 2021, Vol. 713, 207 p. DOI: <https://doi.org/10.1007/978-3-030-60488-2>
10. Baranov S. Logic Synthesis for Control Automata. Dordrecht, Kluwer Academic Publishers, 1994, 312 p.
11. Barkalov A. A. Babakov R. M. Operational formation of state codes in microprogram automata, *Cybernetics and Systems Analysis*, 2011, Volume 47 (2), pp. 193–197. DOI: <https://doi.org/10.1007/s10559-011-9301-y>
12. Barkalov A. A., Titarenko L. A., Babakov R. M. Synthesis of Finite State Machine with Datapath of Transitions According to the Operational Table of Transitions, *Radio Electronics, Computer Science, Control*, 2022, Volume 3 (62), pp. 109–119. DOI: <https://doi.org/10.15588/1607-3274-2022-3-11>
13. Barkalov A. A., Babakov R. M. A Matrix Method for Detecting Formal Solutions to the Problem of Algebraic Synthesis of a Finite-State Machine with a Datapath of Transitions, *Cybernetics and Systems Analysis*, 2023, Volume 59 (2), pp. 190–198. DOI: <https://doi.org/10.1007/s10559-023-00554-6>
14. Babakov R. M., Barkalov A. A., Titarenko L. A., Voitenko M. O. Algorithmic Differences of Complete and Partial Algebraic Synthesis of a Finite State Machine with Datapath of Transitions, *Radio Electronics, Computer Science, Control*, 2024, Volume 4, pp. 143–152. DOI: <https://doi.org/10.15588/1607-3274-2024-4-14>
15. Kubica M., Kania D., Kulisz J. A Technology Mapping of FSMs Based on a Graph of Excitations and Outputs, *IEEE Access*, 2019, Volume 7, pp. 16123–16131. DOI: <https://doi.org/10.1109/ACCESS.2019.2895206>
16. Baranov S. Finite State Machines and Algorithmic State Machines. Seattle, WA, USA: Amazon, 2018, 185 p.

Received 14.07.2025.
Accepted 29.10.2025.

ПСЕВДОВИПАДКОВЕ КОДУВАННЯ СТАНІВ В АЛГОРИТМІ АЛГЕБРАІЧНОГО СИНТЕЗУ МІКРОПРОГРАМНОГО АВТОМАТА

Бабаков Р. М. – д-р техн. наук, доцент, професор кафедри інформаційних технологій Донецького національного університету імені Василя Стуса, м. Вінниця, Україна.

Баркалов О. О. – д-р техн. наук, професор, професор Інституту комп'ютерних наук та електроніки університету Зеленогурського, м. Зельона Гура, Польща.

Тітаренко Л. О. – д-р техн. наук, професор, професор Інституту комп'ютерних наук та електроніки університету Зеленогурського, м. Зельона Гура, Польща.

АНОТАЦІЯ

Актуальність. Розглянуто задачу алгебраїчного синтезу мікропрограмного автомата з операційним автоматом переходів. Схема цього автомата може потребувати менших витрат апаратури і мати меншу вартість у порівнянні зі схемами інших класів цифрових пристроїв керування. Об'єктом дослідження є процес пошуку повних і часткових розв'язків задачі алгебраїчного синтезу автомата із використанням спеціалізованих алгоритмів. Одним із таких алгоритмів є раніше відомий алгоритм повного послідовного перебору варіантів кодування станів при фіксованій множині операцій переходів. У переважній більшості випадків повний послідовний перебір виконується надто довго, що унеможлиблює його практичне застосування в процесі синтезу автоматів з операційним перетворенням кодів станів. В даній роботі пропонується новий підхід, який полягає у заміні повного послідовного перебору варіантів кодування станів на псевдовипадкове кодування. Це дозволяє збільшити кількість кодів станів, що змінюються у кожній ітерації алгоритму, і може сприяти більш швидкому пошуку задовільних розв'язків задачі алгебраїчного синтезу.

Мета. Розробка і дослідження алгоритму пошуку розв'язків задачі алгебраїчного синтезу мікропрограмного автомата з операційним автоматом переходів на основі псевдовипадкового вибору кодів станів.

Метод. В основу дослідження покладено структуру мікропрограмного автомата з операційним автоматом переходів. Синтез схеми автомата передбачає обов'язковий етап алгебраїчного синтезу, результатом якого є поєднання певного способу кодування станів із заставленням арифметико-логічних операцій автоматним переходом. Таке поєднання називається розв'язком задачі алгебраїчного синтезу. В загальному випадку для заданого автомата існує багато розв'язків, кожен з яких може бути як повним (коли операції зіставлені усім переходам) або частковим (коли частина переходів не може бути реалізована за допомогою жодної із заданих операцій). Чим більше переходів реалізуються заданими операціями, тим менше апаратних витрат потребуватиме реалізація схеми автомата і тим кращим є знайдений розв'язок. Пошук кращих розв'язків потребує розгляду великої кількості можливих варіантів кодування станів. В роботі здійснено модифікацію раніше відомого алгоритму, яка полягає у заміні повного послідовного перебору варіантів кодування станів на псевдовипадкову генерацію кодів. Обидва алгоритми були реалізовані програмно за допомогою мови Python і протестовані на прикладі мікропрограмного автомата, що імплементує абстрактний алгоритм керування. У процесі експериментів досліджувалось, який із алгоритмів знайде кращий розв'язок задачі алгебраїчного синтезу за фіксований час. Експерименти повторювались для різних наборів операцій переходів. Метою експериментів було оцінити, яка із стратегій завдання кодів станів є більш ефективною: послідовний перебір кодів станів чи їх псевдовипадкова генерація.

Результати. На прикладі абстрактного алгоритму керування продемонстровано, що загалом псевдовипадкове завдання кодів станів дозволяє за однаковий час знайти кращі розв'язки задачі алгебраїчного синтезу, ніж послідовний перебір кодів станів. Такі фактори, як швидкодія комп'ютера чи метод псевдовипадкової генерації кодів станів, не мають суттєвого впливу на результат експериментів. Перевага псевдовипадкової генерації кодів станів зберігається при використанні різних наборів операцій переходів.

Висновки. В основі алгебраїчного синтезу мікропрограмного автомата з операційним автоматом переходів лежить алгоритм пошуку розв'язків задачі алгебраїчного синтезу. В роботі запропоновано алгоритм пошуку таких розв'язків, оснований на псевдовипадковому кодуванні станів автомата. Програмна реалізація цього алгоритму довела, що такий підхід в загальному випадку є кращим за послідовний перебір варіантів кодування станів, оскільки дозволяє знайти кращі розв'язки (розв'язки з меншою кількістю операційно нереалізованих переходів) за той самий час. Псевдовипадкове завдання кодів станів може бути покладене в основу майбутніх алгоритмів алгебраїчного синтезу мікропрограмних автоматів.

КЛЮЧОВІ СЛОВА: мікропрограмний автомат, операційний автомат переходів, арифметико-логічні операції, алгоритм алгебраїчного синтезу, послідовний перебір, псевдовипадкова генерація кодів станів.

ЛІТЕРАТУРА

1. Bailliul J. Encyclopedia of Systems and Control / J. Bailliul, T. Samad. – Springer: London, UK, 2015. – 1554 p. DOI: <https://doi.org/10.1007/978-1-4471-5058-9>
2. Czerwinski R. Finite state machines logic synthesis for complex programmable logic devices / R. Czerwinski, D. Kania. – Berlin : Springer, 2013. – 172 p. DOI: <https://doi.org/10.1007/978-3-642-36166-1>
3. Baranov S. Logic and System Design of Digital Systems / S. Baranov. – Tallin : TUTPress, 2008. – 267 p.
4. Micheli G. Synthesis and Optimization of Digital Circuits / G. Micheli. – McGraw-Hill : Cambridge, MA, USA, 1994. – 579 p.

5. Minns P. FSM-Based Digital Design Using Verilog HDL / P. Minns, I. Elliot. – JohnWiley and Sons : Hoboken, NJ, USA, 2008. – 408 p. DOI: <https://doi.org/10.1002/9780470987629>
6. Skliarova I. Design of FPGA-based circuits using hierarchical finite state machines / I. Skliarova, V. Sklyarov, A. Sudnitson. – Tallinn : TUT Press, 2012. – 240 p.
7. Klimovich A. S. Minimization of Mealy finite-state machines by internal states gluing / A. S. Klimovich, V. V. Solov'ev // Journal of Computer and Systems Science International. – 2012. – Volume 51. – P. 244–255. DOI: <https://doi.org/10.1134/S1064230712010091>
8. Grout I. Digital Systems Design with FPGAs and CPLDs / I. Grout. – Elsevier Science: Amsterdam, The Netherlands, 2008. – 784 p. DOI: <https://doi.org/10.1016/B978-0-7506-8397-5.X0001-3>
9. Kubica M. Technology Mapping for LUT-Based FPGA / M. Kubica, A. Opara, and D. Kania // Lecture Notes in Electrical Engineering. – 2021. – Vol. 713. – 207 p. DOI: <https://doi.org/10.1007/978-3-030-60488-2>
10. Baranov S. Logic Synthesis for Control Automata / S. Baranov. – Dordrecht : Kluwer Academic Publishers, 1994. – 312 p.
11. Barkalov A. A. Operational formation of state codes in microprogram automata / A. A. Barkalov, R. M. Babakov // Cybernetics and Systems Analysis. – 2011. – Volume 47 (2). – P. 193–197. DOI: <https://doi.org/10.1007/s10559-011-9301-y>
12. Barkalov A. A. Synthesis of Finite State Machine with Datapath of Transitions According to the Operational Table of Transitions / A. A. Barkalov, L. A. Titarenko, R. M. Babakov // Radio Electronics, Computer Science, Control. – 2022. – Volume 3 (62). – P. 109–119. DOI: <https://doi.org/10.15588/1607-3274-2022-3-11>
13. Barkalov A. A. A Matrix Method for Detecting Formal Solutions to the Problem of Algebraic Synthesis of a Finite-State Machine with a Datapath of Transitions / A. A. Barkalov, R. M. Babakov // Cybernetics and Systems Analysis. – 2023. – Volume 59 (2). – P. 190–198. DOI: <https://doi.org/10.1007/s10559-023-00554-6>
14. Algorithmic Differences of Complete and Partial Algebraic Synthesis of a Finite State Machine with Datapath of Transitions / [R. M. Babakov, A. A. Barkalov, L. A. Titarenko, M. O. Voitenko] // Radio Electronics, Computer Science, Control. – 2024. – Volume 4. – P. 143–152. DOI: <https://doi.org/10.15588/1607-3274-2024-4-14>
15. Kubica M. A Technology Mapping of FSMs Based on a Graph of Excitations and Outputs / M. Kubica, D. Kania, J. Kulisz // IEEE Access. – 2019. – Volume 7. – P. 16123–16131. DOI: <https://doi.org/10.1109/ACCESS.2019.2895206>
16. Baranov S. Finite State Machines and Algorithmic State Machines / S. Baranov. – Seattle, WA, USA : Amazon, 2018. – 185 p.

A METHOD FOR DETERMINING THE FUZZY DISCRETE FRÉCHET DISTANCE

Berezsky O. M. – Dr. Sc., Professor, Professor of the Department of Computer Engineering, West Ukrainian National University, Ternopil, Ukraine.

Berezkyi M. O. – Post-graduate student of the Department of Computer Engineering, West Ukrainian National University, Ternopil, Ukraine.

Zarichnyi M. M. – Dr. Sc., Professor, Professor of the Department of Algebra, Topology and Foundations of Mathematics, Ivan Franko Lviv National University, Lviv, Ukraine.

ABSTRACT

Context. The article addresses the problem of image similarity assessment based on the Fréchet distance metric and its modifications. In this context, images are approximated by polygonal curves. The problem arises from the need to quantitatively evaluate image similarity for tasks such as image generation, clustering, and recognition. Quantitative assessment of the proximity of bio-medical images supports decision-making in automated diagnostic systems. The object of the study is the process of image similarity evaluation. The subject of the study is the Fréchet distance metric and its modifications.

Objective. To develop a method for determining the fuzzy discrete Fréchet distance, to evaluate the computational complexity of the proposed method, to implement the algorithm for determining the fuzzy discrete Fréchet distance in software, and to conduct computational experiments to evaluate the fuzzy discrete Fréchet distance between polygons.

Method. The article presents a method for determining the fuzzy discrete Fréchet distance based on the fuzzy Fréchet metric between polygonal curves. The fuzzy Fréchet metric is grounded in the classical Fréchet distance defined on the space of parameterized curves. The required approximation for practical applications is achieved through the discretization of the fuzzy Fréchet metric. The developed method estimates the fuzzy discrete Fréchet distance between polygonal curves by adapting the algorithm for computing the classical discrete Fréchet distance.

Results. The computer experiments were conducted on a set of predefined regions approximated by polygonal curves. Based on the proposed method, an algorithm was developed to evaluate the discrete fuzzy Fréchet distance. The developed algorithm exhibits low computational complexity, equal to the product of the discretized segments of the polygonal curves: $O(Cm \cdot n)$. This enables the estimation of the discrete Fréchet distance with a specified similarity threshold. The software implementation of the method is intended to be integrated into an automatic medical diagnostic system.

Conclusions. The results obtained in the study allow recommending the developed method for evaluating image similarity based on the fuzzy discrete Fréchet distance for broad application in computer vision systems, including image generation, clustering, and recognition.

KEYWORDS: metric, fuzzy Fréchet metric, fuzzy discrete Fréchet metric, image, polygonal curves.

ABBREVIATIONS

MRI is a Magnetic Resonance Imaging;
RAM is a Random Access Memory;
GB is a gigabyte;
Java is a programming language;
OpenCV is an Open Source Computer Vision Library;
HDD is a Hard Disk Drive;
Windows is a Microsoft's operating system;
Linux is a family of open-source operating systems;
MacOS is an Apple's operating system;
OpenCV 3.0 is a version 3 of the OpenCV library;
Java8+ is a Java version 8 or higher;
intelliJIDEA is an Integrated Development Environment.

NOMENCLATURE

X is a topological space;
 (X, d) is a metric space;
 $C(X)$ is a set of parametric curves in X ;
 $d(x, y)$ is a distance between points x and y ;
 I is a segment $[0, 1]$;
 $H(I)$ is a group of homeomorphisms of the segment I ;
 γ_1 is a parametric curve;
 γ_2 is a parametric curve;

$\sigma(\gamma_1)$ is a polygonal approximations of γ_1 ;
 $\sigma(\gamma_2)$ is a polygonal approximations of γ_2 ;
 $\alpha_{1,2}$ is an arbitrary continuous non-decreasing function on the interval $[0,1]$, such that $\alpha(0) = 0$ and $\alpha(1) = 1$;
 s is a parameter defined on the interval $[0, 1]$;
 d_F is a discrete Fréchet distance;
 M_d is a fuzzy metric in the sense of George and Veeramani;
 M_F is a fuzzy Fréchet metric;
 Md_F is a discrete fuzzy Fréchet metric;
 $t \in (0; \infty)$ is a threshold parameter for fuzzy distance;
 n is a number of segments of curve γ_1 ;
 m is a number of segments of curve γ_2 ;
 x is an independent variable;
 y is an independent variable;
 z is an independent variable;
 $dC(X)$ is a set of discrete curves in X .
 q is a number of segments of parameter t ;
 p is an index of curve pairs γ_1 and γ_2 ;
 L is a set of node-like (nodal) pairs of curves γ_1 and γ_2 ;
 k is a number of curve pairs γ_1 and γ_2 ;
 i, j are indices of the sequences of points of the pair of curves γ_1 and γ_2 ;

$O_1, O_2 \subseteq \mathbb{R}^2$ are convex regions;
 $\partial O_1, \partial O_2$ are boundaries of regions O_1, O_2 ;
 $O(d_F)$ is a computational complexity of discrete Fréchet distance;
 N is a fuzzy norm.

INTRODUCTION

At the low level, the input is an image. After applying spatial and frequency filtering algorithms [1], as well as edge enhancement algorithms, the output is an enhanced image.

At the middle level, the input is the enhanced image, and the output consists of segmented regions. Edge and texture extraction procedures are applied to these segments. Upon describing the contours and textures, feature vectors are obtained.

At the high level, the input data comprise contour and texture feature vectors. The output of this level is a set of class labels assigned to the input images.

Thus, for each level of computer vision processing, it is necessary to quantitatively evaluate the quality of image transformation algorithms.

Fuzzy metrics are commonly employed for these tasks.

At the low level of image processing, a primary application of fuzzy metrics is noise reduction. Nebojša M. Ralević, Marija Paunović, and Bratislav Iričanin introduced the notions of fuzzy T-metrics and fuzzy S-metrics, and demonstrated their application for image denoising [2].

In another study [3], the authors presented examples of fuzzy metrics in the sense of George and Veeramani, and provided applications of such metrics in color image processing.

In [4], fuzzy wavelets are introduced, which are employed in neural network training.

In [5], the notions of fuzzy T-metric and fuzzy S-metric are connected with image quality indicators, and the proposed approach is used for image filtering.

At the middle level, fuzzy metrics are typically applied for color image comparison, segmentation, and skeletonization of images.

In [6], the notion of fuzzy colors and fuzzy color spaces is introduced and applied to color image comparison. Authors S. Vinurajkumar and S. Anandhavelu applied fuzzy metrics to the segmentation of MRI images [7]. In [8], the notion of a fuzzy skeleton is introduced and applied for image comparison.

The high level of computer vision is concerned with clustering and classification of images based on fuzzy metrics.

In [9], fuzzy clustering procedures are described. In [10], researchers utilize fuzzy metrics to estimate distance measures when aligning a 3D skull model with facial photographs. The authors of [11] propose a novel fuzzy clustering method, Fuzzy C-Ordered-Means, which performs well on noisy data. In [12], an integrated clustering approach called Rough Fuzzy K-Modes is introduced. For

fuzzy cluster optimization, simulated annealing and random forest algorithms are applied. In [13], a fuzzy granulation-degranulation criterion is proposed to evaluate the quality of data separation. The authors of [14] used fuzzy models to construct cascade classifiers that were applied to biomedical image classification.

As the analysis of computer vision levels demonstrates, the problem of image assessment using fuzzy metrics remains highly relevant.

The subject of the research is the process of determining the fuzzy discrete Fréchet distance between images.

The object of the research is fuzzy metrics.

The purpose of the research is to develop a method for computing the fuzzy discrete Fréchet distance.

1 PROBLEM STATEMENT

Let (X, d) be a given metric space. Additionally, suppose that two continuous parametric curves γ_1 and γ_2 are defined.

Then, the Fréchet distance between these curves is defined as follows:

$$d_F(\gamma_1, \gamma_2) = \inf \sup_{t \in I} \{d(\gamma_1(\alpha_1(t)), \gamma_2(\alpha_2(t)))\},$$

where the infimum is taken over all reparameterizations $\alpha_1, \alpha_2: [0, 1] \rightarrow [0, 1]$.

Let the two continuous parametric curves γ_1 and γ_2 be approximated by piecewise linear curves. Then we obtain polygonal curves defined as:

$$\sigma(\gamma_1) = \{\gamma_1(1), \gamma_1(2), \dots, \gamma_1(n)\}, \\ \sigma(\gamma_2) = \{\gamma_2(1), \gamma_2(2), \dots, \gamma_2(m)\}.$$

In [15], an analogue of the Fréchet distance in fuzzy metric spaces was introduced. That paper considers both monotonic and non-monotonic cases of the fuzzy Fréchet distance $Md(x, y, t)$. In [16], an algorithm was developed for computing the discrete Fréchet distance d_F . In [17], the notion of the discrete fuzzy Fréchet metric Md_F was introduced. That study proved that the Fréchet distance between piecewise linear curves in fuzzy normed spaces is equal to the discrete Fréchet distance between discrete curves formed by their vertices.

Therefore, having the definition of the discrete fuzzy Fréchet distance Md_F , it is necessary to develop a method for computing this distance between polygonal curves.

2 REVIEW OF THE LITERATURE

The authors of present paper have developed metrics for evaluating image quality at the middle level of computer vision.

In [18], the application of classical metrics such as the Hausdorff distance, the Fréchet distance, and their modifications – Gromov-Fréchet [19] and Gromov-Hausdorff distances – was analyzed in the context of image processing tasks. The papers [20, 21] demonstrate applications of the Fréchet distance between trees to image segmentation problems.

A number of studies are dedicated to theoretical aspects of fuzzy metrics. For example, [22] presents an alternative approach to the notion of fuzzy metric spaces. Dušan Repovš, Aleksandr Savchenko, and Mykhailo Zarichnyi introduced a fuzzy metric on the space of probability measures, known as the fuzzy Prokhorov metric [23]. In [24], the authors describe a Hausdorff-type fuzzy metric space defined over the set of non-empty closed subsets. This approach has found applications in computer science.

Several publications are focused on improving algorithms for computing the discrete Fréchet distance. In [25], Omrit Filtse proposed a universal approximate simplification for polygonal curves. Bringmann and Mulzer [26] introduced an α -approximation algorithm for each $\alpha \in [1, n]$ with time complexity $O(n \log n + n^2/\alpha)$. In [27], a more efficient α -approximation algorithm was developed with lower computational complexity. Tim Wylie and Binhai Zhu [28] studied the subset chain matching problem for polygonal curves based on the discrete Fréchet distance and proved that three variants of this problem are NP-complete. In [29], the notion of a mean curve was introduced, consisting of the vertices of input curves. The authors aimed to minimize the maximum discrete Fréchet distance to all input curves. In [30], an improved algorithm was proposed for computing the Fréchet distance between closed curves. In [31], the authors developed a subquadratic algorithm for computing the discrete Fréchet distance between two sequences of points in the plane.

The analysis of the literature confirms that the discrete Fréchet distance is widely used in computer vision to measure the distance between images. However, a significant research gap remains in the development of a fuzzy variant of the discrete Fréchet distance.

3 MATERIALS AND METHODS

Let X be an arbitrary set. A function $d: X \times X \rightarrow R$ is called a metric on X if, for all $x, y, z \in X$, the following conditions are satisfied [32]:

- 1) $d(x, y) \geq 0$ (non-negativity);
- 2) $d(x, y) = 0$ if and only if $x = y$ (identity of indiscernibles);
- 3) $d(x, y) = d(y, x)$ (symmetry);
- 4) $d(x, y) \leq d(x, z) + d(z, y)$ (triangle inequality).

If only conditions (2), (3), and (4) are satisfied, the function is called a pseudometric. If only conditions (1), (2), and (4) are satisfied, the function is called a quasimetric.

Let (X, d) be a metric space. Let $C(X)$ denote the set of parametric curves $\gamma: [0, 1] \rightarrow X$ in the space X .

Define $H(I)$ as the set of non-decreasing continuous self-maps of I that preserve the endpoints.

For two given continuous parametric curves $\gamma_i: I \rightarrow X_i, i = 1, 2$, the Fréchet distance is defined as:

$$d_F(\gamma_1, \gamma_2) = \inf \sup_{t \in I} \{d(\gamma_1(\alpha_1(t)), \gamma_2(\alpha_2(t)))\},$$

where the infimum is taken over all reparameterizations $\alpha_1, \alpha_2 \in H(I)$.

It is known that the function d_F is a metric on the set $C(X)$.

In [17], the notion of a fuzzy Fréchet metric was introduced.

The notion of a fuzzy metric uses the notion of a t -norm, i.e. an associative, commutative, monotone, continuous operation $*: [0, 1] \times [0, 1] \rightarrow [0, 1]$, for which 1 is a neutral element [33].

A function $M: X \times X \times (0; \infty) \rightarrow [0, 1]$ is called a fuzzy metric in the sense of George and Veeramani if the following conditions hold:

- 1) $M(x, y, t) > 0$;
- 2) $M(x, y, t) = 1$ for all $t > 0$ if and only if $x = y$;
- 3) $M(x, y, t) = M(y, x, t)$;
- 4) $M(x, y, t) * M(y, z, t) \leq M(x, z, t + s)$;
- 5) the function $M(x, y, -): [0, \infty) \rightarrow [0, 1]$ is continuous.

Given a metric d on the set X , define the function $M_d: X \times X \times (0; \infty) \rightarrow [0, 1]$:

$$M_d(x, y, t) = \frac{t}{d(x, y) + t}.$$

This function is a fuzzy metric [32].

Let X be a (finite-dimensional) linear space R^n . A fuzzy norm on X is a function

$$N: X \times [0, \infty) \rightarrow [0, 1]$$

satisfying the following conditions [34]:

- 1) $N(x, 0) = 0$;
- 2) $N(x, t) \equiv 1$ if and only if $x = 0$;
- 3) $N(\lambda x, t) = N(x, t/|\lambda|)$ for all $t \geq 0, x \in X$ and $\lambda \neq 0$;
- 4) $N(x + y, t + s) \geq \min\{N(x, t), N(y, s)\}$ for all $x, y \in X, t, s \geq 0$;
- 5) $\lim_{t \rightarrow \infty} N(x, t) = 1$.

Each fuzzy norm N induces a fuzzy metric M on X (under the min t -norm) via:

$$M(x, y, t) = M(x - y, t), x, y \in X, t \in [0, \infty).$$

Moreover, any (classical) norm $\|\cdot\|$ on X induces a fuzzy norm $N_{\|\cdot\|}$ on X defined by the formula:

$$N_{\|\cdot\|}(x, t) = \frac{t}{t + \|x\|}, N_{\|\cdot\|}(0, 0) = 0.$$

This fuzzy norm is called the standard fuzzy norm [35].

In addition to the standard fuzzy norm, article [36] considers a class of fuzzy norms generated by quasiconcave functions.

In the context of image recognition, the case $n = 2$ is of particular importance.

Let us consider two convex regions $O_1, O_2 \subset R^2$. It is known that the Hausdorff distance between O_1 and O_2 is

equal to the Fréchet distance between their boundaries $\partial O_1, \partial O_2$ [37]. One of our results is an extension of this equality to the case of fuzzy norms on $\mathbf{R}^2$ and the fuzzy Fréchet distance.

For functions $\gamma_i(t)$, $i = 1, 2$, and for any $\alpha \in H(t)$, we have: $\inf_{s \in I} M(\gamma_1(\alpha(s), \gamma_2(s), t)) > 0$ and $M_F(\gamma_1, \gamma_2, t) > 0$.

Then, the following theorem holds:

Theorem. The function M_F is a fuzzy metric on the set $C(X)$.

Let us now define a discrete curve in a fuzzy metric space (X, M, t) as a mapping $\gamma: \{1, 2, \dots, n\} \rightarrow X$, where $n \in \mathbf{N}$ [33].

For n , ($n \in \mathbf{N}$), define a sequence of index pairs $(i_1, j_1), (i_2, j_2), \dots, (i_k, j_k)$ of distinct points satisfying the following conditions:

- 1) $(i_1, j_1) = (1, 1)$;
- 2) $(i_k, j_k) = (n, n)$;
- 3) $i_{p+1} \in \{i_p, i_{p+1}\}$;
- 4) $i_{q+1} \in \{i_q, i_{q+1}\}$.

Let the set of all such sequences be denoted by $C(n, m)$.

For discrete curves $\gamma_i: \{1, 2, \dots, n\} \rightarrow X$, $i = 1, 2$, in a fuzzy metric space, we consider a pair:

$$L \in C(n_1, n_2), L = ((i_1, j_1), (i_2, j_2), \dots, (i_k, j_k)), t \in (0; \infty).$$

Define:

$$\|L, t\| = \min\{M(\gamma_1(i_p), \gamma_2(i_p), t) \mid p \in \{1, 2, \dots, k\}\}.$$

Then, the discrete fuzzy Fréchet distance is defined by the formula: $Md_F(\gamma_1, \gamma_2, t) = \max_{L \in C(n_1, n_2)} \|L(t)\|$, which is

defined over the set $dC(X)$ of discrete curves in X .

Based on the introduced metrics – namely, the fuzzy Fréchet metric and the fuzzy discrete Fréchet metric – a method is developed for computing the fuzzy discrete Fréchet distance between arbitrary polygonal curves.

The method consists of the following steps:

- 1) Given two arbitrary polygonal curves $\gamma_1: [0, n] \rightarrow X$, $\gamma_2: [0, m] \rightarrow X$.

For each polygonal curve, define the corresponding sequences:

$$\begin{aligned}\sigma(\gamma_1) &= \{\gamma_1(1), \gamma_1(2), \dots, \gamma_1(n)\}, \\ \sigma(\gamma_2) &= \{\gamma_2(1), \gamma_2(2), \dots, \gamma_2(m)\}.\end{aligned}$$

These sequences satisfy the conditions (1–4) defined previously.

- 2) Define the parameter t via a sequence:

$$\sigma(t) = \{t_1, t_2, \dots, t_q\}.$$

- 3) Construct the set L :

$$L = ((i_1, j_1), (i_2, j_2), \dots, (i_k, j_k)).$$

- 4) Compute the value:

$$\|L, t\| = \min\{M(\gamma_1(i_p), \gamma_2(i_p), t_q \mid p \in \{1, 2, \dots, k\}\}.$$

- 5) Determine the fuzzy discrete Fréchet distance using the formula:

$$Md_F(\gamma_1, \gamma_2, t) = \max_{L \in C(n_1, n_2)} \|L(t)\|.$$

4 EXPERIMENTS

To conduct computational experiments, a method for computing the fuzzy discrete Fréchet distance was implemented in software.

For this purpose, a software module was developed in Java using the OpenCV library. The minimum system requirements are as follows: RAM – 1 GB; HDD – 100 GB; operating system – Windows, Linux, or macOS; programming language – Java 8 or higher; OpenCV library version – 3.0; development environment – IntelliJ IDEA.

The dataset of polygonal curves consists of a set of grayscale images with a resolution of 174×120 pixels. During preprocessing, threshold segmentation is applied to separate the background from the object. In the subsequent step, contours are extracted and approximated using piecewise linear representations.

The threshold t is implemented as a separate variable and, depending on the computed Fréchet distance, the result is displayed in the console in the appropriate format. The threshold t is a dynamic variable and is entered manually upon request by the program.

The pseudocode for computing the fuzzy discrete Fréchet distance Md_F is presented below:

Function $Md_F(\gamma_1, \gamma_2, t)$: real;

input polygonal sequence

$\sigma(\gamma_1) = \{\gamma_1(1), \gamma_1(2), \dots, \gamma_1(n)\}$;

$\sigma(\gamma_2) = \{\gamma_2(1), \gamma_2(2), \dots, \gamma_2(m)\}$;

$\sigma(t) = \{t_1, t_2, \dots, t_q\}$;

function d_F (computation of the discrete Fréchet distance)

begin

if $fd_F < t_i$, then $Md_F = d_F$

else $Md_F = \infty$

end.

Pseudocode for computing the function d_F is provided in [16].

As stated in [16], the computational complexity of calculating the discrete Fréchet distance is:

$$O(d_F) = O(m \cdot n).$$

When computing the fuzzy discrete Fréchet distance, the parameter t is evaluated, which defines the boundary of the Fréchet distance.

Thus, the computational complexity of calculating the fuzzy discrete Fréchet distance is:

$$O(Md_F) = O(Cm \cdot n),$$

where C is a constant.

5 RESULTS

Based on the developed software module for computing the fuzzy discrete Fréchet distance, computational experiments were conducted between polygonal curves.

Table 1 presents the results of calculating the discrete Fréchet distance between the given polygons.

Table 1 – Polygons and the discrete Fréchet distance between them

Image 1	Image 2	Fréchet Distance
		71.84
		130.86
		55.54
		36.06
		69.23

Table 2 presents the results of calculating the fuzzy discrete Fréchet distance for three specified threshold values: t_1 , t_2 and t_3 .

Table 2 – Fuzzy discrete Fréchet distance for the specified threshold values t_1 , t_2 , t_3

Experiment No	Fuzzy Fréchet Distance	Parameter t		
		$t_1=40$	$t_2=70$	$t_3=100$
1	71.84	–	–	+
2	117.27	–	–	–
3	130.86	–	–	–
4	55.54	–	+	–
5	36.06	+	–	–
6	67.18	–	+	–
7	69.23	–	+	–
8	46.52	–	+	–
9	46.04	–	+	–
10	50.21	–	+	–
11	69.23	–	+	–
12	46.40	–	+	–

Continuation of Table 1

13	35.22	+	–	–
14	48.79	–	+	–
15	80.95	–	–	+
16	70.57	–	–	+
17	64.14	–	+	–
18	54.0	–	+	–
19	34.18	+	–	–
20	83.59	–	–	+

6 DISCUSSION

The fuzzy Fréchet metric was developed in [15]. In [16], a method and an efficient algorithm for computing the discrete Fréchet distance with low computational complexity were proposed. In the present work, a method for computing the fuzzy discrete Fréchet distance is developed. Based on this method, an algorithm is constructed for calculating the fuzzy discrete Fréchet distance, which builds upon the classical discrete Fréchet algorithm and also maintains low computational complexity.

The evaluation of distance based on the fuzzy discrete Fréchet metric (FDFM) makes it possible to identify objects that are within a specified threshold t .

Analysis of Table 1 shows that the discrete Fréchet distance effectively estimates image similarity based on contours: the smaller the distance value, the more similar the image contours are, and vice versa.

Table 2 presents the values of the fuzzy discrete Fréchet distance for given thresholds t , which essentially represent the fuzzy variant of the discrete Fréchet distance. Distance evaluation based on the fuzzy discrete Fréchet distance enables the ranking and selection of images whose contour distances are less than or equal to a specified threshold t .

CONCLUSIONS

In computer vision, there exists a fundamental problem of estimating image similarity. To address this challenge, various distance metrics are employed. Therefore, the development of new metrics for measuring image similarity remains a relevant and important research problem. Metrics for evaluating image similarity are typically based on classical distances, such as the Fréchet distance and the Hausdorff distance. The Fréchet distance is commonly used to assess the similarity between image contours, whereas the Hausdorff distance is applied to evaluate the proximity between image regions.

The authors have developed a number of metrics for image distance estimation [18–20]. The current study continues the research presented in [18–20] and builds upon the previously proposed fuzzy Fréchet metrics [17, 19].

The developed method and algorithm for computing the fuzzy discrete Fréchet distance allow the estimation of discrete contour distances under a specified threshold. This can be applied to image clustering and classification tasks.

The **scientific novelty** of the study lies in the development of a method for evaluating the fuzzy discrete Fréchet distance.

The **practical significance** consists in the algorithmic formulation and software implementation of the proposed method, as well as in the performance of computer experiments on a dataset of polygonal curves.

Directions for future research include the application of the fuzzy discrete Fréchet distance for image clustering problems, based on the proposed fuzzy discrete metric.

ACKNOWLEDGEMENTS

The authors of the article express their gratitude to the Department of Pathological Anatomy with a Sectional Course and Forensic Medicine of I. Ya. Horbachevsky Ternopil National Medical University (Ternopil, Ukraine) for providing histopathological images of breast cancer and personally to Doctor of Medical Sciences, Professor, Head of the Department Petro Romanovych Selskyi.

REFERENCES

- Gonzalez R. C., Woods R. E. Digital Image Processing. Fourth Edition, Pearson Education, 2018, 1020 p.
- Ralević N. M., Paunović M. V., Iričanin B. D. Fuzzy metric space and applications in image processing, *Math. Mon.*, 2020, Vol. 48, pp. 103–117. DOI:10.20948/mathmontis-2020-48-9
- Gregori V., Morillas S., Sapena A. Examples of fuzzy metrics and applications, *Fuzzy Sets and Systems*, 2011, Vol. 170, № 1, pp. 95–111. DOI: 10.1016/j.fss.2010.10.019
- Golestaneh P., Zekri M., Sheikholeslam F. Fuzzy wavelet extreme learning machine, *Fuzzy Sets and Systems*, 2018, Vol. 342, pp. 90–108. DOI: 10.1016/j.fss.2017.12.006
- Ralević N. M., Karaklić D., Pištanjat N. Fuzzy metric and its applications in removing the image noise, *Soft Computing*, 2019, Vol. 23, № 22, pp. 12049–12061. DOI: 10.1007/s00500-019-03762-5
- Soto-Hidalgo J. M., Sánchez D., Chamorro-Martínez J., Martínez-Jiménez P. M. Color comparison in fuzzy color spaces, *Fuzzy Sets and Systems*, 2020, Vol. 390, pp. 160–182. DOI: 10.1016/j.fss.2019.09.013
- Vinurajkumar S., Anandhavelu S. An Enhanced Fuzzy Segmentation Framework for extracting white matter from T1-weighted MR images, *Biomedical Signal Processing and Control*, 2022, Vol. 71, P. 103093. DOI: 10.1016/j.bspc.2021.103093
- Bloch I. Fuzzy skeleton and skeleton by influence zones: a review. In *Skeletonization: Theory, Methods and Applications*. Elsevier, Academic Press, 2017, pp. 71–87. DOI: 10.1016/B978-0-08-101291-8.00004-3

9. Soto J., Flores-Sintas A., Palarea-Albaladejo J. Improving probabilities in a fuzzy clustering partition, *Fuzzy Sets and Systems*, 2008, Vol. 159, № 4, pp. 406–421. DOI: 10.1016/j.fss.2007.08.016
10. Campomanes-Álvarez C., Campomanes-Álvarez B. R., Guadarrama S., Ibáñez O., Cerdón O. An experimental study on fuzzy distances for skull-face overlay in craniofacial superimposition, *Fuzzy Sets and Systems*, 2017, Vol. 318, pp. 100–119. DOI: 10.1016/j.fss.2016.06.015
11. Leski J.M. Fuzzy c-ordered-means clustering, *Fuzzy Sets and Systems*, 2016, Vol. 286, pp. 114–133. DOI: 10.1016/j.fss.2014.12.007
12. Saha I., Sarkar J. P., Maulik U. Integrated Rough Fuzzy Clustering for Categorical Data Analysis, *Fuzzy Sets and Systems*, 2019, Vol. 361, № 15, pp. 1–32. DOI: 10.1016/j.fss.2018.02.007
13. Bandyopadhyay S., Saha S., Pedrycz W. Use of a fuzzy granulation-degranulation criterion for assessing cluster validity, *Fuzzy Sets and Systems*, 2011, Vol. 170, № 1, pp. 22–42. DOI: 10.1016/j.fss.2010.11.015
14. Das A., Mohapatra S. K., Mohanty M. N. Design of deep ensemble classifier with fuzzy decision method for biomedical image classification, *Applied Soft Computing*, 2022, Vol. 115, P. 108178. DOI: 10.1016/j.asoc.2021.108178
15. Bazylevych L., Berezsky O., Zarichnyi M. Fréchet fuzzy metric, *Matematychni Studii*, 2022, Vol. 57, No. 2, pp. 210–215. DOI: 10.30970/ms.57.2.210-215.
16. Eiter, T.; Mannila, H. Computing Discrete Fréchet Distance. Tech. Report CD-TR 94/64, Christian Doppler Laboratory for Expert Systems, TU Vienna, Austria, 1994.
17. Berezkyi M., Berezsky O., Zarichnyi M. Discrete Fréchet fuzzy metric, *Fuzzy Sets and Systems*, 2025. (submitted).
18. Berezsky O., Zarichnyi M. In: Shakhovska N., Medyk-ovskyy M. O. (eds) Metric Methods in Computer Vision and Pattern Recognition, *Advances in Intelligent Systems and Computing V. CSIT 2020. Advances in Intelligent Systems and Computing*, V. 1293. Springer, Cham. DOI: 10.1007/978-3-030-63270-0_13.
19. Berezsky O., Zarichnyi M. Fréchet distance between weighted rooted trees, *Matematychni Studii*, 2017, Vol. 48, No.2, pp. 165–170. DOI: 10.15330/ms.48.2.165-170
20. Berezsky O., Zarichnyi M. Gromov-Fréchet distance between curves, *Matematychni Studii*, 2018, Vol. 50, No. 1, pp. 88–92.
21. Berezsky O., Zarichnyi M., Pitsun O. Development of a metric and the methods for quantitative estimation of the segmentation of biomedical images, *Eastern-European Journal of Enterprise Technologies*, 2017, Vol. 6, № 4, pp. 4–11. DOI: 10.15587/1729-4061.2017.119493
22. Sostak A. George-Veeramani Fuzzy Metrics Revised, *Axioms*, 2018, Vol. 7, № 3, P. 60. DOI: 10.3390/axioms7030060
23. Repovš D., Savchenko A., Zarichnyi M. Fuzzy Prokhorov metric on the set of probability measures, *Fuzzy Sets and Systems*, 2011, Vol. 175, № 1, pp. 96–104. DOI: 10.1016/j.fss.2011.02.014
24. Gutiérrez García J., Romaguera S., Sanchis M. An identification theorem for the completion of the Hausdorff fuzzy metric, *Fuzzy Sets and Systems*, 2013, Vol. 227, pp. 96–106. DOI: 10.1016/j.fss.2013.04.012
25. Filtser O. Universal Approximate Simplification Under the Discrete Fréchet Distance, *Information Processing Letters*. – 2018, Vol. 132, pp. 22–27. DOI: 10.1016/j.ipl.2017.10.002
26. Bringmann K., Mulzer W. Approximability of the discrete Fréchet distance, *Journal of Computational Geometry*, 2016, Vol. 7, № 2, pp. 46–76. DOI: 10.20382/jocg.v7i2a4
27. Chan T. M., Rahmati Z. An Improved Approximation Algorithm for the Discrete Fréchet Distance, *Information Processing Letters*, 2018, Vol. 138, pp. 72–74. DOI: 10.1016/j.ipl.2018.06.011
28. Wylie T., Zhu B. Following a Curve with the Discrete Fréchet Distance, *Theoretical Computer Science*, 2014, Vol. 556, pp. 34–44. DOI: 10.1016/j.tcs.2014.06.026
29. Ahn H.-K., Alt H., Buchin M., Oh E., Scharf L., Wenk C. Middle Curves Based on Discrete Fréchet Distance, *Computational Geometry*, 2020, Vol. 89, P. 101621. DOI: 10.1016/j.comgeo.2020.101621
30. Vodolazskiy E. Discrete Fréchet Distance for Closed Curves, *Computational Geometry*, 2023, Vol. 111, P. 101967. DOI: 10.1016/j.comgeo.2022.101967
31. Agarwal P.K., Ben Avraham R., Kaplan H., Sharir M. Computing the Discrete Fréchet Distance in Subquadratic Time, *SIAM Journal on Computing*, 2014, Vol. 43, No. 2, pp. 429–449. DOI: 10.1137/130920526
32. Deza M. M., Deza E. Encyclopedia of Distances. Springer-Verlag Berlin Heidelberg, 2013, 583 p.
33. Schweizer B., Sklar A. Statistical metric spaces, *Pacific Journal of Mathematics*, 1960, Vol. 10, pp. 313–334. DOI: 10.2140/pjm.1960.10.313
34. Bag T., Samanta S. K. Finite dimensional fuzzy normed linear spaces, *Journal of Fuzzy Mathematics*, 2003, Vol. 11, pp. 687–705.
35. Font J. J., Galindo J., Macario S., Sanchis M. Mazur-Ulam type theorems for fuzzy normed spaces, *Journal of Nonlinear Sciences and Applications*, 2017, Vol. 10, № 8, pp. 4499–4506. DOI: 10.22436/jnsa.010.08.41
36. Sánchez J. C., González D. M. A natural correspondence between quasiconcave functions and fuzzy norms, *Fuzzy Sets and Systems*, 2023, Vol. 466, P. 108413. DOI: 10.1016/j.fss.2022.10.005
37. Godau M. On the complexity of measuring the similarity between geometric objects in higher dimensions. PhD thesis, Freie Universität Berlin, 1999. URL: <https://doi.org/10.17169/refubium-7780>

Received 18.06.2025.
Accepted 25.09.2025.

МЕТОД ВИЗНАЧЕННЯ НЕЧІТКОЇ ДИСКРЕТНОЇ ВІДСТАНІ ФРЕШЕ

Березький О. М. – д-р техн. наук, професор, професор кафедри комп'ютерної інженерії Західноукраїнського національного університету, Тернопіль, Україна.

Березький М. О. – аспірант кафедри комп'ютерної інженерії Західноукраїнського національного університету, Тернопіль, Україна.

Зарічний М. М. – д-р фіз.-мат. наук, професор, професор кафедри алгебри, топології та основ математики Львівського національного університету імені Івана Франка, Львів, Україна.

АНОТАЦІЯ

Актуальність. У статті розглянуто проблему оцінки подібності зображень на основі метрики Фреше та її модифікацій. При цьому зображення апроксимовані полігональними кривими. Проблема обумовлена необхідністю кількісно оцінювати подібність зображень для задач генерування, кластеризації та розпізнавання зображень. Кількісна оцінка близькості біомедичних зображень допомагає підтримувати прийняття рішення в системах автоматичного діагностування. Об'єктом дослідження є процес оцінки подібності зображень. Предметом дослідження є метрика Фреше та її модифікації.

Мета роботи – розробити метод визначення нечіткої дискретної відстані Фреше, оцінити обчислювальну складність розробленого методу, програмно реалізувати алгоритм методу визначення нечіткої дискретної відстані Фреше та провести комп'ютерні експерименти з визначення нечіткої дискретної відстані Фреше між полігонами.

Метод. У статті розроблено метод визначення дискретної нечіткої відстані Фреше, що базується на нечіткій метриці Фреше між полігональними кривими. В основі нечіткої метрики Фреше лежить метрика Фреше на просторі параметризованих кривих. Потрібна для застосувань апроксимація задається процедурою дискретизації нечіткої метрики Фреше. Розроблений метод оцінює дискретну нечітку відстань Фреше між полігональними кривими на основі алгоритму обчислення дискретної відстані Фреше.

Результати. Комп'ютерні експерименти проводились на наборі заданих областей, апроксимованих полігональними кривими. На основі запропонованого методу розроблено алгоритм оцінки дискретної нечіткої відстані Фреше. Розроблений алгоритм має низьку обчислювальну складність, яка рівна добутку дискретизованих відрізків полігональних кривих: $O(Cm \cdot n)$. Це дало змогу оцінювати дискретну відстань Фреше із заданим порогом подібності. Програмна реалізація методу буде інтегрована в систему автоматичного медичного діагностування.

Висновки. Отримані результати дослідження дозволяють рекомендувати розроблений метод оцінки близькості зображень на основі нечіткої дискретної відстані Фреше в системах комп'ютерного зору для широкого застосування: генерування, кластеризації та розпізнавання зображень.

КЛЮЧОВІ СЛОВА: метрика, нечітка метрика Фреше, дискретна нечітка метрика Фреше, зображення, полігональні криві.

ЛІТЕРАТУРА

1. Gonzalez R. C. Digital Image Processing / R. C. Gonzalez, R. E. Woods. – Fourth Edition, Pearson Education, 2018. – 1020 p.
2. Ralević N. M. Fuzzy metric space and applications in image processing / N. M. Ralević, M. V. Paunović, B. D. Iričanin // Math. Mon. – 2020. – Vol. 48. – P. 103–117. DOI:10.20948/mathmontis-2020-48-9
3. Gregori V. Examples of fuzzy metrics and applications / V. Gregori, S. Morillas, A. Sapena // Fuzzy Sets and Systems. – 2011. – Vol. 170, № 1. – P. 95–111. DOI: 10.1016/j.fss.2010.10.019
4. Golestaneh P. Fuzzy wavelet extreme learning machine / P. Golestaneh, M. Zekri, F. Sheikholeslam // Fuzzy Sets and Systems. – 2018. – Vol. 342. – P. 90–108. DOI: 10.1016/j.fss.2017.12.006
5. Ralević N. M. Fuzzy metric and its applications in removing the image noise / N. M. Ralević, D. Karaklić, N. Pištinjat // Soft Computing – 2019. – Vol. 23, № 22. – P. 12049–12061. DOI: 10.1007/s00500-019-03762-5
6. Color comparison in fuzzy color spaces / [J. M. Soto-Hidalgo, D. Sánchez, J. Chamorro-Martínez, P. M. Martínez-Jiménez] // Fuzzy Sets and Systems. – 2020. – Vol. 390. – P. 160–182. DOI: 10.1016/j.fss.2019.09.013
7. Vinurajkumar S. An Enhanced Fuzzy Segmentation Framework for extracting white matter from T1-weighted MR images / S. Vinurajkumar, S. Anandhavelu // Biomedical Signal Processing and Control. – 2022. – Vol. 71. – P. 103093. DOI: 10.1016/j.bspc.2021.103093
8. Bloch I. Fuzzy skeleton and skeleton by influence zones: a review / I. Bloch. – In Skeletonization: Theory, Methods and Applications. Elsevier – Academic Press, 2017. – P. 71–87. DOI: 10.1016/B978-0-08-101291-8.00004-3
9. Soto J. Improving probabilities in a fuzzy clustering partition / J. Soto, A. Flores-Sintas, J. Palarea-Albaladejo // Fuzzy Sets and Systems. – 2008. – Vol. 159, № 4. – P. 406–421. DOI: 10.1016/j.fss.2007.08.016
10. An experimental study on fuzzy distances for skull-face overlay in craniofacial superimposition / [C. Campomanes-Álvarez, B. R. Campomanes-Álvarez, S. Guadarrama, O. Ibáñez, O. Córdón] // Fuzzy Sets and Systems. – 2017. – Vol. 318. – P. 100–119. DOI: 10.1016/j.fss.2016.06.015
11. Leski J.M. Fuzzy c-ordered-means clustering / J.M. Leski // Fuzzy Sets and Systems. – 2016. – Vol. 286. – P. 114–133. DOI: 10.1016/j.fss.2014.12.007
12. Saha I. Integrated Rough Fuzzy Clustering for Categorical Data Analysis / I. Saha, J. P. Sarkar, U. Maulik // Fuzzy Sets and Systems. – 2019. – Vol. 361, №15. – P. 1–32. DOI: 10.1016/j.fss.2018.02.007
13. Bandyopadhyay S. Use of a fuzzy granulation-degranulation criterion for assessing cluster validity / S. Bandyopadhyay, S. Saha, W. Pedrycz // Fuzzy Sets and Systems. – 2011. – Vol. 170, № 1. – P. 22–42. DOI: 10.1016/j.fss.2010.11.015
14. Das A. Design of deep ensemble classifier with fuzzy decision method for biomedical image classification / A. Das, S. K. Mohapatra, M. N. Mohanty // Applied Soft Comput.

- ing. – 2022. – Vol. 115. – P. 108178. DOI: 10.1016/j.asoc.2021.108178
15. Bazylevych L. Fréchet fuzzy metric / L. Bazylevych, O. Berezsky, M. Zarichnyi // *Matematychni Studii*. – 2022. – Vol. 57, No. 2. – P. 210–215. DOI: 10.30970/ms.57.2.210-215.
16. Eiter T. Computing Discrete Fréchet Distance. Tech. Report CD-TR 94/64, Christian Doppler Laboratory for Expert Systems / T. Eiter, H. Mannila. – TU Vienna, Austria, 1994.
17. Berezkyi M. Discrete Fréchet fuzzy metric / M. Berezkyi, O. Berezsky, M. Zarichnyi // *Fuzzy Sets and Systems*. – 2025. (submitted).
18. Berezsky O. Metric Methods in Computer Vision and Pattern Recognition / O. Berezsky, M. Zarichnyi – In: Shakhovska N., Medykovsky M. O. (eds) // *Advances in Intelligent Systems and Computing V. CSIT 2020. Advances in Intelligent Systems and Computing*. – V. 1293. – Springer, Cham. DOI: 10.1007/978-3-030-63270-0_13.
19. Berezsky O. Fréchet distance between weighted rooted trees / O. Berezsky, M. Zarichnyi // *Matematychni Studii*. – 2017. – Vol. 48, No. 2. – P. 165–170. DOI: 10.15330/ms.48.2.165-170
20. Berezsky O. Gromov-Fréchet distance between curves / O. Berezsky, M. Zarichnyi // *Matematychni Studii*. – 2018. – Vol. 50, No. 1. – P. 88–92.
21. Berezsky O. Development of a metric and the methods for quantitative estimation of the segmentation of biomedical images / O. Berezsky, M. Zarichnyi, O. Pitsun // *Eastern-European Journal of Enterprise Technologies*. – 2017. – Vol. 6, № 4. – P. 4–11. DOI: 10.15587/1729-4061.2017.119493
22. Šostak A. George-Veeramani Fuzzy Metrics Revised / A. Šostak // *Axioms*. – 2018. – Vol. 7, № 3. – P. 60. DOI: 10.3390/axioms7030060
23. Repovš D. Fuzzy Prokhorov metric on the set of probability measures / D. Repovš, A. Savchenko, M. Zarichnyi // *Fuzzy Sets and Systems*. – 2011. – Vol. 175, № 1. – P. 96–104. DOI: 10.1016/j.fss.2011.02.014
24. Gutiérrez García J. An identification theorem for the completion of the Hausdorff fuzzy metric / J. Gutiérrez García, S. Romaguera, M. Sanchis // *Fuzzy Sets and Systems*. – 2013. – Vol. 227. – P. 96–106. DOI: 10.1016/j.fss.2013.04.012
25. Filtser O. Universal Approximate Simplification Under the Discrete Fréchet Distance / O. Filtser // *Information Processing Letters*. – 2018. – Vol. 132. – P. 22–27. DOI: 10.1016/j.ipl.2017.10.002
26. Bringmann K. Approximability of the discrete Fréchet distance / K. Bringmann, W. Mulzer // *Journal of Computational Geometry*. – 2016. – Vol. 7, № 2. – P. 46–76. DOI: 10.20382/jocg.v7i2a4
27. Chan T. M. An Improved Approximation Algorithm for the Discrete Fréchet Distance / T. M. Chan, Z. Rahmati // *Information Processing Letters*. – 2018. – Vol. 138. – P. 72–74. DOI: 10.1016/j.ipl.2018.06.011
28. Wylie T. Following a Curve with the Discrete Fréchet Distance / T. Wylie, B. Zhu // *Theoretical Computer Science*. – 2014. – Vol. 556. – P. 34–44. DOI: 10.1016/j.tcs.2014.06.026
29. Middle Curves Based on Discrete Fréchet Distance / [H.-K. Ahn, H. Alt, M. Buchin et al.] // *Computational Geometry*. – 2020. – Vol. 89. – P. 101621. DOI: 10.1016/j.comgeo.2020.101621
30. Vodolazskiy E. Discrete Fréchet Distance for Closed Curves / E. Vodolazskiy // *Computational Geometry*. – 2023. – Vol. 111. – P. 101967. DOI: 10.1016/j.comgeo.2022.101967
31. Computing the Discrete Fréchet Distance in Subquadratic Time / [P. K. Agarwal, R. Ben Avraham, H. Kaplan, M. Sharir] // *SIAM Journal on Computing*. – 2014. – Vol. 43, No. 2. – P. 429–449. DOI: 10.1137/130920526
32. Deza M. M. Encyclopedia of Distances / M. M. Deza, E. Deza. – Springer-Verlag Berlin Heidelberg, 2013. – 583 p.
33. Schweizer B. Statistical metric spaces / B. Schweizer, A. Sklar // *Pacific Journal of Mathematics*. – 1960. – Vol. 10. – P. 313–334. DOI: 10.2140/pjm.1960.10.313
34. Bag T. Finite dimensional fuzzy normed linear spaces / T. Bag, S. K. Samanta // *Journal of Fuzzy Mathematics*. – 2003. – Vol. 11. – P. 687–705.
35. Mazur-Ulam type theorems for fuzzy normed spaces / [J. J. Font, J. Galindo, S. Macario, M. Sanchís] // *Journal of Nonlinear Sciences and Applications*. – 2017. – Vol. 10, № 8. – P. 4499–4506. DOI: 10.22436/jnsa.010.08.41
36. Sánchez J. C. A natural correspondence between quasiconcave functions and fuzzy norms / J. C. Sánchez, D. M. González // *Fuzzy Sets and Systems*. – 2023. – Vol. 466. – P. 108413. DOI: 10.1016/j.fss.2022.10.005
37. Godau M. On the complexity of measuring the similarity between geometric objects in higher dimensions / M. Godau. – PhD thesis, Freie Universität Berlin, 1999. URL: <https://doi.org/10.17169/refubium-7780>

ABOUT SPECIAL CASES OF LAGRANGIAN INTERSTRIPATION OF APPROXIMATIONS OF FUNCTIONS OF TWO VARIABLES

Slavik O. – PhD, Assistant at the Department of Applied Mathematics, Kharkiv National University of Radio Electronics, Kharkiv, Ukraine; Associate Professor at the Department of Information Computer Technologies and Mathematics, Educational and Research Institute “Ukrainian Engineering and Pedagogical Academy” of V. N. Karazin Kharkiv National University, Kharkiv, Ukraine.

ABSTRACT

Context. The problem of approximating the values of continuous functions of two variables based on known information about them on stripes, the boundaries of which are parallel to the coordinate axes, is considered. The object of the study is the process of approximating the values of functions based on incomplete information about them, which is given on the system of stripes.

Objective. The goal of the work is the review of information operators of Lagrangian interstripation and features of the construction of information approximation operators for some cases of the mutual arrangement of stripes in some region, which allow to significantly simplify the calculation of approximate values of the function in unknown subregions of the region.

Method. Methods for approximating the values of continuous functions of two variables with incomplete information about them on some limited area are proposed. Information about the function is known only on a system of stripes limited by straight lines parallel to the coordinate axes. A method for approximating the values of continuous functions of two variables, information about which is known on two stripes, as a result of union of which only some rectangular subregion remains unknown in the region, is proposed. A method for approximating the values of continuous functions of two variables, information about which is known on three stripes, as a result of union of which only some rectangular subregion remains unknown in the region, is proposed. A method for approximating the values of continuous functions of two variables, information about which is known on four stripes, as a result of union of which only some rectangular subregion remains unknown in the region, is proposed. For all the considered cases, approximation operators are given that allow calculating the approximate form of the function in the unknown subregions in the analytical form.

Results. The information operators of Lagrangian interstripation are implemented programmatically and investigated in problems of approximating the values of functions of two variables from known information about them on the systems of stripes.

Conclusions. The experiments confirmed the accuracy of approximation of the values of continuous functions of two variables of the proposed information interstripation operators for different systems of stripes. Approximation operators are given for special cases of the location of stripes in the region, the difference of which from the information interstripation operators of the general form lies in the significant simplification of the approximation operators without losing the accuracy of the approximation with a smaller number of arithmetic operations, which can be a decisive factor in some cases. Prospects for further research lie in the application of the proposed information operators in the problems of digital image processing, seismic mineral exploration data and remote sensing data etc.

KEYWORDS: numerical methods, mathematical modeling, information operators, interlination, interstripation.

NOMENCLATURE

$f(x, y)$ is a function of two variables;

D is a rectangular region;

$S_{x,k}$ is a known data stripe with restrictions on the variable x , k – serial number of the strip;

$S_{y,p}$ is a known data stripe with restrictions on the variable y , p – serial number of the strip;

$\bar{S}_{x,k,k+1}$ is an unknown data stripe with restrictions on the variable x , which is located between k and $k + 1$ strips;

$\bar{S}_{y,p,p+1}$ is an unknown data stripe with restrictions on the variable y , which is located between p and $p + 1$ strips;

$\{x_k\}_{k=1}^n$ is a sequence of indexed variables x from 1 to n ;

$f(x, y)|_S$ is a trace of the function $f(x, y)$ on the stripe S ;

$f(x)|_\Gamma$ is a trace of the function $f(x, y)$ on the line Γ , which is a function of one variable x ;

$f(y)|_\Gamma$ is a trace of the function $f(x, y)$ on the line Γ , which is a function of one variable y ;

$Of(x, y)$ is an interstripation operator;

$Lf(x, y)$ is an approximation operator.

INTRODUCTION

Methods of approximation of functions with incomplete information are an important part of modern applied mathematics and are widely used in various fields of science and technology. Their main goal is to restore or simulate a function for which only a limited number of values are known, or these values contain noise or errors. Such situations arise in real conditions, when a complete description of an object or process is not available due to technical, financial or physical constraints.

One of the key areas of application of these methods is signal and image processing. For example, in the compression or restoration of digital images, approximation methods are used to obtain the most accurate reconstruction from partial data. In machine learning, function approximation is the basis for building models that predict or classify data based on incomplete or limited samples.

In natural and technical sciences, approximation methods help solve inverse problems – for example, in medical tomography or geophysical research, where it is necessary to reconstruct the internal structure of an object from indirect measurements. In economics, these methods are used to forecast market indicators, estimate demand or production functions based on incomplete statistical data.

Due to their versatility and ability to work with imprecise or incomplete data, function approximation methods have become a powerful tool in conditions of uncertainty, opening up new possibilities for analysis, modeling, and decision-making in a wide variety of applied problems.

The object of study is the process of approximating functions based on incomplete information about them, given on a system of stripes.

The subject of study is mathematical methods and methods of information operators of O. M. Lytvyn, which allow building approximate models of functions based on partial data.

The purpose of the work is to analyze the main methods of approximating functions in cases of incomplete information, to investigate their effectiveness and areas of application, and to justify the choice of approaches for practical use in conditions of limited data.

1 PROBLEM STATEMENT

Let $f(x, y)$ be a function of two variables defined on some rectangular region $D = [a, b] \times [c, d]$, called the interstripation region. Suppose that the values of this function are known exactly on some system of stripes:

$$S_k = \{(x, y) : \alpha_k \leq \omega_k(x, y) \leq \beta_k\}, \quad k \in \mathbb{N}.$$

Let's assume that the boundaries of the stripes are straight lines parallel to the coordinate axes. In this case, the stripe system can be written in a simplified form if the stripe boundaries are parallel to each other and parallel to the axis Oy:

$$S_{x,k} = \{(x, y) : a_k \leq x \leq b_k\}, \quad k = \overline{1, n}$$

or, if the boundaries of the stripes are parallel to each other and parallel to the Ox axis:

$$S_{y,p} = \{(x, y) : c_p \leq y \leq d_p\}, \quad p = \overline{1, m}.$$

Let's consider the following sequences of values, which contain the coordinates of the boundaries of the

interstripation region and the boundaries of the stripes of the form:

$$\begin{aligned} \{x_k\}_{k=1}^{2n} : a_1 < b_1 \leq a_2 < b_2 \leq \dots \leq a_{n-1} < b_{n-1} \leq a_n < b_n, \\ \{y_p\}_{p=1}^{2m} : c_1 < d_1 \leq c_2 < d_2 \leq \dots \leq c_{m-1} < d_{m-1} \leq c_m < d_m. \end{aligned}$$

Therefore, at points in the region D that do not belong to any of the above stripes $S_{x,k}$ or $S_{y,p}$ the information about the function is unknown. Let us denote such regions by:

$$\begin{aligned} \bar{S}_{x,k,k+1} &= \{(x, y) : b_k < x < a_{k+1}\}, \quad k = \overline{1, n-1}, \\ \bar{S}_{y,p,p+1} &= \{(x, y) : d_p < y < c_{p+1}\}, \quad p = \overline{1, m-1}. \end{aligned}$$

The problem is to approximate the values of the function in those parts of the region D where information about it is unknown.

2 REVIEW OF THE LITERATURE

The problem of function approximation with incomplete, sparse, or noisy data is relevant in many fields of science and technology, therefore its research is interdisciplinary in nature. These may be, for example, values of a function at individual points, approximate observations, or partial information about derivatives. In such cases, various approximation methods are used, which can be classified according to analytical, statistical, or numerical approaches [1–2].

The first systematic approaches to function approximation were developed within the framework of classical analysis and numerical mathematics. A significant contribution to the development of this field was made by such scientists as Weierstrass, Chebyshev, Hilbert, Gauss and others. The Weierstrass and Stone-Weierstrass theorems became the basis for the formal justification of the possibility of approximating continuous functions by polynomials. Chebyshev's work on optimal approximations was also of great importance, especially for the construction of uniform approximations [3–4].

One of the classical methods is interpolation, in which a function is constructed that passes exactly through given points. In the simplest case, polynomial interpolation is used: the function is approximated by a polynomial that passes through all given values. Among the popular forms are Lagrange's interpolation formula and Newton's formula. However, as the number of points increases, these methods become unstable: the Runge effect appears, when the polynomial fluctuates strongly between nodes [1], [5].

To avoid these problems, splines are used – piecewise polynomial functions that combine local approximations with smoothness conditions. The most common is the cubic spline, which ensures the continuity of the function and its derivatives up to the second order. Splines work well for large sets of points, allowing you to preserve both accuracy and smoothness [5–6].

In the case when the data are noisy, it is more expedient to use approximation methods, in particular the least squares method. In this case, it is not required to pass through all points exactly, but only to minimize the mean square error. This provides a smooth approximation that is less sensitive to random errors in the data. Regression methods – both linear and nonlinear – are also used to build approximating models that generalize the behavior of the function [7–8].

In many applied problems, function recovery is accompanied by instability problems, when a small change in the input data can cause large deviations in the result. For such cases, regularization is used – the introduction of additional restrictions or penalty terms into the functional being minimized. The most famous method is Tikhonov regularization, which provides a compromise between accuracy and stability of the solution [9–11].

When not only a sample of function values but also some a priori knowledge about its properties is available, statistical and probabilistic methods, in particular Gaussian process models, are effective. In this approach, the function is considered as a random process for which not only the expected value but also an estimate of the error can be calculated. This is especially important in conditions of uncertainty or data limitations [12].

In recent years, machine learning methods have been actively developed, demonstrating high efficiency in the approximation of complex functions. In particular, artificial neural networks are able to approximate arbitrary continuous functions by learning from a large number of examples. They are extremely flexible, but require significant computational resources and careful tuning [13].

Among modern research, it is also worth noting works devoted to compressed sensing, which are based on the assertion that a function can be accurately reconstructed even from a much smaller number of measurements than required by traditional methods, provided that it has a certain “sparseness” or structure. This direction has been actively developing since the early 2000s [10].

Also, in cases where only general properties of a function are known (for example, monotonicity, convexity, or boundedness), optimization methods are used to find a function that best matches the available information and satisfies the given constraints. In such problems, the approximation of a function is formulated as a problem of minimizing the functional considering structural requirements [14].

A review of the literature shows that the choice of approximation method depends on the nature of the available information about the function: the amount of data, the presence of noise, requirements for accuracy, smoothness or stability of the solution. Modern approaches increasingly combine classical mathematical and modern machine learning tools to achieve maximum accuracy under conditions of limited resources and data.

A relatively new direction in information operators are information operators proposed by O. M. Lytvyn [15–16]. The proposed information operators allow using informa-

tion about the object of study not only on a system of points, but also on lines, stripes and flats.

Such information operators are used for a wide class of applied problems in various fields of science and technology.

3 MATERIALS AND METHODS

Let's introduce the concept of the trace of a function on a strip and on a line.

The trace of a function $f(x, y)$ on a strip S is called a function of two variables $f(x, y)|_S$, which at each point of this stripe takes the same values as the function $f(x, y)$, and outside the stripe is unknown.

The trace of a function $f(x, y)$ on a line Γ is a function of one variable $f(x)|_\Gamma$ or $f(y)|_\Gamma$, which at each point of this line takes the same values as the function $f(x, y)$, and outside the line is unknown.

The interstripation of a continuous function of two variables $f(x, y)$ on a system of stripes S_k , $k \in \mathbb{N}$ is called its recovery using its traces on the strips $f(x, y)|_{S_k}$, $k \in \mathbb{N}$.

The general form of the interstripation operator, which on each stripe returns the value of the function in the form of a trace of the function on the stripe, and between stripes approximates the function according to the known information about it on the stripe system, has the form:

$$Of(x, y) = \begin{cases} f(x, y)|_{S_{x,k}}, & (x, y) \in S_{x,k}, \\ f(x, y)|_{S_{y,p}}, & (x, y) \in S_{y,p}, \\ Lf(x, y), & (x, y) \in D \setminus (S_{x,k} \cup S_{y,p}), \end{cases}$$

$$k = \overline{1, n}, \quad p = \overline{1, m}.$$

The most general case is when the information about the function is known on a system of stripes whose boundaries are mutually perpendicular to each other. That is, the traces of the function are known on the stripes $S_{x,k}$ whose boundaries are parallel to the axis Oy and on the strips $S_{y,p}$ whose boundaries are parallel to the axis Ox . The union of the information from all stripes forms a system of intersecting stripes with rectangular regions where information about the function is unknown:

$$\bar{S}_{k,k+1,p,p+1} = \left\{ (x, y) : (x, y) \in (b_k, a_{k+1}) \times (d_p, c_{p+1}) \right\},$$

$$k = \overline{1, n-1}, \quad p = \overline{1, m-1}.$$

To approximate the values of the function using information from such a system of stripes, it is proposed to use an interstripation operator in the Lagrange form:

$$Lf(x, y) = \sum_{k=1}^{2n} f(x_k, y) \prod_{i=1, i \neq k}^{2n} \frac{x - x_i}{x_k - x_i} + \\ + \sum_{p=1}^{2m} f(x, y_p) \prod_{j=1, j \neq p}^{2m} \frac{y - y_j}{y_p - y_j} - \\ - \sum_{k=1}^{2n} \sum_{p=1}^{2m} f(x_k, y_p) \prod_{i=1, i \neq k}^{2n} \frac{x - x_i}{x_k - x_i} \prod_{j=1, j \neq p}^{2m} \frac{y - y_j}{y_p - y_j}.$$

Properties of the interstripation operator for approximating function values on a system of stripes whose boundaries are mutually perpendicular and parallel to the coordinate axes:

$$Of(x, y)|_{S_{x,k}} = f(x, y)|_{S_{x,k}}, \quad k = \overline{1, n}, \\ Of(x, y)|_{S_{y,p}} = f(x, y)|_{S_{y,p}}, \quad p = \overline{1, m}.$$

In case if $b_k = a_{k+1}$, $k = \overline{1, n-1}$ and $d_p = c_{p+1}$, $p = \overline{1, m-1}$, then the interstripation operator will exactly approximate the function using the given traces on each stripe $S_{x,k}$, $S_{y,p}$.

This approximation can be used to arbitrarily divide a region into stripes whose boundaries are parallel to the coordinate axes. However, for some special cases, the form of the information interstripation operator can be significantly simplified.

Let's assume that the information about the function $f(x, y)$ is given by its traces on two stripes $S_{x,1}$ and $S_{y,1}$. Moreover, it is known that for each of the stripes one of the boundaries whose belong to the interstripation region D , and the other on the boundary of the interstripation region ∂D .

As a result of the union of the information about the function from both stripes, we obtain some rectangular region in D , where the information about the function is unknown. The operator $Lf(x, y)$ in the interstripation operator in this case can be simplified to the following form:

$$Lf(x, y) = f(y)|_{\Gamma_1} + f(x)|_{\Gamma_2} - f(\tilde{x}, \tilde{y}), \\ \Gamma_1 : x = \tilde{x}, \Gamma_2 : y = \tilde{y}, \\ \tilde{x} = \begin{cases} a_1, a_1 \notin \partial D, \\ b_1, b_1 \notin \partial D, \end{cases} \\ \tilde{y} = \begin{cases} c_1, c_1 \notin \partial D, \\ d_1, d_1 \notin \partial D. \end{cases}$$

Let's assume that the information about the function $f(x, y)$ is given by its traces on three stripes $S_{x,1}$, $S_{x,2}$ and $S_{y,1}$. Moreover, as in the previous case, it is known that for each of the strips one of the boundaries whose

belong to the interstripation region D , and the other on the boundary of the interstripation region ∂D .

As a result of the union of the information about the function from the three stripes, we obtain some rectangular region in D , where the information about the function is unknown. The operator $Lf(x, y)$ in the interstripation operator in this case can be simplified to the following form:

$$Lf(x, y) = \frac{x - \tilde{x}_2}{\tilde{x}_1 - \tilde{x}_2} \left(f(y)|_{\Gamma_1} - f(\tilde{x}_1, \tilde{y}) \right) + \\ + \frac{x - \tilde{x}_1}{\tilde{x}_2 - \tilde{x}_1} \left(f(y)|_{\Gamma_2} - f(\tilde{x}_2, \tilde{y}) \right) + f(x)|_{\Gamma_3}, \\ \Gamma_1 : x = \tilde{x}_1, \Gamma_2 : x = \tilde{x}_2, \Gamma_3 : y = \tilde{y}, \\ \tilde{x}_1 = \begin{cases} a_1, a_1 \notin \partial D, \\ b_1, b_1 \notin \partial D, \end{cases} \\ \tilde{x}_2 = \begin{cases} a_2, a_2 \notin \partial D, \\ b_2, b_2 \notin \partial D, \end{cases} \\ \tilde{y} = \begin{cases} c_1, c_1 \notin \partial D, \\ d_1, d_1 \notin \partial D. \end{cases}$$

Note that a similar operator $Lf(x, y)$ can be obtained for the stripes $S_{x,1}$, $S_{y,1}$ and $S_{y,2}$, for which similar conditions on the location of the stripe boundaries in the interstripation region D will be satisfied.

Let's assume that the information about the function $f(x, y)$ is given by its traces on the four stripes $S_{x,1}$, $S_{x,2}$, $S_{y,1}$ and $S_{y,2}$. Moreover, as in the previous case, it is known that for each of the stripes one of the boundaries whose belong to the interstripation region D , and the other on the boundary of the interstripation region ∂D .

As a result of the union of the information about the function from the four stripes, we obtain some rectangular region in D , where the information about the function is unknown. The operator $Lf(x, y)$ in the interstripation operator in this case can be simplified to the following form:

$$Lf(x, y) = \frac{x - \tilde{x}_2}{\tilde{x}_1 - \tilde{x}_2} f(y)|_{\Gamma_1} + \frac{x - \tilde{x}_1}{\tilde{x}_2 - \tilde{x}_1} f(y)|_{\Gamma_2} + \\ + \frac{y - \tilde{y}_2}{\tilde{y}_1 - \tilde{y}_2} f(x)|_{\Gamma_3} + \frac{y - \tilde{y}_1}{\tilde{y}_2 - \tilde{y}_1} f(x)|_{\Gamma_4} - \\ - \frac{x - \tilde{x}_2}{\tilde{x}_1 - \tilde{x}_2} \left(\frac{y - \tilde{y}_2}{\tilde{y}_1 - \tilde{y}_2} f(\tilde{x}_1, \tilde{y}_1) + \frac{y - \tilde{y}_1}{\tilde{y}_2 - \tilde{y}_1} f(\tilde{x}_1, \tilde{y}_2) \right) - \\ - \frac{x - \tilde{x}_1}{\tilde{x}_2 - \tilde{x}_1} \left(\frac{y - \tilde{y}_2}{\tilde{y}_1 - \tilde{y}_2} f(\tilde{x}_2, \tilde{y}_1) + \frac{y - \tilde{y}_1}{\tilde{y}_2 - \tilde{y}_1} f(\tilde{x}_2, \tilde{y}_2) \right), \\ \Gamma_1 : x = \tilde{x}_1, \Gamma_2 : x = \tilde{x}_2, \Gamma_3 : y = \tilde{y}_1, \Gamma_4 : y = \tilde{y}_2, \\ \tilde{x}_1 = \begin{cases} a_1, a_1 \notin \partial D, \\ b_1, b_1 \notin \partial D, \end{cases}$$

$$\tilde{x}_2 = \begin{cases} a_2, a_2 \notin \partial D, \\ b_2, b_2 \notin \partial D, \end{cases}$$

$$\tilde{y}_1 = \begin{cases} c_1, c_1 \notin \partial D, \\ d_1, d_1 \notin \partial D, \end{cases}$$

$$\tilde{y}_2 = \begin{cases} c_2, c_2 \notin \partial D, \\ d_2, d_2 \notin \partial D. \end{cases}$$

Let's assume that the information about the function $f(x, y)$ is given by its traces on two stripes $S_{x,1}$ and $S_{y,1}$. Moreover, it is known that for each of the stripes both boundaries belong to the interstripation region D .

As a result of the union of the information about the function from both stripes, we obtain in the general case four rectangular regions in D , where the information about the function is unknown. The operator $Lf(x, y)$ in the interstripation operator in this case can be simplified to the following form:

$$\begin{aligned} Lf(x, y) = & \frac{x-b_1}{a_1-b_1} f(y)|_{\Gamma_1} + \frac{x-a_1}{b_1-a_1} f(y)|_{\Gamma_2} + \\ & + \frac{y-d_1}{c_1-d_1} f(x)|_{\Gamma_3} + \frac{y-c_1}{d_1-c_1} f(x)|_{\Gamma_4} - \\ & - \frac{x-b_1}{a_1-b_1} \left(\frac{y-d_1}{c_1-d_1} f(a_1, c_1) + \frac{y-c_1}{d_1-c_1} f(a_1, d_1) \right) - \\ & - \frac{x-a_1}{b_1-a_1} \left(\frac{y-d_1}{c_1-d_1} f(b_1, c_1) + \frac{y-c_1}{d_1-c_1} f(b_1, d_1) \right), \\ & \Gamma_1: x = a_1, \Gamma_2: x = b_1, \Gamma_3: y = c_1, \Gamma_4: y = d_1. \end{aligned}$$

For all the above interstripation operators, the same properties will be valid as for the general interstripation operator.

4 EXPERIMENTS

For computational experiments, a computer program was developed in the Python programming language using additional libraries (for example, Matplotlib – for visualization, plotting function graphs, SymPy – for finding function approximations in analytical form).

The interstripation region $D = [-3, 3] \times [-3, 3]$ was chosen for computational experiments.

The following function was chosen as the test function:

$$f(x, y) = Ae^{\frac{(x-x_0)^2}{2\sigma_x^2}} + Be^{\frac{(y-y_0)^2}{2\sigma_y^2}}.$$

Test function parameters for computational experiments:

$$A = 0, B = 0, x_0 = 0, y_0 = 0, \sigma_x = 1, \sigma_y = 1.$$

Analytical form of the test function after parameter substitution:

$$f(x, y) = e^{-0.5x^2} + e^{-0.5y^2}.$$

For each computational experiment, a system of stripes was given. The function itself $f(x, y)$ was chosen in the traces of the function on each of the stripes in the systems (i.e., the function on the stripes was considered to be known exactly, without errors).

Four computational experiments were conducted for each of the above special cases of the location of stripes on the interstripation region.

For the first computational experiment, the following system of stripes was specified:

$$\begin{aligned} S_{x,1} &= \{(x, y) : -3 \leq x \leq 0\}, \\ S_{y,1} &= \{(x, y) : 0 \leq y \leq 3\}. \end{aligned}$$

For the second computational experiment, the following system of stripes was specified:

$$\begin{aligned} S_{x,1} &= \{(x, y) : -3 \leq x \leq -1\}, \\ S_{x,2} &= \{(x, y) : 1 \leq x \leq 3\}, \\ S_{y,1} &= \{(x, y) : 1 \leq y \leq 3\}. \end{aligned}$$

For the third computational experiment, the following system of stripes was specified:

$$\begin{aligned} S_{x,1} &= \{(x, y) : -3 \leq x \leq -1\}, \\ S_{x,2} &= \{(x, y) : 1 \leq x \leq 3\}, \\ S_{y,1} &= \{(x, y) : -3 \leq y \leq -1\}, \\ S_{y,2} &= \{(x, y) : 1 \leq y \leq 3\}. \end{aligned}$$

For the fourth computational experiment, the following system of stripes was specified:

$$\begin{aligned} S_{x,1} &= \{(x, y) : -1 \leq x \leq 1\}, \\ S_{y,1} &= \{(x, y) : -1 \leq y \leq 1\}. \end{aligned}$$

5 RESULTS

For each of the computational experiments conducted, an analytical form of the information approximation operator $Lf(x, y)$ was obtained, and graphs of approximations and absolute errors of approximations were also constructed.

Case one. Traces of the function on the boundaries of stripes which belong to the interstripation region ($\tilde{x} = 0$, $\tilde{y} = 0$):

$$\begin{aligned} f(y)|_{\Gamma_1} &= f(0, y) = 1 + e^{-0.5y^2}, \\ f(x)|_{\Gamma_2} &= f(x, 0) = e^{-0.5x^2} + 1, \\ \Gamma_1 : x &= 0, \quad \Gamma_2 : y = 0. \end{aligned}$$

The value at the intersection point of the boundaries of stripes which belong to the interstripation region:

$$f(\tilde{x}, \tilde{y}) = f(0, 0) = 2.$$

Fig. 1 graphically illustrates the union of the test data on the stripes.

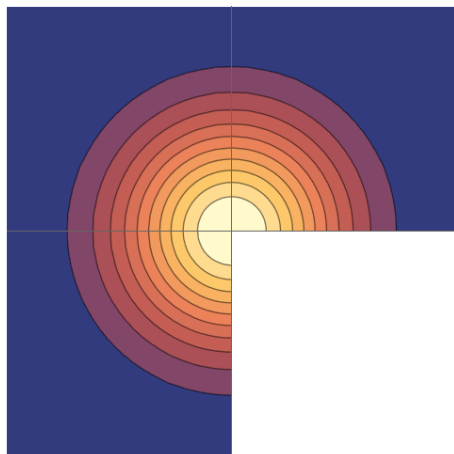


Figure 1 – The known information about the test function based on traces on two stripes for the first computational experiment

After substituting the test data into the corresponding formula for the operator $Lf(x, y)$, we obtain the analytical form of the approximation of the function in the subregion of the region D , where the function is unknown:

$$Lf(x, y) = e^{-0.5x^2} + e^{-0.5y^2}.$$

Case two. Traces of the function on the boundaries of stripes which belong to the interstripation region ($\tilde{x}_1 = -1, \tilde{x}_2 = 1, \tilde{y} = 1$):

$$\begin{aligned} f(y)|_{\Gamma_1} &= f(-1, y) = e^{-0.5} + e^{-0.5y^2}, \\ f(y)|_{\Gamma_2} &= f(1, y) = e^{-0.5} + e^{-0.5y^2}, \\ f(x)|_{\Gamma_3} &= f(x, 1) = e^{-0.5x^2} + e^{-0.5}, \\ \Gamma_1 : x &= -1, \quad \Gamma_2 : x = 1, \quad \Gamma_3 : y = 1. \end{aligned}$$

The values at the intersection points of the boundaries of stripes which belong to the interstripation region:

$$f(\tilde{x}_1, \tilde{y}) = f(-1, 1) = 2e^{-0.5},$$

$$f(\tilde{x}_2, \tilde{y}) = f(1, 1) = 2e^{-0.5}.$$

Fig. 2 graphically illustrates the union of the test data on the stripes.

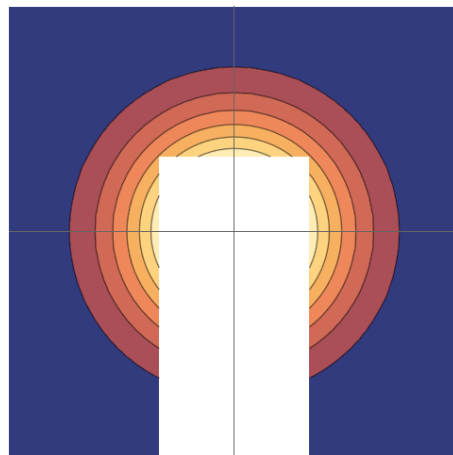


Figure 2 – The known information about the test function based on traces on three stripes for the second computational experiment

After substituting the test data into the corresponding formula for the operator $Lf(x, y)$, we obtain the analytical form of the approximation of the function in the subregion of the region D , where the function is unknown:

$$\begin{aligned} Lf(x, y) &= -\frac{x-1}{2} \left(e^{-0.5} + e^{-0.5y^2} - 2e^{-0.5} \right) + \\ &+ \frac{x+1}{2} \left(e^{-0.5} + e^{-0.5y^2} - 2e^{-0.5} \right) + e^{-0.5x^2} + e^{-0.5}. \end{aligned}$$

After combining the similar ones, we get the following expression for $Lf(x, y)$:

$$Lf(x, y) = e^{-0.5x^2} + e^{-0.5y^2}.$$

Case three. Traces of the function on the boundaries of stripes which belong to the interstripation region ($\tilde{x}_1 = -1, \tilde{x}_2 = 1, \tilde{y}_1 = -1, \tilde{y}_2 = 1$):

$$\begin{aligned} f(y)|_{\Gamma_1} &= f(-1, y) = e^{-0.5} + e^{-0.5y^2}, \\ f(y)|_{\Gamma_2} &= f(1, y) = e^{-0.5} + e^{-0.5y^2}, \\ f(x)|_{\Gamma_3} &= f(x, -1) = e^{-0.5x^2} + e^{-0.5}, \\ f(x)|_{\Gamma_4} &= f(x, 1) = e^{-0.5x^2} + e^{-0.5}, \\ \Gamma_1 : x &= -1, \quad \Gamma_2 : x = 1, \quad \Gamma_3 : y = -1, \quad \Gamma_4 : y = 1. \end{aligned}$$

The values at the intersection points of the boundaries of stripes which belong to the interstripation region:

$$\begin{aligned}f(\tilde{x}_1, \tilde{y}_1) &= f(-1, -1) = 2e^{-0.5}, \\f(\tilde{x}_1, \tilde{y}_2) &= f(-1, 1) = 2e^{-0.5}, \\f(\tilde{x}_2, \tilde{y}_1) &= f(1, -1) = 2e^{-0.5}, \\f(\tilde{x}_2, \tilde{y}_2) &= f(1, 1) = 2e^{-0.5}.\end{aligned}$$

Fig. 3 graphically illustrates the union of the test data on the stripes.

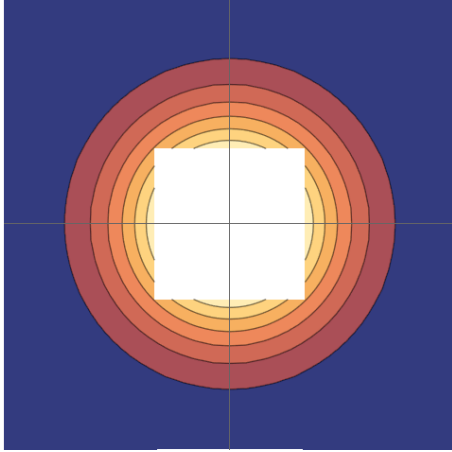


Figure 3 – The known information about the test function based on traces on four stripes for the third computational experiment

After substituting the test data into the corresponding formula for the operator $Lf(x, y)$, we obtain the analytical form of the approximation of the function in the subregion of the region D , where the function is unknown:

$$\begin{aligned}Lf(x, y) &= -\frac{x-1}{2}\left(e^{-0.5} + e^{-0.5y^2}\right) + \frac{x+1}{2}\left(e^{-0.5} + e^{-0.5y^2}\right) - \\&\quad -\frac{y-1}{2}\left(e^{-0.5x^2} + e^{-0.5}\right) + \frac{y+1}{2}\left(e^{-0.5x^2} + e^{-0.5}\right) + \\&\quad + \frac{x-1}{2}\left(-(y-1)e^{-0.5} + (y+1)e^{-0.5}\right) - \\&\quad -\frac{x+1}{2}\left(-(y-1)e^{-0.5} + (y+1)e^{-0.5}\right).\end{aligned}$$

After combining the similar ones, we get the following expression for $Lf(x, y)$:

$$Lf(x, y) = e^{-0.5x^2} + e^{-0.5y^2}.$$

Case four. Traces of the function on the boundaries of stripes:

$$\begin{aligned}f(y)|_{\Gamma_1} &= f(-1, y) = e^{-0.5} + e^{-0.5y^2}, \\f(y)|_{\Gamma_2} &= f(1, y) = e^{-0.5} + e^{-0.5y^2},\end{aligned}$$

$$f(x)|_{\Gamma_3} = f(x, -1) = e^{-0.5x^2} + e^{-0.5},$$

$$f(x)|_{\Gamma_4} = f(x, 1) = e^{-0.5x^2} + e^{-0.5},$$

$$\Gamma_1 : x = -1, \Gamma_2 : x = 1, \Gamma_3 : y = -1, \Gamma_4 : y = 1.$$

The values at the intersection points of the boundaries of stripes:

$$\begin{aligned}f(-1, -1) &= 2e^{-0.5}, \\f(-1, 1) &= 2e^{-0.5}, \\f(1, -1) &= 2e^{-0.5}, \\f(1, 1) &= 2e^{-0.5}.\end{aligned}$$

Fig. 4 graphically illustrates the union of the test data on the stripes.

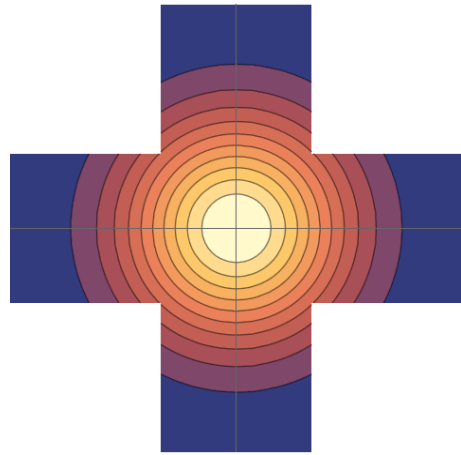


Figure 4 – The known information about the test function based on traces on two stripes for the fourth computational experiment

After substituting the test data into the corresponding formula for the operator $Lf(x, y)$, we obtain the analytical form of the approximation of the function in the subregion of the region D , where the function is unknown:

$$\begin{aligned}Lf(x, y) &= -\frac{x-1}{2}\left(e^{-0.5} + e^{-0.5y^2}\right) + \frac{x+1}{2}\left(e^{-0.5} + e^{-0.5y^2}\right) - \\&\quad -\frac{y-1}{2}\left(e^{-0.5x^2} + e^{-0.5}\right) + \frac{y+1}{2}\left(e^{-0.5x^2} + e^{-0.5}\right) + \\&\quad + \frac{x-1}{2}\left(-(y-1)e^{-0.5} + (y+1)e^{-0.5}\right) - \\&\quad -\frac{x+1}{2}\left(-(y-1)e^{-0.5} + (y+1)e^{-0.5}\right).\end{aligned}$$

After combining the similar ones, we get the following expression for $Lf(x, y)$:

$$Lf(x, y) = e^{-0.5x^2} + e^{-0.5y^2}.$$

6 DISCUSSION

As can be seen from the results of the computational experiments, for all special cases of the location of the stripes in the interstripation region, the given operators accurately restore the value of the function from the given traces of the function on the system of stripes.

This result is achieved due to the fact that for the experiments the function traces on the stripes were chosen in such a way that they exactly coincide with the value of the function.

In practice, to describe the traces of a function on stripes in this way is a problematic and sometimes even impossible task. Therefore, the traces of a function that one has to deal with in practice already contain the quality of inaccuracy and error. Or they are given on some system of points and to apply the informational operators of interstripation it is necessary to first apply interpolation operators, which also approximate the values with some error.

However, interstripation information operators can potentially be applied in many fields of science and technology.

CONCLUSIONS

This work is devoted to methods for approximating the values of continuous functions of two variables, which are given on a system of stripes.

The scientific novelty of the obtained results lies in the fact that for the first time the approximation of continuous functions of two variables, given by their traces on systems of stripes of a special form, was considered. Their difference from information operators of interstripation of a general form lies in the significant simplification of the approximation operators without losing the accuracy of the approximation with a smaller number of arithmetic operations, which can be a decisive factor in some cases.

The practical significance of the obtained results lies in the fact that software has been developed that implements the approximation of continuous functions of two variables by information operators of interstripation in the Lagrange form. The results of the experiments prove the accuracy of the approximation of functions on a given system of strips, the boundaries of which are parallel to the coordinate axes.

Prospects for further research are the application of the proposed information operators in the tasks of digital image processing, seismic exploration data, remote sensing data etc.

REFERENCES

1. Costabile F. A., Gualtieri M. I. Umbral interpolation: A survey, *Mathematics*, 2025, Vol. 13(2), P. 271. DOI: 10.3390/math13020271
2. Liptaj A. General approach to function approximation, *Mathematics*, 2024, Vol. 12(23), P. 3702. DOI: 10.3390/math12233702
3. Malachivskyy P., Melnychok L., Pizyur Ya. Chebyshev approximation of multivariable functions with the interpolation, *Mathematical Modeling and Computing*, 2022, Vol. 9 (3), pp. 757–766. DOI: 10.23939/mmc2022.03.757
4. Malachivskyy P. S., Pizyur Y. V., Malachivskiy R. P., Ukhanska O. M. Chebyshev approximation of functions of several variables, *Cybernetics and Systems Analysis*, 2020, Vol. 56 (1), pp. 76–86. DOI: 10.1007/s10559-020-00227-8
5. Boor de C. R. Topics in multivariate approximation theory, *Topics in Numerical Analysis*, 2006, Vol. 965, pp. 39–78. DOI: 10.1007/BFb0063200
6. Boor de C. R. A practical guide to spline, *Mathematics of Computation*, 1978, Vol. 27(149), pp. 325–326. DOI: 10.2307/2006241
7. Viwatwongkasem C. EM Algorithm for Truncated and Censored Poisson Likelihoods, *Procedia Computer Science*, 2016, Vol. 86, pp. 240–243. DOI: 10.1016/j.procs.2016.05.109
8. Vecchia B. D., Mastroianni G., Vertesi P. Hermite-Fejer interpolation of higher order with varying weights, *Rendiconti del Circolo Matematico di Palermo*, 2005, Vol. 54(1), pp. 37–49. DOI: 10.1007/BF02875742
9. Goodman J., Lin K., Morzfeld M. Small-Noise Analysis and Symmetrization of Implicit Monte Carlo Samplers, *Communications on Pure and Applied Mathematics*, Vol. 69(10), pp. 1924–1951. DOI: 10.1002/cpa.21592
10. Donoho D. L. Compressed sensing, *IEEE Transactions on Information Theory*, 2006, Vol. 52 (4), pp. 1289–1306. DOI: 10.1109/TIT.2006.871582
11. Candes E., Romberg J., Tao T. Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information, *IEEE Transactions on Information Theory*, 2006, Vol. 52(2), pp. 489–509. DOI: 10.1109/TIT.2005.862083
12. Rasmussen C. Williams C. Gaussian Processes for Machine Learning. The MIT Press, 2005, 266 p. DOI: 10.7551/mitpress/3206.001.0001
13. Gosciwski D., Gerus-Gosciwska M., Szczepanska A. Application of polynomial interpolation for iterative complementation of the missing nodes in a regular network of squares used for the construction of a digital terrain model, *Remote sensing*, 2024, Vol. 16(6), pp. 1–24. DOI: 10.3390/rs16060999
14. Franke R., Nielson G. Scattered data interpolation and applications: a tutorial and survey, *Computer Graphics – Systems and Applications*, 1991, pp. 131–160. DOI: 10.1007/978-3-642-76404-2_6
15. Lytvyn O. M., Lytvyn O. O., Slavik O. V. Generalized interstripation of functions of two variables, *Cybernetics and Systems Analysis*, 2018, № 54(3), pp. 465–475. DOI: 10.1007/s10559-018-0047-7
16. Mezhyuev V., Lytvyn O. N., Lytvyn O. O., Pershyna Yu. I., Nechuviter O. P. Algorithm for the reconstruction of the discontinuous structure of a body by its projections along mutually perpendicular lines, *Conference Abstract: 2018 7th International Conference on Software and Computer Applications (ICSCA 2018), February 8–10, 2018, P. 26. DOI: 10.1145/3185089.3185115*

Received 26.06.2025.
Accepted 25.09.2025.

ПРО ОСОБЛИВИ ВИПАДКИ ЛАГРАНЖЕВОЇ ІНТЕРСТРІПАЦІЇ НАБЛИЖЕННЯ ФУНКЦІЙ ДВОХ ЗМІННИХ

Славик О. В. – канд. фіз.-мат. наук, асистент кафедри прикладної математики, Харківський національний університет радіоелектроніки, м. Харків, Україна, доцент кафедри інформаційних комп'ютерних технологій і математики, Навчально-науковий інститут «Українська інженерно-педагогічна академія» Харківського національного університету імені В. Н. Каразіна, м. Харків, Україна.

АНОТАЦІЯ

Актуальність. Розглянуто задачу наближення значень неперервних функцій двох змінних за відомою інформацією про неї на смугах, границі яких паралельні осям координат. Об'єктом дослідження є процес наближення значень функцій на основі неповної інформації про них, що задана на системі смуг.

Мета роботи – огляд інформаційних операторів інтерстріпації Лагранжа та особливостей побудови інформаційних операторів наближення для деяких випадків взаємного розташування смуг у просторі, які дозволяють значно спростити обчислення наближених значень функції у невідомих підобластях області дослідження.

Метод. Запропоновано методи наближення значень неперервних функцій двох змінних за неповної інформацією про неї на деякій обмеженій області. Інформація про функцію є відомою лише на системі смуг, обмежених прямими, що паралельні осям координат. Запропоновано метод наближення значень неперервних функцій двох змінних, інформація про яку є відомою на двох смугах, в результаті об'єднання яких у області залишається невідомою лише деяка прямокутна підобласть. Запропоновано метод наближення значень неперервних функцій двох змінних, інформація про яку є відомою на трьох смугах, в результаті об'єднання яких у області залишається невідомою лише деяка прямокутна підобласть. Запропоновано метод наближення значень неперервних функцій двох змінних, інформація про яку є відомою на чотирьох смугах, в результаті об'єднання яких у області залишається невідомою лише деяка прямокутна підобласть. Запропоновано метод наближення значень неперервних функцій двох змінних, інформація про яку є відомою на двох смугах, в результаті об'єднання яких у області залишається невідомими чотири прямокутні підобласті. Для всіх розглянутих випадків наведено оператори наближення, які дозволяють обчислити наближений вигляд функції у невідомих підобластях у аналітичному вигляді.

Результати. Інформаційні оператори інтерстріпації Лагранжа реалізовані програмно і досліджені в задачах наближення значень функцій двох змінних за відомою інформацією про неї на системах смуг.

Висновки. Проведені експерименти підтвердили точність наближення значень неперервних функцій двох змінних запропонованих інформаційних операторів інтерстріпації для різних систем смуг. Наведено оператори наближення для особливих випадків розташування смуг у області дослідження, відмінність яких від інформаційних операторів інтерстріпації загального вигляду полягає у значному спрощенні операторів наближення не втрачаючи при цьому точності наближення при меншій кількості арифметичних операцій, що може бути вирішальним фактором в деяких випадках. Перспективи подальших досліджень полягають у застосуванні запропонованих інформаційних операторів у задачах цифрової обробки зображень, даних сейсмічної розвідки та даних дистанційного зондування планети.

КЛЮЧОВІ СЛОВА: чисельні методи, математичне моделювання, інформаційні оператори, інтерплінація, інтерстріпація.

ЛІТЕРАТУРА

1. Costabile F. A. Umbral interpolation: A survey / F. A. Costabile, M. I. Gualtieri // *Mathematics*. – 2025. – Vol. 13(2). – P. 271. DOI: 10.3390/math13020271
2. Liptaj A. General approach to function approximation / A. Liptaj // *Mathematics*. – 2024. – Vol. 12(23). – P. 3702. DOI: 10.3390/math12233702
3. Malachivskyy P. Chebyshev approximation of multivariable functions with the interpolation / P. Malachivskyy, L. Melnychok, Ya. Pizyur // *Mathematical Modeling and Computing*. – 2022. – Vol. 9 (3). – P. 757–766. DOI: 10.23939/mmc2022.03.757
4. Chebyshev approximation of functions of several variables / [P. S. Malachivskyy, Y. V. Pizyur, R. P. Malachivskiy, O. M. Ukhanska] // *Cybernetics and Systems Analysis*. 2020. – Vol. 56 (1). – P. 76–86. DOI: 10.1007/s10559-020-00227-8
5. Boor de C. R. Topics in multivariate approximation theory / C. R. de Boor // *Topics in Numerical Analysis*. – 2006. – Vol. 965. – P. 39–78. DOI: 10.1007/BFb0063200
6. Boor de C. R. A practical guide to spline / C. R. de Boor // *Mathematics of Computation*. – 1978. – Vol. 27(149). – P. 325–326. DOI: 10.2307/2006241
7. Viwatwongkasem C. EM Algorithm for Truncated and Censored Poisson Likelihoods / C. Viwatwongkasem // *Procedia Computer Science*. – 2016. – Vol. 86. – P. 240–243. DOI: 10.1016/j.procs.2016.05.109
8. Vecchia B. D. Hermite-Fejer interpolation of higher order with varying weights / B. D. Vecchia, G. Mastroianni, P. Vertesi // *Rendiconti del Circolo Matematico di Palermo*. – 2005. – Vol. 54(1). P. 37–49. DOI: 10.1007/BF02875742
9. Goodman J. Small-Noise Analysis and Symmetrization of Implicit Monte Carlo Samplers / J. Goodman, K. Lin, M. Morzfeld // *Communications on Pure and Applied Mathematics*. – Vol. 69(10). – P. 1924–1951. DOI: 10.1002/cpa.21592
10. Donoho D. L. Compressed sensing / D. L. Donoho // *IEEE Transactions on Information Theory*. – 2006. – Vol. 52 (4). – P. 1289–1306. DOI: 10.1109/TIT.2006.871582
11. Candes E. Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information / E. Candes, J. Romberg, T. Tao // *IEEE Transactions on Information Theory*. – 2006. – Vol. 52(2). – P. 489–509. DOI: 10.1109/TIT.2005.862083
12. Rasmussen C. Gaussian Processes for Machine Learning / C. Rasmussen, C. Williams. – The MIT Press, 2005. – 266 p. DOI: 10.7551/mitpress/3206.001.0001
13. Gosciwski D. Application of polynomial interpolation for iterative complementation of the missing nodes in a regular network of squares used for the construction of a digital terrain model / D. Gosciwski, M. Gerus-Gosciwska, A. Szczepanska // *Remote sensing*. – 2024. – Vol. 16(6). – P. 1–24. DOI: 10.3390/rs16060999
14. Franke R. Scattered data interpolation and applications: a tutorial and survey / R. Franke, G. Nielson // *Computer Graphics – Systems and Applications*. – 1991. – P. 131–160. DOI: 10.1007/978-3-642-76404-2_6
15. Lytvyn O. M. Generalized interstripation of functions of two variables / O. M. Lytvyn, O. O. Lytvyn, O. V. Slavik // *Cybernetics and Systems Analysis*. – 2018. – № 54(3). – P. 465–475. DOI: 10.1007/s10559-018-0047-7
16. Algorithm for the reconstruction of the discontinuous structure of a body by its projections along mutually perpendicular lines / [V. Mezhyuev, O. N. Lytvyn, O. O. Lytvyn et al.] // *Conference Abstract: 2018 7th International Conference on Software and Computer Applications (ICSCA 2018)*, February 8–10. – 2018. – P. 26. DOI: 10.1145/3185089.3185115

OPTIMIZATION OF PERMANENT DECOMPOSITION PROCEDURES USING PARALLELIZATION ALGORITHM

Turbal Y. V. – Dr. Sc., Professor, Head of the Department of computer science and applied mathematics, National University of Water and Environmental Engineering, Rivne, Ukraine.

Moroziuk A. Y. – Post-graduate student of the Department of computer science and applied mathematics, National University of Water and Environmental Engineering, Rivne, Ukraine.

ABSTRACT

Context. The problem of efficiently finding all permutations of a list of N elements is a key problem in many areas of computer science, such as combinatorics, optimization, cryptography, and machine learning. The object of the study was to analyze the procedure of permanent decomposition and propose an algorithm for its parallelization using modern features of working with threads in C#.

Objective. The goal of the work is the creation of the algorithm for parallelizing the generation of permutations using permanent decomposition processes.

Method. The main research method is the comparison of various algorithms with proposed parallelized algorithm, taking into account such criteria as accuracy and speed. Scientific works [10, 9, 17] present algorithms, including the regular permanent decomposition algorithm and the Johnson-Trotter algorithm. The Johnson-Trotter's algorithm is one of the most effective, so it has been taken as some kind of benchmark.

It is worth mentioning that each paralleling process has its own disadvantages, including additional resources needed for data synchronization between threads. This can be minimized using both technical abilities of modern programming languages and optimization of the algorithm itself.

Results. The developed parallelized algorithm have improved performance of the regular permanent decomposition algorithm for solving the problem of finding all permutations.

Conclusions. The conducted experiments have confirmed the proposed parallelized algorithm's version is better from a performance standpoint than the regular one. The prospects for further research may include the application of the parallelized version of the algorithm to some practical tasks and comparison of the results.

KEYWORDS: parallelization, permanent decomposition, multi thread, algorithm, C#.

ABBREVIATIONS

TPL is a Task Parallel Library.

NOMENCLATURE

A is an input data list;

a_i is an i -th element of the input data list;

B is an initial list for parallel algorithm;

N is a number of elements in the input data list;

$O()$ is an algorithm's complexity;

$perm()$ is a permanent of the matrix;

S_N is a set of all permutations of the indices of the rows or columns of a matrix;

x is a number of all permutations for an initial list for parallel algorithm;

X_0 is a list of all permutations for an initial list for parallel algorithm;

X_j is an extended list of all permutations.

INTRODUCTION

In the modern theory of combinatorics and object generation algorithms, methods that allow solving complex problems, including permutations, combinations, and other structures, are of great importance. One of these technologies is the method of permanent decomposition, which has proven to be effective [17]. However, the process of generating combinatorial objects is a difficult task,

especially for large matrices, since the number of possible permutations of matrix elements grows exponentially with its size (complexity $O(N!)$ [3]). Parallelization of computations can significantly increase the efficiency of this process. In this article, we will consider approaches to parallelizing the permanent decomposition algorithm and the main technical aspects of this process.

The object of study is the problem of generating all permutations of a list of N elements.

This is a fundamental problem in computer science with applications in areas such as combinatorics, optimization, cryptography, and algorithm design. Therefore, to improve the speed of the permanent decomposition algorithm it is possible to parallelize it and make some performance measurements.

The subject of study is the parallelization techniques that can be applied to permanent decomposition algorithm.

For the effectiveness comparison Johnson-Trotter and regular (sequential) version of algorithm will be used.

The purpose of the work is to implement a parallel version of the algorithm and compare performance with other algorithms.

1 PROBLEM STATEMENT

Suppose given the original list of N elements $A = [a_1, a_2, \dots, a_N]$.

For a given list A the problem of generating all permutations can be divided into pieces and parallelized, to improve execution speed for this process using modern multi-core computers.

In turn, the problem of parallelizing permanent decomposition process is to implement algorithm of dividing original list into sub-lists and combine results from each thread/operation.

2 REVIEW OF THE LITERATURE

Sequential algorithms for generating permutations have been widely studied. One of the earliest and most notable methods is the Johnson-Trotter algorithm, introduced by Johnson [9] and Trotter [16]. This algorithm generates permutations using adjacent transpositions and is efficient in terms of memory usage. However, its computational complexity becomes a limiting factor for large values of n . Another significant contribution is Heap's algorithm, proposed by Heap [6], which generates permutations in a non-lexicographic order and is celebrated for its simplicity and in-place nature. Knuth, in *The Art of Computer Programming* [10], provides an extensive analysis of these and other permutation generation algorithms, highlighting their theoretical foundations and practical limitations.

With the rise of multi-core and distributed computing systems, researchers have turned their attention to parallel algorithms to enhance the efficiency of combinatorial generation. Akl, in *The Design and Analysis of Parallel Algorithms* [1], discusses the potential of parallel computing for solving combinatorial problems. More recently, in *Structured Parallel Programming* [12], author emphasize the importance of leveraging modern threading models and parallel frameworks to optimize computationally intensive tasks.

Permanent decomposition is a mathematical technique with applications in combinatorial optimization and matrix theory. Valiant, in his seminal work on the complexity of computing the permanent [18], highlighted its relevance in graph theory and statistical physics. Glynn [4] later proposed an efficient algorithm for computing the permanent of a matrix, which has been adapted for combinatorial generation tasks. Despite these advancements, the parallelization of permanent decomposition procedures remains an understudied area, with limited research addressing its implementation in multi-core systems.

Modern programming languages, such as C#, offer robust support for parallel computing through threading libraries and task-based parallelism. Richter, in his book "CLR via C#" [13], explores the threading capabilities of C# and their application to computationally intensive tasks. Similarly, Toub, in his lecture [11], provides practical insights into designing scalable parallel algorithms using the TPL [15]. These resources have been instrumental in developing efficient parallel implementations of combinatorial algorithms.

The generation of permutations and combinations has numerous applications across various domains. Garey and Johnson, in *Computers and Intractability* [8], discuss the

role of combinatorial algorithms in solving NP-hard problems. In bioinformatics, Gusfield [5], highlights the use of combinatorial generation for sequence alignment and genome analysis. Babych and Turbal [17] show how the permanent decomposition methods might be applied to solve schedule generation problems.

3 MATERIALS AND METHODS

The permanent of a square matrix is a generalization of the determinant without taking into account the change of permutation signs. The algorithmic complexity of calculating the permanent grows exponentially with the size of the matrix, which limits its practical application for large matrices. The standard approach, based on trying all possible permutations of rows or columns, has a complexity of $O(N!)$. This makes the task inefficient for large matrices. To speed up the computation, efficient parallelization algorithms are needed.

The permanent of a matrix A of size $N \times N$ is defined as follows:

$$\text{perm}(A) = \sum_{\sigma \in S_N} \prod_{i=1}^N a_{i, \sigma(i)}. \quad (1)$$

Due to the absence of sign changes (as in the determinant), the constant has no analytically simple ways to compute it, so computing the constant is a combinatorially difficult task.

The method of permanent decomposition is used to efficiently generate combinatorial objects such as permutations, combinations with and without repetitions, and various subsets [17]. The main idea of the method is to perform a recursive process with gradual memorization of element indices instead of directly calculating the algebraic permanent of the incidence matrix. This reduces the number of operations and optimizes the generation process.

An example is the generation of permutations of the set $\{1, 2, 3\}$. Instead of checking all possible variants, the permanent decomposition method uses a special incidence matrix to build permutations step by step, which reduces computational complexity. As a result, the method offers an effective alternative to classical approaches such as the lexicographic method or the Johnson-Trotter algorithm.

This approach is especially useful in problems with a large number of objects or complex constraints, such as planning or resource optimization, where it is necessary to compute many possible options in a short time.

Parallelization of computations can significantly speed up the process of calculating a constant, especially when there are a large number of permutations or subtasks. The main approach to parallelization is based on dividing the computation of a constant into several independent threads or processes that can run simultaneously on different processor cores or cluster nodes.

Distribution of permutations: Since a persistent depends on all possible permutations of rows or columns, these permutations can be distributed among multiple

threads. Each thread computes its own part of the permutations, and then the results are combined. This approach allows you to linearly reduce the computation time in the case of perfect parallelization.

Recursive decomposition: Another parallelization approach is based on recursive decomposition of the matrix into smaller blocks. Each block can be computed independently, and these computations can be performed in parallel on different threads or nodes. For example, if we have a matrix of size $N \times N$, it can be divided into four blocks of size $N/2 \times N/2$, and the permanent of each of these blocks can be calculated on a separate thread.

The software implementation was made using the C# programming language, as it provides a wide range of tools for organizing parallel computing through the TPL [15], which allows for the efficient use of multi-core processors. The main tools for implementing parallel execution are the Parallel, Task classes [14] and asynchronous programming concepts.

One of the easiest ways to parallelize the calculation of a constant is to use the Parallel.For loop [2], which allows you to distribute tasks among several threads. In the context of calculating a permanent, each thread can be responsible for calculating the permanent for a certain subset of matrix permutations.

An alternative approach is the use of tasks, which allows for more flexible thread management and asynchronous execution of tasks, which makes the implementation of this approach more complex. It is worth noting that it is precisely because of the additional flexibility and the possibility of fine-tuning that the choice fell on this method of working with parallelism.

As already mentioned, parallelization should speed up the work by dividing it among several performers, while synchronization, resource sharing, and blocking [7] can degrade the performance of the software implementation of the algorithm, so it is important to maintain a balance between parallelization and the excessive costs of such a system. To implement the algorithm, we used both single-threading [2] and multi-threading [7].

The proposed algorithm is based on a combination of several approaches: permutation distribution and recursive decomposition. Take for example a list of N values (for better visualization, it can be a list of numbers from 1 to N). The purpose of a parallelized algorithm is to perform part of the work in parallel, while obtaining a list of all possible permutations for the input list. The algorithm itself in this case is as follows:

Here is the description of the algorithm:

1. The input is a list A of N elements $A = [a_1, a_2, \dots, a_N]$.
2. Generate the initial list B , which will contain $(N-1)$ elements, for example, we take all the elements of the list except the first one – a_1 .

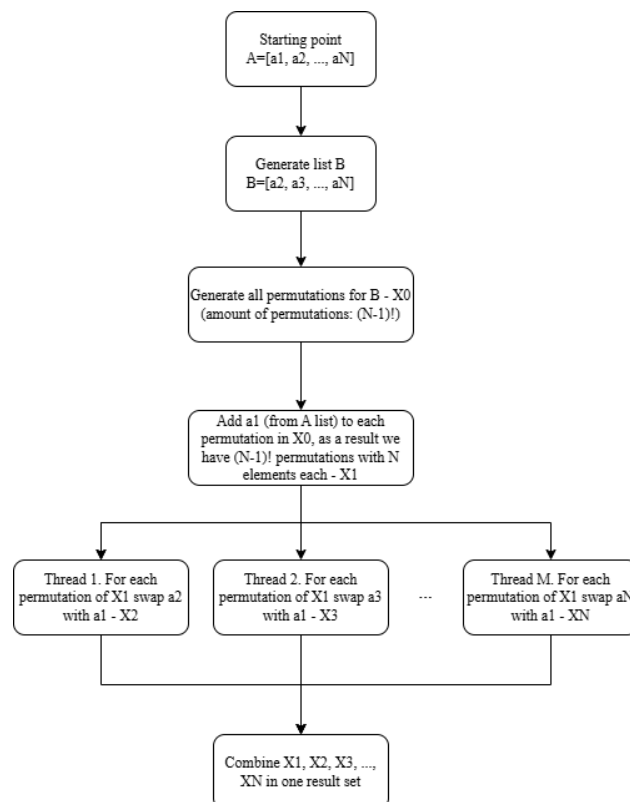


Figure 1 – Parallel permanent decomposition algorithm

3. Generate all possible permutations X_0 for list B in the main thread. As a result of this step, we will get x possible permutations, where $x = (N-1)!$, and the number of elements in each permutation is $(N-1)$.

4. Generate X_1 by adding one more element (a_1) to each permutation, as a result we get x permutations, but with the number of elements N in each.

5. Run a loop through our list A and process each iteration of this loop in parallel, running step 6 in a separate thread.

6. Generate a new list of combinations based on list X_1 , swapping a_i and a_1 .

7. After all parallel tasks are completed, combine the results, and the output is a list of all possible combinations of elements of list A .

The software implementation was made using the tools of the C# programming language. Main method that contains logic of running tasks in different threads is provided on Figure 2.

```
public static async Task<List<List<int>>> Get(int n, int numberOfThreads = 4)
{
    List<int> nums = new List<int>();
    for (int i = 1; i <= n; i++)
    {
        nums.Add(i);
    }
    List<List<int>> permutations = new List<List<int>>();

    var numsToBeProcessed = nums.Except(new List<int> { nums[0] }).ToList();
    var initialSet = Permute(numsToBeProcessed.ToArray());
    initialSet.ForEach(item => item.Add(nums[0]));
    permutations.AddRange(initialSet);
    List<Task> tasks = new List<Task>();

    foreach (var number in nums)
    {
        if (number != nums[0])
        {
            tasks.Add(Task.Run(() => ProcessOneElement(nums, number, initialSet, nums[0])));
        }
    }

    await Task.WhenAll(tasks);
    foreach (Task<List<List<int>>> task in tasks)
    {
        permutations.AddRange(task.Result);
    }

    return permutations;
}
```

Figure 2 – Main method of parallel algorithm implementation

4 EXPERIMENTS

The experimental study compared the performance of three approaches to finding all permutations of a set of N elements:

- Johnson-Trotter method;
- Permanent decomposition method (sequential variant);
- Permanent decomposition method (parallel variant).

The tests were conducted for different values of N (from 9 to 11, since the difference for less than 9 is relatively small and is not representative) on the same hardware using a multi-core processor for parallel execution. The experiments were performed on a computer with the

following characteristics: Intel(R) Core(TM) i7-10510U CPU @ 1.80GHz 2.30GHz, 16 GB of RAM.

5 RESULTS

The fragment of the results of conducted experiments is presented in the Table 1.

The main results of the experiments show that the Johnson-Trotter method demonstrates the best results among the other tested algorithms, but as N grows, its execution time increases exponentially. The sequential variant of the permanent decomposition method also has an exponential increase in execution time, but it is slightly slower. The parallel variant of this method shows a reduction in execution time compared to sequential approaches, but as N grows, the resource consumption for thread synchronization increases.

In terms of memory consumption, the Johnson-Trotter method requires the least resources due to the minimum number of additional structures. Instead, the permanent decomposition methods, especially the parallel one, require more memory to store data on subtasks and control threads.

The efficiency of parallel execution is highest when using 4–8 cores, after which the performance gain decreases due to uneven distribution of tasks and synchronization overhead.

In terms of implementation complexity, the Johnson-Trotter method is the easiest to implement. On the contrary, permanent decomposition methods require a significant amount of code to handle recursions and synchronize threads in parallel execution.

Table 1 – The fragment of experimental results for selected methods, grouped by number of input elements

Number of elements (n)	Permanent decomposition method	Parallelized permanent decomposition method	Johnson-Trotter method
9	0.141s	0.138s	0.129s
10	1.12s	1.065s	0.987s
11	11.825s	11.004s	10.031s

6 DISCUSSION

As it evident from the Table 1, with the increasing of examples elements the complexity of the task is increased significantly.

The experimental results presented in this study provide valuable insights into the performance, scalability, and practicality of the Johnson-Trotter method and permanent decomposition methods for generating permutations. The findings highlight the trade-offs between execution time, memory consumption, and implementation complexity, offering a comprehensive understanding of the strengths and limitations of each approach.

The Johnson-Trotter method demonstrated superior performance in terms of execution time for smaller values of N , outperforming both sequential and parallel variants of the permanent decomposition method. This aligns with

existing literature, which emphasizes the efficiency of the Johnson-Trotter algorithm for small to moderately sized problems [9].

The parallel variant of the permanent decomposition method showed a reduction in execution time compared to its sequential counterpart, particularly when utilizing 4–8 cores. This is consistent with the findings of McCool [12], who emphasize the potential of parallel computing for optimizing computationally intensive tasks. However, the performance gain diminishes as the number of cores increases beyond this range, primarily due to the overhead associated with thread synchronization and uneven task distribution. This observation highlights the challenges of scaling parallel algorithms for combinatorial problems, as noted by Akl [1].

CONCLUSIONS

Parallelization of the permanent decomposition algorithm is an important step to speed up the computation of large matrices. Thanks to modern parallel programming tools, it is possible to reduce the execution time of such algorithms by distributing computational tasks over several threads. The effectiveness of parallelization depends on the structure of the matrix, the chosen decomposition method, the uniformity of load distribution between threads, and the degree of parallelization.

The above algorithm works correctly and has prospects for further use for applied tasks.

The scientific novelty of obtained results is that the algorithm of parallelization of the permanent decomposition method is firstly proposed. It allows us to improve the efficiency of this method and make it more scalable for modern multi-core systems.

The practical significance of obtained results is that the software implementation of proposed algorithm is developed, as well as experiments to study their efficiency are conducted. The experimental results allow to recommend the proposed algorithm for use in practice to gain better results.

Prospects for further research are to improve load balancing methods for parallel algorithms; to adapt and apply developed solution to distributed computing environments.

REFERENCES

1. Akl S. G. The design and analysis of parallel algorithms. Englewood Cliffs, N.J : Prentice-Hall, 1989, 401 p.
2. Albahari J., Albahari B. C# 7.0 in a nutshell: the definitive reference. [S. l.], O'Reilly Media, 2017, 1088 p.
3. Cormen T. H., Stein C. , Rivest R. L. Introduction to algorithms. [S. l.], MIT Press, 2009, 1320 p.
4. Glynn D. G. The permanent of a square matrix [Electronic resource], *European journal of combinatorics*, 2010, Vol. 31, No. 7, pp. 1887–1891. Mode of access: <https://doi.org/10.1016/j.ejc.2010.01.010>
5. Gusfield D. Algorithms on strings, trees, and sequences: computer science and computational biology. Cambridge [England], Cambridge University Press, 1999, 534 p.
6. Heap B. R. Permutations by interchanges [Electronic resource], *The computer journal*, 1963, Vol. 6, No. 3, pp. 293–298. Mode of access: <https://doi.org/10.1093/comjnl/6.3.293>
7. Herlihy M., Shavit N. Art of multiprocessor programming . [S. l.], Elsevier Science & Technology Books, 2011.
8. Johnson D. S., Garey. M. R. Computers and intractability: A guide to the theory of NP-completeness. New York, W.H. Freeman, 1983, 340 p.
9. Johnson S. M. Generation of permutations by adjacent transposition [Electronic resource], *Mathematics of computation*, 1963, Vol. 17, No. 83, pp. 282–285. Mode of access: <https://doi.org/10.1090/s0025-5718-1963-0159764-2>
10. Knuth D. E. The art of computer programming: generating all tuples and permutations. 2nd ed. Upper Saddle River [etc.], Addison-Wesley, 2009, 128 p.
11. Let's talk parallel programming with Stephen Toub [Electronic resource], *Microsoft Learn*. Mode of access: <https://learn.microsoft.com/en-us/shows/on-dotnet/lets-talk-parallel-programming-with-stephen-toub>
12. McCool M., Reinders J., Robison A. Structured parallel programming: patterns for efficient computation. [S. l.], Elsevier Science & Technology Books, 2012, 432 p.
13. Richter J. CLR via C#. [S. l.]: Microsoft Press, 2012, 894 p.
14. Task-based asynchronous programming – .NET [Electronic resource], *Microsoft Learn*. Mode of access: <https://learn.microsoft.com/en-us/dotnet/standard/parallel-programming/task-based-asynchronous-programming>
15. Task parallel library (TPL) – .NET [Electronic resource] // Microsoft Learn. Mode of access: <https://learn.microsoft.com/uk-ua/dotnet/standard/parallel-programming/task-parallel-library-tpl>
16. Trotter H. F. Algorithm 115: perm [Electronic resource], *Communications of the ACM*, 1962, Vol. 5, No. 8, pp. 434–435. Mode of access: <https://doi.org/10.1145/368637.368660>
17. Turbal Y. V., Babysh S. V., Kunanets N. E. Permanent decomposition algorithm for the combinatorial objects generation [Electronic resource], *Radio electronics, computer science, control*, 2022, No. 4, P. 119. Mode of access: <https://doi.org/10.15588/1607-3274-2022-4-10>
18. Valiant L. G. The complexity of computing the permanent [Electronic resource], *Theoretical computer science*, 1979, Vol. 8, No. 2, pp. 189–201. Mode of access: [https://doi.org/10.1016/0304-3975\(79\)90044-6](https://doi.org/10.1016/0304-3975(79)90044-6)

Received 19.05.2025.
Accepted 19.09.2025.

ОПТИМІЗАЦІЯ ПРОЦЕДУР ДЕКОМПОЗИЦІЇ ПЕРМАНЕНТУ З ВИКОРИСТАННЯМ АЛГОРИТМУ РОЗПАРАЛЕЛЮВАННЯ

Турбал Ю. В. – д-р техн. наук, професор, завідувач кафедри комп'ютерних наук та прикладної математики Національного університету водного господарства та природокористування, Рівне, Україна.

Морозюк А. Ю. – аспірант кафедри комп'ютерних наук та прикладної математики Національного університету водного господарства та природокористування, Рівне, Україна.

АНОТАЦІЯ

Актуальність. Задача ефективного знаходження всіх перестановок списку з N елементів є ключовою проблемою в багатьох областях комп'ютерних наук, таких як комбінаторика, оптимізація, криптографія та машинне навчання. Мета дослідження – проаналізувати процедуру перманентної декомпозиції та запропонувати алгоритм для її розпаралелювання з використанням сучасних можливостей роботи з потоками в мові C#.

Мета роботи – метою роботи є створення алгоритму для розпаралелювання генерації перестановок з використанням перманентних процесів декомпозиції.

Метод. Основним методом дослідження є порівняння різних алгоритмів із запропонованим розпаралеленим алгоритмом з урахуванням таких критеріїв, як точність та швидкість. У наукових працях [10, 9, 17] представлено алгоритми, серед яких регулярний алгоритм перманентної декомпозиції та алгоритм Джонсона-Троттера. Алгоритм Джонсона-Троттера є одним з найефективніших, тому його було взято за певний еталон.

Варто зазначити, що кожен процес розпаралелювання має свої недоліки, зокрема, додаткові ресурси, необхідні для синхронізації даних між потоками. Це можна мінімізувати, використовуючи як технічні можливості сучасних мов програмування, так і оптимізацію самого алгоритму.

Результати. Розроблений розпаралелений алгоритм дозволив покращити продуктивність звичайного алгоритму перманентної декомпозиції для розв'язання задачі знаходження всіх перестановок.

Висновки. Проведені експерименти підтвердили, що запропонована розпаралелена версія алгоритму є кращою з точки зору продуктивності, ніж звичайна. Перспективами подальших досліджень може бути застосування розпаралеленої версії алгоритму до деяких практичних задач та порівняння отриманих результатів.

КЛЮЧОВІ СЛОВА: розпаралелювання, постійна декомпозиція, багатопотоковість, алгоритм, C#.

ЛІТЕРАТУРА

1. Akl S. G. The design and analysis of parallel algorithms / Selim G. Akl. – Englewood Cliffs, N.J.: Prentice-Hall, 1989. – 401 p.
2. Albahari J. C# 7.0 in a nutshell: the definitive reference / J. Albahari, B. Albahari. – [S. l.]: O'Reilly Media, 2017. – 1088 p.
3. Cormen T. H. Introduction to algorithms / Thomas H. Cormen, Clifford Stein, Ronald L. Rivest. – [S. l.]: MIT Press, 2009. – 1320 p.
4. Glynn D. G. The permanent of a square matrix [Electronic resource] / David G. Glynn // European journal of combinatorics. – 2010. – Vol. 31, no. 7. – P. 1887–1891. – Mode of access: <https://doi.org/10.1016/j.ejc.2010.01.010>
5. Gusfield D. Algorithms on strings, trees, and sequences: computer science and computational biology / Dan Gusfield. – Cambridge [England]: Cambridge University Press, 1999. – 534 p.
6. Heap B. R. Permutations by interchanges [Electronic resource] / B. R. Heap // The computer journal. – 1963. – Vol. 6, No. 3. – P. 293–298. – Mode of access: <https://doi.org/10.1093/comjnl/6.3.293>
7. Herlihy M. Art of multiprocessor programming / Maurice Herlihy, Nir Shavit. – [S. l.]: Elsevier Science & Technology Books, 2011.
8. Johnson D. S. Computers and intractability: A guide to the theory of NP-completeness / David S. Johnson, M. R. Garey. – New York: W. H. Freeman, 1983. – 340 p.
9. Johnson S. M. Generation of permutations by adjacent transposition [Electronic resource] / Selmer M. Johnson // Mathematics of computation. – 1963. – Vol. 17, No. 83. – P. 282–285. – Mode of access: <https://doi.org/10.1090/s0025-5718-1963-0159764-2>
10. Knuth D. E. The art of computer programming: generating all tuples and permutations / Donald Ervin Knuth. – 2nd ed. – Upper Saddle River [etc.]: Addison-Wesley, 2009. – 128 p.
11. Let's talk parallel programming with Stephen Toub [Electronic resource] // Microsoft Learn. – Mode of access: <https://learn.microsoft.com/en-us/shows/on-dotnet/lets-talk-parallel-programming-with-stephen-toub>
12. McCool M. Structured parallel programming: patterns for efficient computation / Michael McCool, James Reinders, Arch Robison. – [S. l.]: Elsevier Science & Technology Books, 2012. – 432 p.
13. Richter J. CLR via C# / Jeffrey Richter. – [S. l.]: Microsoft Press, 2012. – 894 p.
14. Task-based asynchronous programming – .NET [Electronic resource] // Microsoft Learn. – Mode of access: <https://learn.microsoft.com/en-us/dotnet/standard/parallel-programming/task-based-asynchronous-programming>
15. Task parallel library (TPL) – .NET [Electronic resource] // Microsoft Learn. – Mode of access: <https://learn.microsoft.com/uk-ua/dotnet/standard/parallel-programming/task-parallel-library-tpl>
16. Trotter H. F. Algorithm 115: perm [Electronic resource] / H. F. Trotter // Communications of the ACM. – 1962. – Vol. 5, no. 8. – P. 434–435. – Mode of access: <https://doi.org/10.1145/368637.368660>
17. Turbal Y. V. Permanent decomposition algorithm for the combinatorial objects generation [Electronic resource] / Y. V. Turbal, S. V. Babych, N. E. Kunanets // Radio electronics, computer science, control. – 2022. – No. 4. – P. 119. – Mode of access: <https://doi.org/10.15588/1607-3274-2022-4-10>
18. Valiant L. G. The complexity of computing the permanent [Electronic resource] / L. G. Valiant // Theoretical computer science. – 1979. – Vol. 8, no. 2. – P. 189–201. – Mode of access: [https://doi.org/10.1016/0304-3975\(79\)90044-6](https://doi.org/10.1016/0304-3975(79)90044-6)

НЕЙРОІНФОРМАТИКА ТА ІНТЕЛЕКТУАЛЬНІ СИСТЕМИ

NEUROINFORMATICS AND INTELLIGENT SYSTEMS

UDC 004.9

LOGIC-ONTOLOGICAL RECONSTRUCTION OF SCIENTIFIC DISCOURSE AND ITS IMPLEMENTATION IN AN AI-BASED REVIEWING SYSTEM

Bedratyuk L. P. – Dr. Sc., Professor, Head of the Department of Software Engineering, Khmelnytskyi National University, Khmelnytskyi, Ukraine.

ABSTRACT

Context. The growing number of scientific publications and the emergence of tools based on large language models (LLMs) highlight the need for automated verification of the structural quality of scientific texts. Most existing solutions focus on surface-level linguistic analysis and do not account for logical-discursive integrity – specifically, whether the text includes a hypothesis, method, results, conclusions, and whether these elements are connected by normative relationships.

Objective. The aim of this study is to develop an ontology-driven approach for the formalized verification of scientific text structures by constructing an ontological knowledge graph and evaluating its compliance with a predefined normative model of scientific discourse.

Method. A model is proposed based on two interrelated ontologies: “Scientific Publication” (defining node types and their roles) and “Reviewing” (defining logical-discursive requirements). The text is represented as a graph where nodes are formed through semantic markup using an LLM, and connections are verified according to a set of normative rules. A specialized GPT agent capable of dynamically applying ontological knowledge during analysis and review generation is employed for implementation.

Results. The model enables automatic detection of discourse structure violations: the absence of key elements, logical discontinuities, substitution of scientific novelty with practical significance, and incorrect interpretation of results. The proposed metrics quantitatively capture the level of structural completeness and consistency. Provided examples of graphs and reviews demonstrate that the system can detect non-obvious, latent logical inconsistencies even in formally complete texts.

Conclusions. The scientific novelty of the study lies in introducing the ontological graph as an interpretable model of scientific argumentation, used in tandem with a large language model. The practical significance lies in establishing a foundation for semi-automated reviewing, structural analysis of publications, and academic writing training. The methodology is scalable to other genres of scientific texts and can potentially be integrated into editorial platforms.

KEYWORDS: ontology, ontological graph, GPT agent, reviewing, semantic analysis, scientific publication, logical-discursive structure, machine learning, algorithms.

ABBREVIATIONS

BFO is a Basic Formal Ontology;
COPE is a Committee on Publication Ethics;
GPT is a Generative Pre-trained Transformer;
IMRaD is A standard structure of academic papers, dividing them into Introduction, Methods, Results, and Discussion sections;
LLM is a Large Language Model;
UFO is an Unified Foundational Ontology.

NOMENCLATURE

A is an artificially generated texts with controlled defects;
 B is a publications from journals excluded from scientometric databases;
 C is a benchmark articles;
 E is a set of directed edges;
 FN is a set of articles in which actual defects were missed;
 FP is a set of articles in which non-defective elements were incorrectly flagged as defective;

G is a ontological knowledge graph;
 L is a node labeling function;
 S_i is a i -th fragment of a scientific text;
 S is an ordered set of text fragments;
 t_i is a i -th type of a node;
 v_i is a i -th node in an ontological graph;
 V is a set of nodes in an ontological graph;
 V_{expected} is a set of ontological nodes expected in a reference annotation;
 V_{correct} is a set of ontological nodes correctly identified by the system;
 $\mathcal{R}$ is a set of admissible relation types between nodes;
 T is a set of ontological terms;
 TN is a set of articles correctly identified as having no defects;
 TP is a set of articles in which defects were correctly detected;
 Φ is a set of logical-discursive rules;

$\phi_i(G)$ is a value of a Boolean predicate;
 Δ is a set of violated rules.

INTRODUCTION

In the context of a growing volume of scientific publications and the widespread adoption of large language models (LLMs), the issue arises not only of ensuring linguistic correctness but also of maintaining the structural and logical coherence of scientific discourse. Peer review, as a key instrument of scholarly communication, requires significant time and cognitive resources, especially given the thematic diversity and the increasing number of manuscript submissions. This underscores the need for tools supporting preliminary or automated assessment based on a formal structural model of a scientific text.

Most existing solutions focus on superficial analysis or review generation, without considering the logical-discursive structure of the publication. They lack an explicit knowledge model of a scientific article as a system of interconnected semantic components – namely, hypothesis, method, results, and conclusions. Attempts to integrate ontologies into the analysis of scientific texts remain largely theoretical and do not cover the full pipeline – from semantic markup to the generation of a verifiable review. In contrast, modern LLMs enable the integration of generative capabilities with formal ontologies, paving the way toward interpretable and logically controlled reviewing systems.

The object of this study is the process of expert evaluation of scientific publications based on their logical-discursive structure.

The subject of the study is an ontological model of a scientific publication represented as a knowledge graph $G = (V, E, L)$ with fixed node types and normative links, as well as its integration into an automated reviewing process using a GPT-based agent.

The goal of this work is to design and implement an ontology-driven model for the analysis of scientific texts, which allows for the automatic generation of a preliminary review by detecting structural completeness, logical consistency, scientific novelty, argumentative soundness, and typical markers of pseudo-scientific imitation.

To achieve this goal, the following tasks must be accomplished:

- to construct an ontology of a scientific publication that reflects key nodes and the logical dependencies among them;
- to formulate a reviewing ontology as a set of predicates and normative relationships;
- to develop a review template aligned with both ontologies;
- to implement a GPT-based agent that performs semantic annotation, graph construction, and logical validation;
- to conduct an experiment with a controlled corpus and evaluate the accuracy of defect detection;
- to assess the applicability of the proposed approach in editorial and educational contexts.

The proposed approach combines the formal semantics of ontologies with the flexibility of generative language models, forming a new class of hybrid systems for structural-logical reviewing. The developed system not only identifies key structural elements of the text but also evaluates their coherence, scientific relevance, and compliance with the standards of academic reasoning. Such a tool can be useful to editors and reviewers of scientific journals for pre-screening submitted manuscripts, detecting potential structural flaws, and ensuring a more objective and rapid evaluation process. The implementation as a standalone GPT agent and the possibility of API integration open up perspectives for embedding the system into editorial platforms or using it in educational settings – particularly in courses on academic writing and scholarly communication.

The GPT agent is implemented within the ChatGPT environment and can assist journal editors and reviewers in the preliminary screening of submitted manuscripts, detection of potential structural deficiencies, and facilitation of more objective and efficient assessments.

1 PROBLEM STATEMENT

This study addresses the problem of logical-ontological analysis of a scientific article, represented as an ordered set of textual fragments $S = \{s_1, s_2, \dots, s_n\}$,

where each fragment s_i is a logically complete unit of scientific discourse. The ultimate objective is to construct a graph-based representation of the article's content in the form of an ontological graph $G = (V, E, L)$, where the set of nodes V corresponds to conceptual units of the text, the set of edges $E \subseteq V \times V$ reflects logical and argumentative relationships, and the labeling function $L: V \rightarrow T$ assigns each node an ontological type from a predefined set T (e.g., Hypothesis, Method, Result, Conclusion).

The input to the model includes the article text S , the set of ontological terms, and a set of verification rules $\Phi = \{\phi_1, \dots, \phi_k\}$, each represented as a Boolean function applied to the graph G , enabling assessment of structural completeness and logical coherence.

The expected output of the task is an ontological graph G that reflects the logical-discursive structure of the scientific article. Additionally, the truth value $\phi_i \in \Phi$ is computed for each rule, and an aggregate consistency metric is calculated as the arithmetic mean of all fulfilled rules:

$$\text{Consistency}(G) = \frac{1}{k} \sum_{i=1}^k \phi_i(G).$$

The problem is formalized with consideration of several constraints. First, each textual fragment may be assigned no more than one ontological type from the set CCC.

Second, the graph must be connected or consist of connected components that include nodes of the key types – Hypothesis, Method, Result, and Conclusion. Third, the relations between nodes must conform to the

admissible links defined in the corresponding ontology, such as Method validates Hypothesis, Result supports Hypothesis, and Conclusion follows from Result.

2 REVIEW OF THE LITERATURE

The issue of knowledge formalization in the form of ontologies has become the subject of intensive research in philosophy, logic, and computer science. The fundamental principles of the ontological approach were formulated in works [1] and [2], which proposed upper-level ontologies (in particular, UFO) that formed the basis for semantic modeling in systems analysis. These works devote significant attention to the rigor of concept classification and the relationships between them. An extension toward natural language analysis is proposed in [3], where natural language ontology is considered as a means of formalizing the meaning of linguistic units within a philosophical-semantic context.

The issues of the truth of statements, truth-bearers, and the structure of argumentation are analyzed in works [4,5], which propose approaches for formalizing a logical-semantic model of text aimed at evaluating the soundness of scientific claims. The application of these ideas in domain-specific computational modeling is also supported in [6], where the use of tropes and semantic spaces is suggested for refining concepts within models.

Technical aspects of automatically deriving ontologies from texts are addressed in [7], which describes the capabilities and limitations of ontology learning methods from text, especially relevant in the case of scientific publications.

A separate group of studies focuses on the application of large language models (LLMs), particularly GPT, for generating reviews and evaluating scientific articles. In [8], the use of GPT for preliminary reviewing is proposed, though without clear logical-structural control over the text. A similar approach is discussed in [9], which demonstrates that LLMs can produce stylistically convincing, yet often superficial reviews.

The evaluation of factual consistency of claims in scientific publications is implemented in [10], where it is verified whether a specific claim is supported by data. However, this model does not assess the structure of scientific discourse as a coherent argumentative entity. An extended study of inconsistency risks in language models such as GPT-3 was conducted in [11], which highlights potential errors in generating complex logical dependencies. In [12], the possibility of using language models to reduce information asymmetry is discussed, yet these approaches do not operate with a structured representation of knowledge and do not include logical-discursive validation of argumentation at the level of a scientific publication, as proposed in our system.

The use of automatic indices and services for analyzing scientific output is also well known, such as Microsoft Academic Service [13], which constructs scientific knowledge graphs but does not perform logical validation of argumentative coherence.

Institutional issues of scientific publication quality are reflected in the lists of discontinued Scopus journals [14], where the reasons for discontinuation are often associated with the absence of proper peer review or with the formal character of publications. An additional ethical guideline is provided by the COPE recommendations [15], which emphasize the necessity of logical consistency and structural completeness in the creation and reviewing of scientific texts.

In this context, our study continues the line of research combining ontologies and language models, but for the first time implements them as a formalized logic-ontological structure applied to the full text corpus for the purpose of detecting structural deficiencies. This makes it possible not only to assess factual accuracy or style, but also to formally verify the presence and coherence of elements of scientific argumentation.

3 MATERIALS AND METHODS

In this study, the object of analysis is a scientific article composed of textual fragments corresponding to the main logical-discursive components: hypothesis, method, results, conclusions, and so forth. The expected outcome is a structured representation of the article in the form of a knowledge graph, which includes formalized nodes and links between them according to the constructed ontology, as well as an analytical assessment of the integrity, completeness, and logical coherence of the scientific argumentation, expressed through a set of quantitative metrics.

The methodology is based on the principle of multi-level system integration of three interrelated components: a formal ontology, logical-discursive verification, and a large language model (LLM). The formal ontology describes the semantic structure of the scientific text in the form of a set of terms and admissible relations among them. The verification logic defines a set of rules that must be fulfilled for the publication to be considered structurally complete. The LLM provides semantic alignment between text fragments and ontological types, as well as models the logical dependencies among them. In this way, the language model functions as an execution mechanism that performs structural analysis and logical validation in accordance with the formal problem specification.

The overall workflow of the study is implemented as a sequence of interconnected stages. At the first stage, the publication text undergoes preprocessing: segmentation into logical fragments (introduction, methods, results, etc.) is performed. At the second stage, each fragment is assigned an appropriate ontological type using the language model and predefined instructions. Then, an ontological knowledge graph is constructed, in which nodes correspond to fragments and directed edges represent logical relations (e.g., “method validates hypothesis”, “conclusion is based on results”). At the next stage, the graph is checked against a set of logical-discursive rules implemented as Boolean predicates. After this, the system computes an aggregated score of logical and structural coherence. The final stage is the generation of a textual review

based on the identified nodes, assessed relations, and detected violations, integrated into a template of an automated report.

This approach ensures control over the structure of the study, reproducibility of the reviewing process, and the potential scalability of the system to other domains of knowledge or types of texts.

Ontology, in the context of this study, is defined as a formal, explicit, and machine-processable specification of a set of concepts and the relationships among them, which describes the structure of knowledge within a specific subject domain. Unlike a taxonomy or glossary, an ontology defines not only a hierarchy of concepts, but also the types of connections, logical dependencies, and constraints that operate between conceptual units.

The goal of constructing an ontology of a scientific publication is to formalize the logical-discursive structure of the text, that is, to identify the key semantic nodes that are necessarily present in a scientific argument and to fix the normative relationships among them. The ontology is implemented as a set of node types (terms) and admissible relations between them, forming a directed semantic graph. A typical set of terms includes such ontological nodes as Hypothesis, Goal, Method, Object, Result, Conclusion, Novelty, Practical significance, and so forth. In the future, individual nodes may be refined with additional attributes (e.g., structure, formal features of formulation, stylistic characteristics, etc.).

Admissible relationships between types of ontological nodes are defined by a set $\mathfrak{R}$, which is explicitly specified as a list of tuples of the form $(t_i, t_j, type)$. Each element $r \in \mathfrak{R}$ describes a permitted normative relation between a node of type t_i and a node of type t_j , with the type of connection indicated. For example, the tuple (“Method”, “Hypothesis”, “validates”) is interpreted as “the method validates the hypothesis”, while (“Conclusion”, “Result”, “justified_by”) means “the conclusion is justified by the results.”

The types of relations are formalized as predicates such as *validates*, *supports*, *derived_from*, *motivates*, *contradicts*, and others. Each of these predicates takes two arguments: the first is the source node, and the second is the target node. For example, the predicate *validates*(Method, Hypothesis) indicates that the method validates the hypothesis, while *supports*(Result, Hypothesis) means that the result supports the hypothesis.

The set $\mathfrak{R}$ is part of the input ontological model provided to the agent along with the set of predicates Φ used for logical-discursive validation. This set is editable and can be adapted to the requirements of a particular journal or research paradigm. It defines the expected structure of scientific argumentation, according to which knowledge graphs are constructed and semantic evaluation of texts is performed.

The developed ontology is not universal – it is tailored to the format of academic articles, the IMRaD struc-

ture, and the recommendations of COPE. During the construction of the ontology, elements from foundational knowledge ontologies (such as BFO and UFO) were employed alongside specialized categories proposed within our model. A distinctive feature of the developed ontology is the inclusion of additional concepts that are typically not represented in the formal structure of scientific articles but are important for peer review. These include features such as scientific relevance, alignment with contemporary scientific discourse, linguistic neutrality, stylistic correctness, and the representativeness of the reference base.

Each ontology node is associated with a fragment of the article text. To achieve this, a labeling function $L: V \rightarrow T$ is introduced, which maps a node v_i from V (extracted fragments) to a type from the set T . All text fragments that are not assigned to a label are considered out-of-scope or noise. The construction of the ontological graph $G = (V, E, L)$ is performed as a result of text interpretation by the language model, which conducts semantic classification of fragments to define the labeling function L and establishes admissible links E according to the specified rules $\mathfrak{R}$.

The model allows for incompleteness: not all nodes may be detected; however, the presence of a certain minimal core is a prerequisite for the formal eligibility of a text for review. Thus, ontology serves as a structural scaffold that defines the expected architecture of scientific discourse and provides a basis for logical-structural validation.

Classical knowledge ontologies such as BFO and UFO are primarily oriented toward modeling universal types of entities and their properties in the real or conceptual world. Their advantages lie in a high level of abstraction, ontological rigor, and stability. However, when modeling the logical-discursive structure of a scientific text, this level of abstraction is insufficient to capture the dynamics of argumentation, semantic dependencies between text fragments, and elements of publication quality assessment.

Within the scope of our study, several essential extensions to the base ontological schema were introduced. First, dedicated nodes were added for concepts that are not structurally mandatory in a paper but are critically important for peer review: Scientific relevance, Practical significance, Embeddedness in contemporary scientific discourse, Novelty, Completeness of justification, and Interpretation of results. These nodes enable the system to track not only the presence of standard elements (hypothesis, method, result) but also the degree of their substantiation and interconnectedness.

Second, our focus shifts from the ontological hierarchy of types to the dynamics of relationships between them. Instead of classical typing with subclasses and object properties, we use a model of logical predicates over node pairs that capture discursive logic: for example, *supports*(H, R), *derived_from*(M, H), *justified_by*(C, R). In this way, the ontology em-

phasizes functional relations that reflect the logic of scientific argument construction.

Third, meta-concepts have been introduced, allowing not only structural elements but also their quality to be described. For instance, a node of type *Hypothesis* may have attributes such as explicit, testable, narrow vs broad. This makes it possible to combine structural analysis with content-based criteria evaluation.

These extensions exemplify the adaptation of foundational ontologies to the applied task of automated analysis of scientific texts for the purpose of assessing their logical-discursive completeness. The proposed modifications do not contradict the foundations of BFO or UFO but rather function as applied overlays tailored to the needs of modern scholarly communication and automated reviewing.

Unlike the general ontology of a scientific publication, which captures the structure of scientific discourse as such, the reviewing ontology formalizes the requirements for its individual components and defines criteria for logical validation of their presence, interconnection, and semantic completeness. It is used not to describe the article, but to evaluate it in accordance with expert standards of academic quality.

At the first level, such ontology contains a specification of requirements for each key node of a scientific publication. For the Hypothesis, the requirements include clarity of formulation, testability, connection to a problem situation, and justification of relevance. The Method must be related to the hypothesis, described in sufficient detail to enable reproducibility, and include a rationale for its selection. The Results must be obtained through the application of the method and presented in a form suitable determines whether the conclusions are connected to the results through an admissible relation of type **justified_by**.

In the implemented system, the set of predicates $\Phi = \{\phi_1, \dots, \phi_k\}$ is defined as an explicit collection of Boolean conditions formulated based on the ontological model (node types and admissible links) and the structure of the review template (logical- for analysis. The Conclusions should follow from the results, generalize them, and not merely restate the hypothesis verbatim.

In addition to these core nodes, the system separately checks for Novelty, Practical significance (applicability), Justification (sufficient argumentation), and Scientific relevance (connection to the current state of the problem in the global context).

All these requirements are formalized as a set of Boolean predicates:

$$\Phi = \{\phi_1, \phi_2, \dots, \phi_k\}, \phi_i : G \rightarrow \{0,1\},$$

where each ϕ_i is a logical-discursive rule that verifies whether a specific condition is met within the ontological graph GGG. For example, the rule $\phi_{\text{hypothesis_present}}$ checks for the presence of at least one node labeled with

the type Hypothesis; the rule $\phi_{\text{method_mathces_hypothesis}}$ verifies whether a logical connection of type **validates** exists between the method and the hypothesis; and the rule $\phi_{\text{result_supports_conclusion}}$ discursive requirements for the scientific text). Each rule ϕ_i checks the fulfillment of a specific condition in the graph $G = (V, E, L)$ – for example, the presence of a node of type Hypothesis, or a link of type **validates** between the method and the hypothesis. At the same time, the system supports the possibility of dynamically extending the set of predicates, considering the editorial policy of a journal, domain-specific features, or individual review settings. This enables flexible adaptation of the logical-discursive validation to the context of the submitted publication, while preserving the transparency and formalization of the process.

A key stage is internal self-verification: prior to generating the review, the agent explicitly computes the value of each rule $\phi_i(G)$ and creates a logical mask of violations. For each violated rule $\phi_i(G) = 0$, an explanation is generated, which is subsequently integrated into the corresponding section of the review. In this way, a link is established between structural validation and natural language justification, enhancing the interpretability of the system.

A violation of any rule from the set Φ is interpreted as a specific anomaly in the structure of the scientific text, indicating logical incompleteness, lack of argumentative coherence, or a missing key component of the publication. The totality of such violations forms the set:

$$\Delta = \{\phi_i \in \Phi \mid \phi_i(G) = 0\},$$

which represents a list of specific logical-discursive deficiencies. This set serves as the basis for constructing an interpretable review or a formalized analytical report.

The overall measure of a manuscript's compliance with the requirements of the reviewing ontology is defined as the proportion of fulfilled rules:

$$\text{Consistency}(G) = \frac{|\Phi| - |\Delta|}{|\Phi|},$$

where $|\Phi|$ is the total number of rules, and $|\Delta|$ is the number of violations. This metric is used as a numerical indicator of structural completeness and logical consistency.

Thus, the agent not only generates the text of the review but also performs a formalized logic-ontological analysis based on a normative model of scientific discourse. The proposed approach allows for the validation of both basic structural requirements (such as the presence of key elements) and more complex logical dependencies between them, for instance, the alignment of the method with the hypothesis, the justification of conclusions by results, or the substitution of scientific novelty with practical applicability.

The list of predicates is provided in the appendix as a table, representing the core of the set Φ – that is, the pri-

mary set of Boolean conditions used to assess the structural completeness and logical consistency of a scientific text. This core is passed to the agent as part of the input ontological structure for logic-discursive analysis. If necessary, the list of predicates can be extended or adapted according to the specific subject domain or the editorial policy of a particular scientific journal. Based on the results of predicate evaluation, the system generates a structured report that reflects the degree to which the publication satisfies the expected criteria of argumentativeness, completeness, and coherence of scientific reasoning.

The formalization of a scientific article's text is carried out through the construction of a semantic knowledge graph that models the logical-discursive structure of the publication. The source text S is treated as an ordered set of fragments S_i , which may contain hypotheses, methods, results, conclusions, and other components of scientific argumentation. These fragments are mapped to nodes of the graph $G = (V, E, L)$, which serves as the primary object of analysis. The graph $G = (V, E, L)$ is automatically constructed by the agent based on the article text and the ontological model, which includes the set of terms S and the set of admissible links $\mathfrak{R}$. The graph's nodes are created via semantic classification of text fragments, while the edges are formed by checking for the presence of discursive dependencies according to the ontology.

In the implemented system, the resulting graph can be exported in a standard format (e.g., .json or .dot) or visualized. This enables transparent inspection of the structure G , in particular – the presence of all key nodes, the types of relations, and the possibility of evaluating the logic-discursive predicates. The graph G , serves as the foundation both for structural analysis and for the generation of the review.

The set $V = \{v_1, v_2, \dots, v_n\}$ is the set of nodes in the graph, where each node v_i corresponds to a specific text fragment S_i . Each node is assigned an ontological type using a labeling function:

$$L : V \rightarrow T,$$

where T is the set of ontological terms (such as Hypothesis, Method, Result, Conclusion, Novelty, Significance, etc.). This labeling is performed automatically based on text analysis using a GPT-based language model, which generates the labeling function LLL through semantic alignment of fragments with the definitions of ontological nodes.

The set $E \subseteq V \times V$ defines the logical connections between nodes. Each edge $(v_i, v_j) \in E$ represents a relation of discursive dependency, for example, “the method validates the hypothesis” or “the conclusion is based on the result”. The type of relation is specified by a separate function that assigns predicate-level semantics to the link, according to the reviewing ontology. Thus, the graph

GGG is not merely a textual schema but a structured representation of the argumentative logic of the publication.

The formalized graph enables verification of the presence of all required structural components and the logical dependencies among them. For this purpose, the previously defined numerical model of structural consistency – Consistency Score – is used.

Thanks to the construction of the knowledge graph G , it becomes possible not only to detect formal deficiencies in the text but also to perform automated generation of a review based on the structure of the graph, its compliance with the expected logic, and the violations identified. This approach allows for a shift from surface-level analysis of style and content to formalized semantic verification of the scientific text.

The construction of the knowledge graph G plays a central role in the architecture of the proposed system: it provides a formal transition from the linguistic representation of the text to a structured ontological model suitable for logical-discursive analysis. The graph structure makes it possible not only to record the presence or absence of key elements but also to verify their logical consistency. This, in turn, serves as the basis for computing the Consistency metric and generating an explainable review. Thus, formalization in the form of a graph transforms the evaluation of a scientific text from a heuristic procedure into a formalized and replicable operation.

The constructed graph $G = (V, E, L)$, in this context, is not merely a knowledge graph in the general sense but a specialized ontological graph, since all its elements – both nodes and edges – are formed in accordance with a predefined ontology. Specifically, the node types are restricted to a set of terms T (e.g., Hypothesis, Method, Result), and the links are limited to a set of admissible relations $\mathfrak{R}$. This representation not only preserves the semantic structure of the text but also enables validation of its compliance with expected scientific standards using the set of predicates Φ .

To validate the developed logic-ontological model for analyzing scientific publications, a specialized test corpus was constructed, including texts with deliberately introduced logical defects. This allows for both quantitative and qualitative evaluation of the system on texts for which the violated structural or logical elements are known in advance and thus provides a basis for assessing the model's ability to detect them.

The first part of the corpus consists of artificially generated texts in which one or more structural components were deliberately omitted or altered. Specifically, texts were created without a hypothesis, without a method, with results that do not support the hypothesis, with conclusions that do not generalize the results, with confusion between practical significance and scientific novelty, and with argumentation limited to local context without properly addressing the broader scientific discourse. These texts maintain syntactic correctness, an academic-like style, and a length typical of short articles. They allow precise evaluation of whether the system can detect the

absence of key nodes, logical inconsistencies, or argumentation defects without confusing them with stylistic variation.

The second part of the corpus includes real-world publications from journals that were excluded from reputable international databases due to systematic violations of academic integrity. These publications are treated as a heuristic negative class – a source of texts that may appear formally scientific but typically contain violations of logical-discursive structure, lack of justification, missing hypotheses, misuse of methods, or overgeneralized conclusions.

The initial documents were obtained in PDF and DOCX formats, then converted to plain text to ensure uniformity of the input data. The texts were further preprocessed to standardize formatting, remove unnecessary breaks, and conform to a unified structure. For each document, a manual or semi-automated annotation was created in JSON format, which includes information about the expected structural nodes, their presence or absence, and known logical-discursive defects. These annotations serve as a reference for computing error detection accuracy and validating the model.

Thus, the constructed corpus enables comprehensive testing of the proposed approach using both real and synthetic examples with varying levels of structural and argumentative quality. It also ensures reproducibility of the experiment and independent verification of the model's effectiveness on external material.

To evaluate the effectiveness of the proposed ontology-driven approach to scientific publication analysis, a set of quantitative metrics was used to assess how accurately the system identifies structural nodes, logical dependencies between them, and generates adequate reviews. These metrics incorporate both classical information retrieval indicators and specialized metrics for ontological analysis.

Several key metrics are used to assess the quality of the system's performance. In particular, precision is defined as the proportion of correctly identified defects among all defects flagged by the system. It is calculated using the formula:

$$\text{Precision} = \frac{|TP|}{|TP| + |FP|},$$

where $|TP|$ is the number of true positives (correctly detected violations), and $|FP|$ is the number of false positives.

Recall characterizes the system's ability to detect all existing defects and is defined as the proportion of detected defects among all those that are actually present:

$$\text{Recall} = \frac{|TP|}{|TP| + |FN|},$$

where $|FN|$ is the number of defects that the system failed to detect.

Another important metric, in addition to the Consistency Score, is the Onto-score, which reflects the system's ability to correctly identify key ontological nodes (such as Hypothesis, Method, Result, Conclusion, etc.). This metric is computed as the ratio of correctly identified nodes to the total number of expected nodes:

$$\text{OntoScore} = \frac{|V_{\text{correct}}|}{|V_{\text{expected}}|}.$$

For each test article, the automatically constructed ontological graph is compared with a reference structural model that defines the expected presence and types of nodes according to the ontology. Compliance is assessed by comparing node types (for the Onto-score) and by evaluating the fulfillment or violation of logical rules (for the Consistency metric).

The results of the experiments are recorded in tabular format, where each row corresponds to an article and includes: its identifier, the number of detected nodes, the number of violated rules, and the values of the Precision, Recall, Consistency, and Onto-score metrics. For visualization, bar charts are used (to compare Precision and Recall across articles), heatmaps are used to show the frequency of detection of specific defect types, and pie charts provide an overview of the review results.

Thus, the result verification process is quantitative in nature, ensures reproducibility, and makes it possible to clearly identify the strengths and weaknesses of the ontology-driven reviewing system.

4 EXPERIMENTS

The objective of the experiment was to test the hypothesis that the implemented system is capable of automatically detecting structural and logical-discursive defects in scientific articles based on an ontological knowledge graph, without human intervention.

For each article, the following sequential steps were performed:

- (1) loading and cleaning the text;
- (2) segmenting the text into logical fragments;
- (3) semantically labeling each fragment with an ontological node type using the publication ontology via the GPT agent;
- (4) constructing the graph $G = (V, E, L)$;
- (5) verifying compliance with the set of predicates Φ ;
- (6) calculating the metrics: Precision, Recall, Consistency, and Onto-score;
- (7) automatically generating a draft review.

The processing steps were implemented as sequential functions of the GPT agent, aligned with the logic of distinct modules: semantic labeling, graph construction, rule validation, and review generation.

In total, 120 texts were analyzed, divided into three subsets of 40 texts each:

A – artificially generated texts with controlled defects (removed nodes, broken links);

B – publications from journals excluded from scientometric databases (formally structured but often lacking novelty or argumentative rigor);

C – benchmark articles selected from open-access publications in Q1-category journals indexed in the Scopus scientometric database.

All publications were processed using a single ontological JSON template, in which the set of node types, admissible relations, and validation rules were manually defined. This ensured a unified analysis structure and consistent comparison of results.

The experiments were conducted in the Jupyter Notebook environment (Python 3.10). For graph construction, the networkx library was used; for table processing, pandas; for metrics calculation, scikit-learn; and for visualization, matplotlib and seaborn. The GPT agent was implemented on the ChatGPT GPTs platform (without using an external API), utilizing the ontological structures and review template as initial context.

Validation of the graph's compliance with the reviewing ontology requirements was performed through the logic-ontological core – a module that implements a set of Boolean predicates $\phi_i : G \rightarrow \{0,1\}$ to evaluate the satisfaction of each structural condition.

For each article, a table was generated showing the number of nodes, fulfilled predicates, and computed metrics. The following visualizations were produced:

- (1) bar charts for Precision and Recall across all classes;
- (2) boxplots to compare metrics between classes;
- (3) heatmaps of rule fulfillment across all documents;
- (4) sample ontological graphs for individual articles.

This enabled assessment of the system's quality, robustness, and sensitivity to different types of defects.

5 RESULTS

To conduct the experiments, a corpus was constructed consisting of three subsets of scientific publications, differing in structural completeness and the expected quality of logical-discursive content. General statistics are presented in Table 1. Each subset contains the same number of documents (40), ensuring symmetry for comparative analysis. The average number of extracted fragments (nodes) in documents from subset C was 7.6, corresponding to the typical structure of a full-length scientific article. In subset B, this value was lower – 5.1, and in subset A – 4.2, indicating the presence of structural omissions.

For each subset, the detection frequency of four key ontological nodes was evaluated: Hypothesis, Method, Result, and Conclusion. As shown in Figure 1, in the benchmark articles (subset C), these nodes were identified with high consistency (in 95–100% of cases). In subset B, a significant reduction was observed in the presence of Conclusions (down to 60%) and Methods (down to 55%), whereas in subset A, where the defects were deliberately introduced, the frequency of specific node types could be below 40%.

The obtained data reveal significant differences in structural completeness among the article classes, providing a foundation for further analysis of the effectiveness of logic-ontological verification and the calculation of quality metrics.

One of the key stages of the system's operation is the automatic semantic labeling of text fragments using a large language model in accordance with the ontological schema. To evaluate the quality of this labeling, precision and recall were calculated for each of the main node types: Hypothesis, Method, Result, and Conclusion.

Table 1 – Distribution of Documents Across Corpus Subsets

Parameter	A (Synthetic)	B (Discontinued Journals)	C (Benchmark)
Number of documents	40	40	40
Average number of nodes	4.2	5.1	7.6
Standard deviation	1.1	1.4	0.9

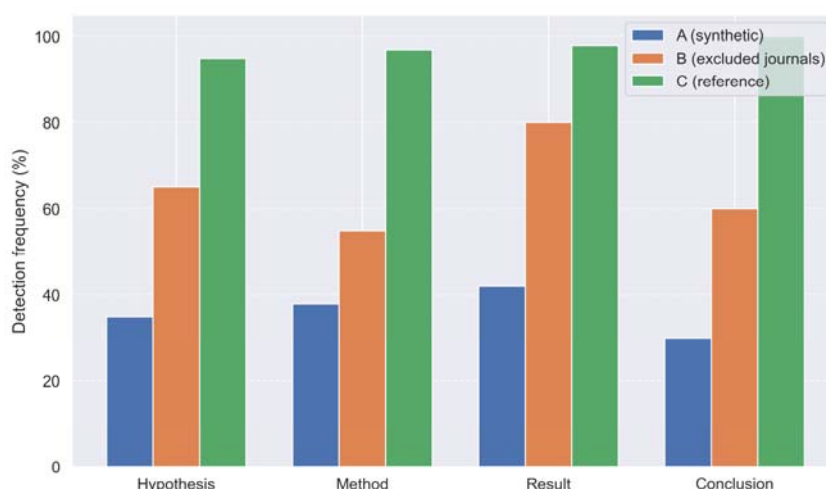


Figure 1 – Frequency of Detection of Core Ontological Nodes Across Corpus Subsets

In the benchmark article class (C), the labeling precision exceeds 0.90 for all four node types, indicating the model's ability to correctly identify the discursive structure of full-format scientific texts. In contrast, subset B shows a decrease in both precision and recall, especially for Result (Precision = 0.74, Recall = 0.68) and Conclusion (Precision = 0.70, Recall = 0.64). The lowest scores were observed in the artificially deformed texts (subset

A), where, for example, conclusion elements were deliberately omitted or replaced with general statements. Here, the precision for Conclusion was only 0.48, and recall – 0.37.

The labeling quality evaluation is presented in Table 2. The values were computed based on comparison with the reference annotation, which included manual labeling of all expected nodes.

Table 2 – Precision and Recall Metrics for Ontological Categories (Averaged Across Subsets)

Node	A		B		C	
	Precision	Recall	Precision	Recall	Precision	Recall
Hypothesis	0.61	0.58	0.78	0.75	0.93	0.91
Method	0.67	0.63	0.80	0.78	0.94	0.92
Result	0.54	0.49	0.74	0.68	0.91	0.89
Conclusion	0.48	0.37	0.70	0.64	0.92	0.90

Examples of constructed knowledge graphs based on semantic labeling are presented in Fig. 2. The left part illustrates the structure of a benchmark article, in which all key ontological nodes are present: Goal, Hypothesis, Method, Result, Conclusion, and the relations between them are clearly defined. The right part shows the structure of an artificially generated publication in which the Result node was deliberately omitted. As a result, the graph appears fragmented and lacks a logical transition to the conclusion, which was accurately detected by the system.

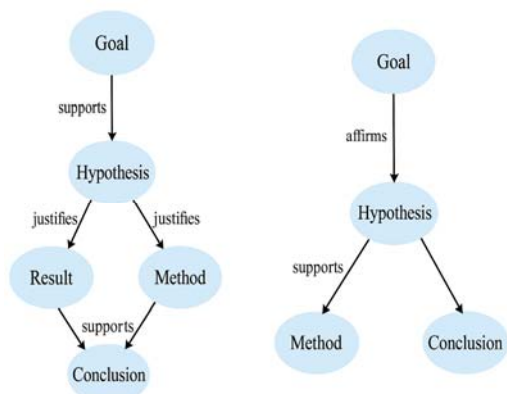


Figure 2 – Knowledge Graphs for a Benchmark Article (left) and a Synthetic Article (right)

These results indicate the model's sufficient capability to automatically recognize the key structural elements of a scientific article. At the same time, the analysis shows that the labeling process is sensitive to the quality of the input text and can detect deviations even when the discursive structure is weakly expressed.

The next stage of evaluating the effectiveness of the ontology-driven model was to determine the degree of compliance of the knowledge graph with the discursive rules defined by the reviewing ontology. To this end, each constructed graph $G = (V, E, L)$ was checked against a set of rules $\Phi = \{\phi_1, \phi_2, \dots, \phi_k\}$, which represent the expected relations between key nodes.

The results revealed a significant difference across the three subsets. In class C (benchmark articles), the average proportion of satisfied rules was 92%, in class B (problematic real-world articles) 71%, and in class A (artificially defective) – only 55%. Thus, the Consistency index reflects the degree of alignment between the internal logic of an article and the formal standards of scientific exposition.

The most frequently violated rules were:

- Method validates Hypothesis – absence of or weak linkage between the method and the hypothesis;
- Conclusion justified by Result – formal conclusions that are not based on the results;
- Hypothesis motivated by Literature Review – hypotheses not substantiated in the introduction or literature review.

Fig. 3 presents a heatmap showing the frequency of rule satisfaction across the different subsets. Each row corresponds to a specific discursive rule, and each column – to a class of articles. The visualization clearly shows consistent rule fulfillment in the control group, whereas synthetic defects systematically disrupt basic relations. Class B exhibits a more heterogeneous pattern: some articles are structurally complete, while others demonstrate significant logical deficiencies.

The obtained results demonstrate the sensitivity of the ontological model to implicit violations in the logic of scientific exposition. Even when the required fragments are formally present, the system is able to detect broken or missing links between them – an essential capability for preliminary review and recommendations for revision.

To quantitatively assess structural completeness and compliance with the logical-discursive structure, two integral metrics were used: Consistency and Onto-score. The first reflects the proportion of satisfied discursive rules from the set Φ within the constructed ontological graph, while the second indicates the proportion of correctly identified structural nodes relative to the expected reference set. In subset C (benchmark articles), the average Consistency score was 0.92 and the Onto-score was 0.94, indicating a stable structure and full presence of key discursive components. In subset B, both metrics showed

a noticeable decline: 0.71 for Consistency and 0.76 for Onto-score. The lowest values were recorded in subset A, where due to controlled defects the metrics did not exceed 0.56 for Consistency and 0.61 for Onto-score.

Fig. 4 presents a comparison of the distributions of Consistency and Onto-score metrics across the three classes using boxplot diagrams. It can be observed that benchmark articles exhibit not only high average values but also low dispersion, while subset B demonstrates a

widespread, indicating varying degrees of structural completeness even within the same class.

In addition, Fig. 5 presents histograms of the Consistency score distribution for all documents in the corpus. This type of visualization makes it possible to assess not only the average but also the distribution density of the metric values within each class. In subset A, most values are concentrated in the interval 0.4-0.6, whereas in class C, the majority of articles exhibit Consistency scores above 0.9.

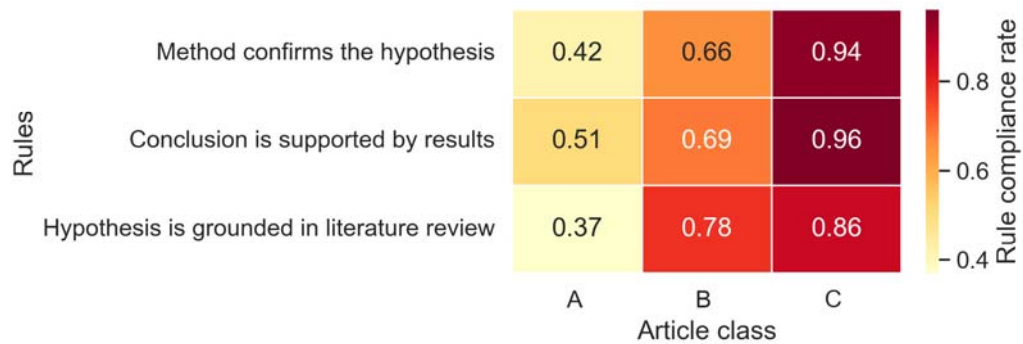


Figure 3 – Heatmap of Logic-Discursive Rule Satisfaction Across the Three Corpus Subsets

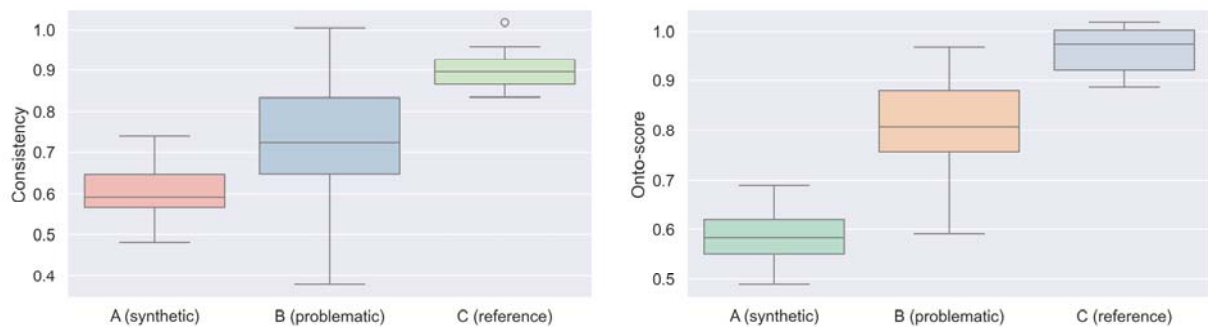


Figure 4 – Boxplots for the Consistency and Onto-score metrics across three subsets

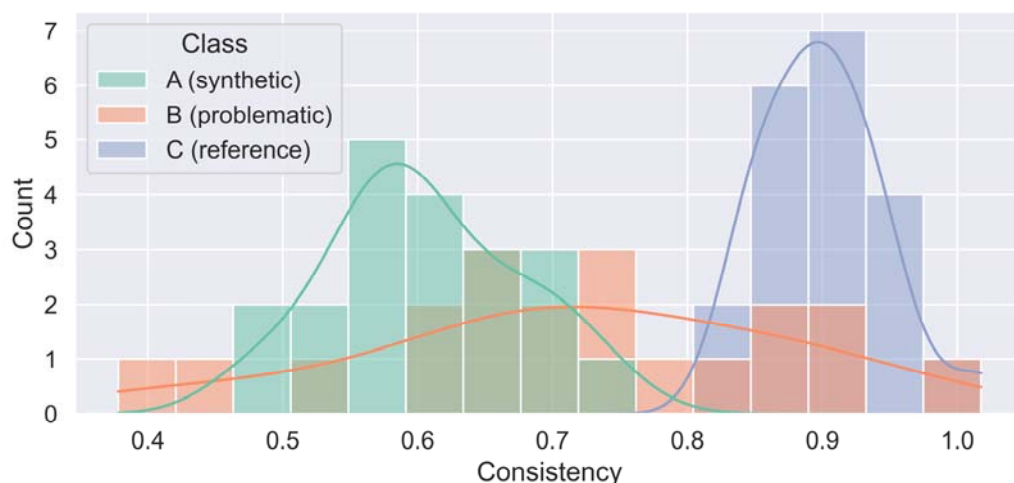


Figure 5 – Histogram of Consistency Score Distribution for All Corpus Articles

The obtained results confirm that both metrics are informative criteria for the preliminary automated assessment of the structural quality of scientific publications. They effectively distinguish between text classes and can serve as quality indicators even in the absence of expert input.

To evaluate the system's ability to accurately detect structural defects in scientific publications, classical information retrieval metrics – precision and recall – were employed. Precision indicates the proportion of truly detected violations among all automatically flagged ones, while recall characterizes the system's ability to identify all defects actually present in the text according to the reference annotation.

In subset A, where the defects were deliberately embedded in the text, the system achieved a precision of 0.82

and recall of 0.79, demonstrating its ability to detect even subtly masked structural issues. In subset B, which contains real articles with potentially poor structure, the values were lower – 0.74 for precision and 0.66 for recall. In the control class C, the system detected only minor deviations – mostly false positives – leading to a reduction in precision to 0.71, while recall remained high at 0.93.

Fig. 6 presents a bar chart comparing the average values of Precision and Recall across each subset. As shown, subsets A and B exhibit similar trends, although the defects in subset A are detected more consistently, due to their formalized nature. In subset C, Recall dominates, reflecting a scenario in which the system tends to “over-react,” responding even to stylistic deviations. Edge cases were analyzed separately.

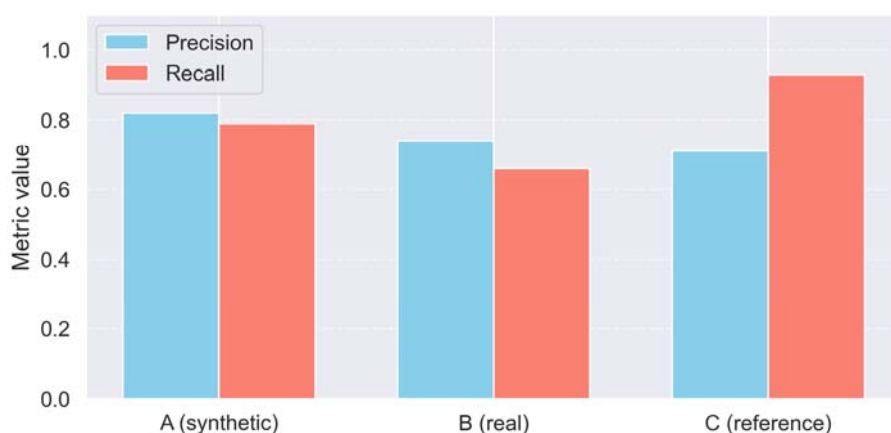


Figure 6 – Comparison of Precision and Recall Across Subsets

The highest precision score 1.0 was recorded for a synthetic article from subset A, in which three critically important nodes (Method, Result, Conclusion) were simultaneously removed. This allowed the system to clearly detect the defects without generating false positives. Conversely, the lowest precision score 0.54 was observed in a real article from subset B, where the system mistakenly interpreted general phrases as structural nodes, resulting in a large number of false positives.

Thus, the use of Precision and Recall metrics not only allows for the evaluation of overall defect detection performance but also reveals specific scenarios in which the model is either too conservative or overly “sensitive” to semantic variation.

To gain deeper insight into the typical types of defects detected by the system, a pie chart was constructed (Fig. 7) showing the distribution of missing ontological nodes in subset A – that is, in the artificially deformed texts. The chart complements the Recall metric: each segment represents the proportion of missing nodes for a given ontological type.

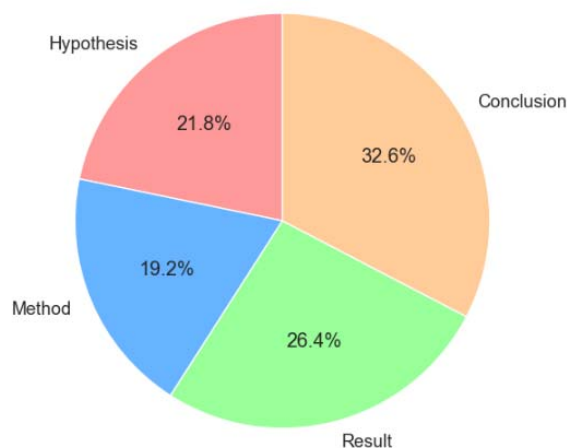


Figure 7 – Distribution of Missing Ontological Nodes in Subset A (Based on Recall Metric)

As shown in the diagram, the largest proportion of missing nodes corresponds to the Conclusion type (63%), indicating the difficulty in identifying logical closures in deformed or overly generalized statements. This is followed by Result (51%) and Hypothesis (42%), which points to challenges in detecting research justification and outcome representation. The fewest losses were recorded

for the Method type (37%), which aligns with the fact that methodological descriptions often have a distinct stylistic structure.

This visualization allows not only for the quantitative assessment of Recall but also for the localization of the most problematic areas in article structures, which is important for further improvement of the recognition model.

To qualitatively illustrate the system's operation, typical examples of knowledge graphs are provided, which in this context are constructed as ontological graphs. Each such graph $G = (V, E, L)$ is built based on the article text and a predefined ontology. It reflects the logical-discursive structure of the article: the nodes V correspond to semantic fragments, the labeling function L assigns them ontological categories, and the edge set E forms the normative logic of argumentation.

This subsection presents one graph for each of the corpus subsets (A, B, C), enabling visual comparison of typical deviations – from structural violations to logical incompleteness and full compliance.

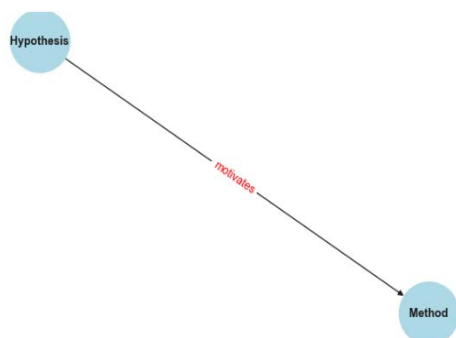


Figure 8 – Ontological Graph for an Article with Controlled Structural Defects (Subset A)

Fig. 8 shows an example graph for an article from subset A, in which the Result and Conclusion nodes were deliberately omitted. As a result, the Hypothesis and Method remained without logical closure: relations of type *derived_from* and *justified_by* are missing. The graph appears fragmented, with a broken argumentative chain, which was correctly detected during logic-discursive validation.

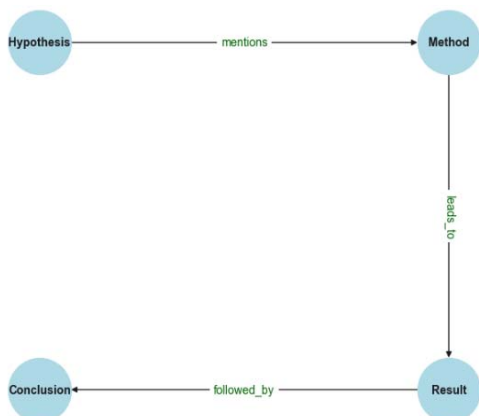


Figure 9 – Graph for a Problematic Article with Violations in Logical Structure (Subset B)

Fig. 9 illustrates an example graph for a publication from subset B. Unlike the previous case, all four key nodes are present here; however, their logical coherence is violated. In particular, the method is not specific to the hypothesis, and the results are not connected to the conclusions. As a result, the graph is formally complete but logically flawed: the *validates* and *justified_by* edges are either missing or lack meaningful justification in the text.

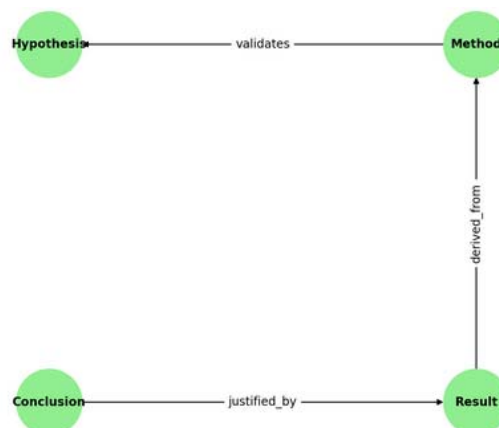


Figure 10 – Fully Coherent Ontological Graph for a High-Quality Article (Subset C)

Fig. 10 presents a sample graph for a benchmark publication from subset C. In this case, all major nodes and the logical links between them are present: Method validates Hypothesis, Results are derived from the method, and Conclusions provide a justified summary of the results. This structure is logically complete and corresponds to the expected model of scientific exposition.

Although the graphs for subsets B and C may appear visually identical – with the same nodes (Hypothesis, Method, Result, Conclusion), the same number of edges, and identical geometric layout – their fundamental difference lies in the semantics of the edges. In graph C, which corresponds to a benchmark publication, all connections are normative: Method validates Hypothesis, Result is derived from Method, and Conclusion is logically justified by Result. In contrast, in graph B, these same connections are replaced by stylistically neutral or semantically weak relations (*mentions*, *leads_to*, *followed_by*) which deprive the graph of logical force and render it formally complete but conceptually flawed.

This difference in the types of relations allows the system to detect violations of normative structure, even when the external form of the graph remains unchanged.

The comparison of graphs across subsets makes it possible to visually identify key differences: from missing nodes (subset A), through inconsistent relations (subset B), to logical completeness (subset C). The graph-based representation not only formalizes the scientific text but also serves as a powerful tool for the automated diagnosis of its structural quality.

6 DISCUSSION

The results obtained confirm that the use of an ontological model combined with logic-discursive validation enables effective detection of structural and semantic defects in scientific articles prior to the peer review stage. High Consistency and Onto-score values in the control group demonstrate that the system correctly identifies complete scientific argument structures, and in cases with intentionally disrupted or implicitly flawed logic, it is capable of detecting missing or inconsistent key components. This highlights not only the technical feasibility of the proposed approach but also its conceptual productivity.

Comparison with related research, particularly with approaches that rely on rule-based or formally structural models without deep semantic interpretation, indicates the advantage of a hybrid methodology that combines ontological reasoning, logical verification, and the generative capabilities of LLMs. Unlike traditional systems based on template-driven segment detection, our model operates at the level of semantic relations and context-sensitive logic of exposition. On the other hand, the presence of false positives in articles from the benchmark subset reveals the sensitivity limits of the model, consistent with findings that emphasize the constraints of LLMs when dealing with highly variable textual forms.

The similarity of results in subsets A and B, despite the different nature of the defects (artificially introduced vs. naturally occurring), indicates that the model is equally capable of responding to both explicit and latent deviations from normative structure. At the same time, in subset B, the system encounters more complex error configurations, explaining the wider spread of metrics and the higher proportion of logically ambiguous cases.

Among the limitations, it is worth noting the model's sensitivity to stylistic variations and non-formalized content presentation. In cases where the author deliberately avoids strict academic structure (e.g., does not explicitly mark the hypothesis or merges conclusions with discussion), the system may lower the evaluation score despite the presence of valid scientific content. This highlights the need for further extension of the ontology to encompass discursive strategies.

The practical applications of the proposed approach include automated support for initial editorial screening, review systems with low expert capacity, and educational platforms for teaching scientific writing. Integration of the model into the reviewer interface enables quick identification of weak points, structured recommendations, and enhanced transparency of the evaluation process.

Promising directions for future research include improving the logical core of the system using formal logics (e.g., description logics), leveraging vector-based ontologies for assessing semantic proximity of nodes, and testing the system on materials from diverse scientific domains. A separate objective is the creation of an open-access corpus of articles with logic-ontological annotation, which would enable standardization and advancement of metareviewing methodologies.

CONCLUSIONS

In this work, the scientific task of formalized automated assessment of the structural quality of scientific publications was addressed by developing and applying an ontology-driven model of discursive analysis. The proposed approach integrates the ontological formalization of a scientific publication, logic-discursive consistency verification, and semantic labeling using a large language model.

The scientific novelty of the research lies in the following. For the first time, an integration of the ontology of a scientific publication and the ontology of peer review is proposed to construct a formalized knowledge graph $G = (V, E, L)$ that captures both the structure and logic of an article. The logic-discursive verification approach has been further developed by defining a set of rules Φ that formalize the requirements of scientific exposition and are algorithmically verifiable. The methodology for computing the structural metrics Consistency and Onto-score has been improved, allowing not only the detection of node presence but also the evaluation of their coherence. Unlike previous approaches focused on template-based detection, our model performs a logical reconstruction of scientific discourse.

The practical significance of the results lies in the creation of a system capable of automatically analyzing scientific texts, detecting structural deficiencies, providing improvement recommendations, and generating preliminary reviews. The system can be integrated into editorial interfaces of scientific journals for preliminary screening of submitted articles or used as an educational tool in academic writing courses.

Among the achieved quantitative results: the precision of structural defect detection in the controlled corpus reached 82%, the Consistency score for benchmark articles was 0.92, and the Onto-score was 0.94. For articles from discontinued journals, these scores were 0.71 and 0.76, respectively, confirming the model's effectiveness in detecting both explicit and latent structural issues.

It is recommended that the obtained results be used to further automate peer review processes, structure scientific knowledge bases, and serve as an expert support tool in digital humanities and interdisciplinary research. In the future, the ontological model will be extended to other genres of scientific texts (e.g., review articles, methodological papers, dissertations), and mechanisms for explainability of logic-ontological decisions will be implemented.

REFERENCES

1. Arp R., Smith B., Spear A. D. Building ontologies with basic formal ontology. Cambridge, MIT Press, 2015, 248 p. DOI: 10.7551/mitpress/9780262527811.001.0001
2. Guizzardi G., Wagner G., Almeida J. P. A., Guizzardi R. S. S. Towards Ontological Foundations for Conceptual Modeling: The Unified Foundational Ontology (UFO) Story, *Applied Ontology*, 2015, Vol. 10, No. 3–4, pp. 259–271. DOI: 10.3233/AO-150157

3. Moltmann F. Natural Language Ontology. Oxford Research Encyclopedia of Linguistics, 2017. DOI: 10.1093/acrefore/9780199384655.013.330
4. Moltmann F. Truth predicates, truth bearers, and their variants, *Synthese*, 2021, Vol. 198, Suppl. 2, pp. 689–716. DOI: 10.1007/s11229-018-1814-8
5. Moltmann F. Relative truth and the first person, *Philosophical Studies*, 2010, Vol. 150, pp. 187–220. DOI: 10.1007/s11098-009-9383-9
6. Guizzardi G., Zamborlini V. Using a trope-based foundational ontology for bridging different areas of concern in ontology-driven conceptual modeling, *Science of Computer Programming*, 2014, Vol. 96, pp. 417–443. DOI: 10.1016/j.scico.2014.02.022
7. Cimiano P., Völker J., Studer R. Ontologies on demand? A description of the state-of-the-art, applications, challenges and trends for ontology learning from text, *Information Wissenschaft und Praxis*, 2006, Vol. 57, No. 6–7, pp. 315–320.
8. Wu S., Ma X., Luo D., Li L., Shi X., Chang X., Gong J. Automated review generation method based on large language models, *arXiv preprint*, 2024, arXiv:2407.20906. DOI: 10.48550/arXiv.2407.20906
9. Liang W., Zhang Y., Cao H., Wang B., Ding D. Y., Yang X., Zou J. Can large language models provide useful feedback on research papers? A large-scale empirical analysis, *NEJM AI*, 2024, Vol. 1, No. 8, Article A10a2400196. DOI: 10.1056/A10a2400196
10. Wadden D., Lin S., Lo K., Wang L. L., Zuylen M. van, Cohen A., Hajishirzi H. Fact or fiction: Verifying scientific claims, *arXiv preprint*, 2020, arXiv:2004.14974. DOI: 10.48550/arXiv.2004.14974
11. Tam D., Mascarenhas A., Zhang S., Kwan S., Bansal M., Raffel C. Evaluating the factual consistency of large language models through news summarization, *arXiv preprint*, 2022, arXiv:2211.08412. DOI: 10.48550/arXiv.2211.08412
12. Rahaman N., Weiss M., Wüthrich M., Bengio Y., Li L. E., Pal C., Schölkopf B. Language Models Can Reduce Asymmetry in Information Markets, *arXiv preprint*, 2024, arXiv:2403.14443. DOI: 10.48550/arXiv.2403.14443
13. Sinha A., Shen Z., Song Y., Ma H., Eide D., Hsu B. J., Wang K. An overview of Microsoft Academic Service (MAS) and applications, *Proceedings of the 24th International Conference on World Wide Web (WWW)*, 18–22 May 2015. Florence, Italy, pp. 243–246. DOI: 10.1145/2740908.2742839
14. Elsevier. List of Scopus discontinued journals [Electronic resource]. Access mode: <https://journalsearches.com/blog/scopus-discontinued-journals-list.php>
15. Committee on Publication Ethics (COPE). Core practices and guidelines [Electronic resource]. Access mode: <https://publicationethics.org/guidance>, 2023.

Received 18.06.2025.
Accepted 22.09.2025.

УДК 004.9

ЛОГІКО-ОНТОЛОГІЧНА РЕКОНСТРУКЦІЯ ДИСКУРСИВНОЇ СТРУКТУРИ НАУКОВОГО ТЕКСТУ ТА ЇЇ РЕАЛІЗАЦІЯ В ШІ-СИСТЕМІ РЕЦЕНЗУВАННЯ

Бедратюк Л. П. – д-р фіз.-мат. наук, професор, завідувач кафедри інженерії програмного забезпечення Хмельницького національного університету, Хмельницький, Україна.

АНОТАЦІЯ

Актуальність. Зростання кількості наукових публікацій і поява інструментів на основі великих мовних моделей (LLM) актуалізують потребу в автоматизованій верифікації структурної якості наукового тексту. Більшість існуючих рішень зосереджені на поверхневому лінгвістичному аналізі та не враховують логіко-дискурсивну цілісність: зокрема, чи присутні в тексті гіпотеза, метод, результати, висновки та чи пов'язані вони між собою нормативними зв'язками.

Мета. Метою дослідження є розроблення онтологічно керованого підходу до формалізованої перевірки структури наукових текстів шляхом побудови онтологічного графа знань і оцінки його відповідності наперед визначеній нормативній моделі наукового дискурсу.

Метод. Запропоновано модель, що ґрунтується на двох взаємопов'язаних онтологіях – «Наукова публікація» (визначає типи вузлів і їх ролі) та «Рецензування» (визначає логіко-дискурсивні вимоги). Текст подається у вигляді графа у якому вузли формуються на основі семантичної розмітки за допомогою LLM, а зв'язки перевіряються відповідно до множини нормативних правил. Для реалізації використано спеціалізованого GPT-агента, здатного динамічно застосовувати онтологічні знання під час аналізу та генерації рецензій.

Результати. Модель дозволяє автоматично виявляти порушення дискурсивної структури: відсутність ключових елементів, логічну розірваність, підміну наукової новизни практичною значущістю, некоректну інтерпретацію результатів. Запропоновані метрики кількісно фіксують рівень структурної повноти та узгодженості. Наведені приклади графів і рецензій демонструють, що система здатна виявляти неочевидні, латентні порушення логіки викладу навіть у формально повних текстах.

Висновки. Наукова новизна дослідження полягає у впровадженні онтологічного графа як інтерпретованої моделі наукової аргументації, що використовується у тандемі з великою мовною моделлю. Практичне значення полягає у створенні основи для напівавтоматизованого рецензування, структурного аналізу публікацій і навчання академічному письму. Методологія є масштабованою на інші жанри наукового тексту та потенційно інтегрованою у редакційні платформи.

КЛЮЧОВІ СЛОВА: онтологія, онтологічний граф, GPT-агент, рецензування, семантичний аналіз, наукова публікація, логіко-дискурсивна структура, машинне навчання, алгоритми.

ЛІТЕРАТУРА

1. Arp R. Building ontologies with basic formal ontology / R. Arp, B. Smith, A. D. Spear. – Cambridge : MIT Press, 2015. – 248 p. DOI: 10.7551/mitpress/9780262527811.001.0001
2. Towards Ontological Foundations for Conceptual Modeling: The Unified Foundational Ontology (UFO) Story / [G. Guizzardi, G. Wagner, J. P. A. Almeida, R. S. S. Guizzardi] // Applied Ontology. – 2015. – Vol. 10, No. 3–4. – P. 259–271. DOI: 10.3233/AO-150157
3. Moltmann F. Natural Language Ontology / F. Moltmann // Oxford Research Encyclopedia of Linguistics. – 2017. DOI: 10.1093/acrefore/9780199384655.013.330
4. Moltmann F. Truth predicates, truth bearers, and their variants / F. Moltmann // Synthese. – 2021. – Vol. 198, Suppl. 2. – P. 689–716. DOI: 10.1007/s11229-018-1814-8
5. Moltmann F. Relative truth and the first person / F. Moltmann // Philosophical Studies. – 2010. – Vol. 150. – P. 187–220. DOI: 10.1007/s11098-009-9383-9
6. Guizzardi G. Using a trope-based foundational ontology for bridging different areas of concern in ontology-driven conceptual modeling / G. Guizzardi, V. Zamborlini // Science of Computer Programming. – 2014. – Vol. 96. – P. 417–443. DOI: 10.1016/j.scico.2014.02.022
7. Cimiano P. Ontologies on demand? A description of the state-of-the-art, applications, challenges and trends for ontology learning from text / P. Cimiano, J. Völker, R. Studer // Information Wissenschaft und Praxis. – 2006. – Vol. 57, No. 6–7. – P. 315–320.
8. Automated review generation method based on large language models / [S. Wu, X. Ma, D. Luo et al.] // arXiv preprint. – 2024. – arXiv:2407.20906. DOI: 10.48550/arXiv.2407.20906
9. Can large language models provide useful feedback on research papers? A large-scale empirical analysis / [W. Liang, Y. Zhang, H. Cao et al.] // NEJM AI. – 2024. – Vol. 1, No. 8. – Article AIoa2400196. DOI: 10.1056/AIoa2400196
10. Fact or fiction: Verifying scientific claims / [D. Wadden, S. Lin, K. Lo et al.] // arXiv preprint. – 2020. – arXiv:2004.14974. DOI: 10.48550/arXiv.2004.14974
11. Evaluating the factual consistency of large language models through news summarization / [D. Tam, A. Mascarenhas, S. Zhang et al.] // arXiv preprint. – 2022. – arXiv:2211.08412. DOI: 10.48550/arXiv.2211.08412
12. Language Models Can Reduce Asymmetry in Information Markets / [N. Rahaman, M. Weiss, M. Wüthrich et al.] // arXiv preprint. – 2024. – arXiv:2403.14443. DOI: 10.48550/arXiv.2403.14443
13. An overview of Microsoft Academic Service (MAS) and applications / [A. Sinha, Z. Shen, Y. Song et al.] // Proceedings of the 24th International Conference on World Wide Web (WWW), 18–22 May 2015. – Florence, Italy. – P. 243–246. DOI: 10.1145/2740908.2742839
14. Elsevier. List of Scopus discontinued journals [Electronic resource]. – Access mode: <https://journalsearches.com/blog/scopus-discontinued-journals-list.php>
15. Committee on Publication Ethics (COPE). Core practices and guidelines [Electronic resource]. – Access mode: <https://publicationethics.org/guidance>, 2023.

METAHEURISTIC FRAMEWORKS FOR PARAMETER ESTIMATION IN APPROXIMATION MODELS

Grygor O. O. – Dr. Sc., Professor, Rector of the Cherkasy State Technological University, Cherkasy, Ukraine.

Fedorov E. E. – Dr. Sc., Professor, Professor of Department of Statistics and Applied Mathematics, Cherkasy State Technological University, Cherkasy, Ukraine.

Leshchenko M. M. – PhD, Associate Professor, Associate Professor of Department of International Economics and Business, Cherkasy State Technological University, Cherkasy, Ukraine.

Rudakov K. S. – PhD, Associate professor, Associate Professor of Department of Robotics and Specialized Computer Systems, Cherkasy State Technological University, Cherkasy, Ukraine.

Sakhno T. A. – PhD, Associate Professor of Department of International Economics and Business, Cherkasy State Technological University, Cherkasy, Ukraine.

ABSTRACT

Context. To enhance the performance of numerical optimization techniques, hybrid approaches integrating probabilistic modeling algorithms with annealing simulation have been introduced. These include Bayesian optimization, Markov-based strategies, and extended compact genetic algorithms, each augmented by annealing mechanisms. Such methods enable more precise search trajectories without requiring fitness function transformation, owing to their ability to explore the global search space in early iterations and refine the directionality of search in later stages.

Objective. The research aims to improve the effectiveness of parameter identification within approximation models of financial indicators by applying metaheuristic algorithms that incorporate probabilistic modeling and annealing-based simulation in intelligent computing systems.

Method. This study employs metaheuristic techniques grounded in probabilistic modeling and annealing-based simulation to enhance the accuracy and efficiency of parameter estimation within economic indicator approximation frameworks. Specifically, it introduces three hybrid strategies: Bayesian-based optimization integrated with annealing simulation, Markov-driven optimization enhanced by annealing, and an extended compact genetic algorithm coupled with annealing mechanisms. These methods enhance the accuracy of the search process by exploring the entire search space in initial iterations and refining the search direction in final iterations. The Bayesian optimization method employs a Bayesian network for structured search and solution refinement. The Markov optimization method integrates Gibbs quantization within a Markov network to improve search precision. The extended compact genetic algorithm utilizes limit distribution models to generate optimal solutions. These methods eliminate the need for fitness function transformation, optimizing computational efficiency. The proposed techniques expand the application of metaheuristics in intelligent economic computer systems.

Results. The implemented optimization strategies significantly enhanced the precision of parameter estimation within intelligent financial computing frameworks. The combination of probabilistic models and annealing simulation enhanced search efficiency without requiring fitness function transformation.

Conclusions. The proposed method expands the application of metaheuristics in economic modeling, increasing computational effectiveness. Further research should explore their implementation across diverse artificial intelligence problems.

KEYWORDS: probabilistic optimization frameworks, hybrid metaheuristic techniques, adaptive search algorithms, annealing-based simulation, computational parameter estimation, parametric identification, economics approximation model.

ABBREVIATIONS

GDP is a gross domestic product;
ICT is an information and communications technology;
BOA is a Bayesian optimization algorithm;
MOA is a Markov optimization algorithm;
BIC is a Bayes information criterion;
BD is a Bayes-Dirichlet;
ECGA is an Extended Compact Genetic Algorithm;
MDL is a Minimum Descriptive Length.

NOMENCLATURE

A^{rp} is a reproduction operator;
 A^{cr} is a probabilistic model creation operator;
 A^{gn} is a descendant generation operator;
 A^{rd} is a reduction operator;

N is a maximum number of iterations;
 K is a population size;
 $\tilde{K}$ is a middle population size;
 $\hat{K}$ is a descendants' number;
 $A^{\max}$ is an edges maximum count;
 M is a length of chromosome;
 E is bits count for gene of each chromosome;
 $x_j^{\min}, x_j^{\max}$ are the minimum and maximum values for a chromosome;
 x is a chromosome real vector;
 b is a chromosome binary vector;
 $U(0,1)$ is a function returning a standard uniformly distributed random value;
 n is an iteration number;
 G is a Bayesian network;

$\tilde{G}$ is an ordered Bayesian network;
 G_j^{in} is the set of vertices that preceding the vertex
 G_j^{num} ;
 G_j^{out} is the set of vertices that following the vertex
 G_j^{num} ;
 α is a significance parameter;
 α_s is a natural parameter;
 $equal()$ is a function returning the matched element
count;
 $round()$ is a function returning rounded number;
 β is a specifying the cooling coefficient;
 T_0 is an initial temperature;
 T^{max} is a maximum temperature;
 N_1 is an iterations maximum count;
 N_2 is an iterations maximum count for Gibbs quanti-
zation;
 P is a population;
 $\tilde{P}$ is a middle population;
 $\hat{P}$ is a descendant's population;
 $\mathbf{w}$ is a vector of weights;
 p_{ij} , $p1_{ij}$, $p2_{ij}$ are a probability;
 a^{avg} is a joint information matrix average value;
 $S1$ is a disjoint set (set of clusters);
 $S2$ is two different clusters set of unions;
 $\mathbf{B}$ is a joint information symmetric matrix;
 $C1_i^{pop}$, $C2_i^{pop}$ are a compressed population complex-
ity;
 $C1_i^{model}$, $C2_i^{model}$ are a model complexity;
 $C1_i$, $C2_i$ are a combined complexity.

INTRODUCTION

The post-industrial era is marked by the expansion of electronic environments and the shift toward an information-driven society. In recent years, digitalization has emerged as a central challenge within the broader context of economic development, particularly in the information and communication technology sector. Digital transformation influences both economic progress and societal acceleration through complex, interconnected mechanisms, highlighting the need for standardized frameworks and reliable metrics to assess the digital economy and facilitate data exchange across industries. In this context, the implementation of parametric identification techniques for economic indicator approximation models – utilized in intelligent financial computing systems informed by digitalization metrics – has become increasingly significant. A key objective is to advance such identification methods to enhance model accuracy and system performance. Traditional random search algorithms lack convergence guarantees, while exact optimization tech-

niques are computationally intensive. Local search-based methods, although more efficient, are prone to entrapment in local optima, underscoring the need for more effective optimization strategies.

Advanced heuristic approaches – commonly known as metaheuristics [1–3] – are utilized to expedite the search process for near-optimal solutions in numerical optimization tasks while reducing the likelihood of convergence to local extrema. These techniques extend classical heuristics by incorporating multiple strategies into a higher-order framework [4–6]. Population-based variants, in particular, are employed to enhance search speed and solution accuracy through collective and adaptive mechanisms [7–9]. The focus of this study is the parametric identification process applied to economic indicator approximation models. The research specifically investigates metaheuristic-based techniques designed to enhance the accuracy and efficiency of such identification methods.

The present research aims to enhance the efficiency of parametric identification applied to approximation models of economic indicators. To achieve this objective, a metaheuristic optimization method is proposed, integrating probabilistic modeling techniques with simulated annealing.

To fulfill the research goal, the following tasks were undertaken:

- 1) the development of a hybrid optimization method combining Bayesian optimization algorithms with annealing simulation;
- 2) the creation of a Markov-based optimization strategy supported by simulated annealing principles;
- 3) the formulation of an extended compact genetic algorithm integrated with annealing mechanisms; and
- 4) the execution of a numerical study to evaluate the proposed optimization techniques.

In parallel with these developments, contemporary trends reveal a growing prevalence of automated image analysis systems powered by machine learning algorithms. These technologies, widespread across domains such as diagnostics, remote sensing, and economic modeling, offer enhanced accuracy and adaptability, reinforcing the relevance and applicability of intelligent computational methods within financial systems.

1 PROBLEM STATEMENT

The core challenge addressed in this research involves enhancing the efficiency of solving numerical optimization tasks by employing probabilistic, multidimensional, and discrete evolutionary metaheuristic frameworks. Formally, the problem is defined as constructing a structured set of operators $\{A^{rp}, A^{cr}, A^{gn}, A^{rd}\}$, whose repeated application leads to a solution x^* satisfying the conditions $F(x^*) \rightarrow \min$ and $T \rightarrow \min$.

2 REVIEW OF THE LITERATURE

Existing metaheuristic frameworks have been observed to exhibit several limitations:

- many approaches provide descriptions tailored to specific problems or rely on abstract generalizations without broader applicability [10–12];
- they often lack support for non-binary candidate solution representations [13–15];
- the influence of iteration count on the search process is frequently overlooked [16–18];
- convergence to an optimal solution is not consistently guaranteed [19–21];
- there is limited capability to address constrained optimization problems [22–24];
- accuracy levels may be insufficient for complex or sensitive optimization tasks [25–27];
- parameter value determination procedures are typically non-automated and require manual tuning [28].

This highlights the ongoing challenge of developing robust and efficient metaheuristic optimization techniques. Among the widely adopted approaches are the Bayesian optimization algorithm [29], the Markov-based optimization framework [30], and the extended compact genetic algorithm [31], each offering distinct advantages in solving complex numerical problems.

3 MATERIALS AND METHODS

The extent of digital technology advancement across countries is primarily shaped by a set of indicators and influencing factors that collectively provide a multidimensional profile of each nation's socioeconomic landscape. These metrics serve to reflect both the degree of digital integration and the overall performance of digital systems within diverse national contexts.

A comprehensive dataset was compiled for the purpose of this study, encompassing macroeconomic indicators that reflect national economic performance – specifically, GDP per capita (final annual values in US dollars, denoted as Y). The dataset also integrates metrics related to digital technology development, which serve as explanatory variables. These include: X_1 – Frontier Technology Readiness Index (ICT component), annual; X_2 – Frontier Technology Readiness Index (Research and Development component), annual; X_3 – Share of ICT goods in total trade, annual percentage; X_4 – International trade in digitally deliverable services, annual value in millions of US dollars. All indicators were collected for the period spanning 2012 to 2022, covering 159 countries within the global economy.

The Bayesian optimization method incorporating annealing simulation was developed to enhance search efficiency and solution quality. The Bayesian Optimization Algorithm (BOA) [29], originally introduced by Pelikan, Goldberg, and Cantú-Paz [29], operates through a multi-phase process: reproduction, probabilistic model construction via Bayesian networks, offspring generation based on the learned network, and a subsequent reduction phase. The underlying probabilistic graphical model employed in this approach is illustrated in Figure 1.

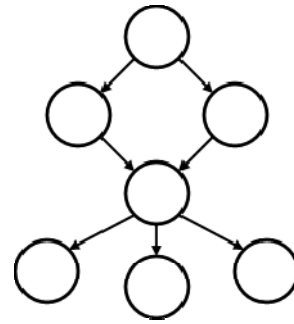


Figure 1 – Probabilistic BOA Graphical Model [29]

The methodology is structured into the following sequential phases:

1. Initial setup.

1.1. Defining the N , K , $\tilde{K}$, $\hat{K}$, $A^{\max}$ integrated into a specific vertex of the Bayesian structure: M , E , $x_j^{\min}$, $x_j^{\max}$, $j \in \overline{1, M}$.

1.2. Defining a fitness function $F(x) \rightarrow \min_x$

1.3. Generation each source population chromosome k , $k \in \overline{1, K}$

a) generation each binary vector component j , $j \in \overline{1, M \cdot E}$

$$\lambda = U(0,1).$$

$$b_{kj} = \begin{cases} 1, & \lambda < 0.5 \\ 0, & \lambda \geq 0.5 \end{cases}$$

b) creation of a real vector

$$x_{kj} = x_j^{\min} + (x_j^{\max} - x_j^{\min}) \frac{\sum_{e=1}^E (2^{E-e} \cdot b_{k,(j-1)E+e+1-e})}{2^E - 1},$$

$j \in \overline{1, M}$.

c) when $(b_k, x_k) \notin P$, $P = P \cup \{(b_k, x_k)\}$.

1.4. Finding the best chromosome

$$k^* = \arg \min_k F(x_k), \quad x^* = x_{k^*}$$

2. The count of iteration $n = 1$.

3. Replication and generation middle population $\tilde{P}$.

3.1. To organize P by fitness function, $F(x_k) < F(x_{k+1})$.

3.2. A new population $\tilde{P} = \{(\tilde{b}_k, \tilde{x}_k)\}$ of defined size $\tilde{K}$ is generated through a hybrid approach that combines simulated annealing with linear order-based selection.

Calculation the choosing each i -th chromosome probability:

$$p(x_i) = \frac{1}{K} \exp(-1/T(n)) + \frac{1}{K} \left(a - (2a - 2) \frac{i-1}{K-1} \right) (1 - \exp(-1/T(n))),$$

$$T(n) = \beta T(n-1), \quad 0 < \beta < 1, \quad T(0) = T_0, \quad T_0 > 0.$$

At both the initial and final phases of the genetic algorithm, distinct selection strategies are employed to balance exploration and exploitation. During the early iterations, equiprobable selection is applied, enabling broad coverage of the search space through random chromosome sampling. In contrast, the later stages utilize a linearly ranked selection mechanism that prioritizes high-performing chromosomes, thereby guiding the search toward promising regions. This dual-phase approach eliminates the need for fitness function transformation while maintaining search efficiency.

4. To develop a Bayesian network G .

4.1. Creating elements of Bayesian Network.

$$G_j = (G_j^{in}, G_j^{out}, G_j^{num}),$$

$$G_j^{in} = \emptyset, G_j^{out} = \emptyset, G_j^{num} = j, j \in \overline{1, M \cdot E}.$$

4.2. Directed connection in the Bayesian structure count $m=1$.

4.3. $w^{\max} = -1, v^{from} = 0, v^{to} = 0$

4.4. The size of the vertex set defining the Bayesian network count $i=1$.

4.5. Creation each viable descendant nodes j , $j \in \overline{1, M \cdot E}$, for the i -th parent node.

a) initialization of the visited vertices empty set $V^{viable} = \emptyset$, set a vertex j in the empty vertex stack S^V .

b) when the vertex i is on the S^V , $f^{pe} = 1$, move to stage j).

c) to pop a vertex from S^V and set it v^{cur} .

d) when $v^{cur} \in V^{visited}$, move to stage b)

e) $V^{visited} = V^{visited} \cup \{v^{cur}\}$.

f) when $|G_{v^{cur}}^{out}| = 0$, move to stage h)

g) for each $l, l \in \overline{1, |G_{v^{cur}}^{out}|}$, when $g_{v^{cur}l}^{out} \notin V^{visited}$,

move S^V at the top $g_{v^{cur}l}^{out}$.

h) when the S^V is not empty, move to stage b)

i) $f^{pe} = 0$.

j) when $G_i^{num} \neq j \wedge j \notin G_i^{out} \wedge f^{pe} = 0$, then $V^{viable} = V^{viable} \cup \{j\}$.

4.6. To calculate each weights vector component j , $j \in \overline{1, M \cdot E}$, based on $K2$ or BIC or BD .

When $|G_j^{in}| = A^{\max}$ or $j \notin V^{viable}$, $w_j = -1$ then

a) $C = \{G_i^{num}\} \cup G_j^{in} \cup \{j\}$.

b) element sets $Q_1, \dots, Q_{2^{|C|}}$ are derived, consisting of $\tilde{b}_k$ chromosomes, where the gene values are defined with values derived from the specified set C (that is, the val-

ues assigned to individual genes $\tilde{b}_{k,c_1}, \dots, \tilde{b}_{k,c_{|C|}}$) match set

Q_s with value s for $\tilde{b}_k$ chromosome was defined as

$$s = \sum_{e=1}^{|C|} 2^{|C|-e} \tilde{b}_{k, c_{E+1-e}}, \text{ for chromosomes in } Q_{2s-1}, Q_{2s}$$

sets, the values assigned to individual genes from C set (genes values is $\tilde{b}_{k,c_1}, \dots, \tilde{b}_{k,c_{|C|}}$) different corresponding to a single gene.

c) to assign a weight w_j to an edge (i, j) , the following basis is applied:

– BD metric

$$w_j = \prod_{s=1}^{2^{|C|}/2} \frac{(\alpha_{2s-1} + \alpha_{2s} - 1)!}{(\alpha_{2s-1} - 1)! (\alpha_{2s} - 1)!} \cdot \frac{|Q_{2s-1} + \alpha_{2s-1} - 1|! |Q_{2s} + \alpha_{2s} - 1|!}{(|Q_{2s-1}| + |Q_{2s}| + \alpha_{2s-1} + \alpha_{2s} - 1)!},$$

when $\alpha_s = 1$, BD matches with $K2$.

– $K2$ metric

$$w_j = \prod_{s=1}^{2^{|C|}/2} \frac{|Q_{2s-1}|! |Q_{2s}|!}{(|Q_{2s-1}| + |Q_{2s}| + 1)!};$$

– BIC metric

$$w_j = \sum_{s=1}^{2^{|C|}/2} \left(|Q_{2s-1}| \ln \frac{|Q_{2s-1}|}{|Q_{2s-1}| + |Q_{2s}|} + |Q_{2s}| \ln \frac{|Q_{2s}|}{|Q_{2s-1}| + |Q_{2s}|} \right) - \left(\frac{1}{2} \ln \tilde{K} \right) 2^{|C|}.$$

4.7. Determination a Best Edge (v^{from}, v^{to})

For each Bayesian network vertex count j , $j \in \overline{1, M \cdot E}$, when $w^{\max} < w_j$, then $w^{\max} = w_j$, $v^{from} = i$, $v^{to} = j$.

4.8. When $i < M \cdot E$, perform $i = i + 1$, move to stage 4.5.

4.9. When $w^{\max} > 0$, move to the most informative edge (v^{from}, v^{to}) for enhancing the Bayesian model:

$$G_{v^{to}}^{in} = v^{from}, G_{v^{from}}^{out} = v^{to}.$$

4.10. When $m < A^{\max} \tilde{K}$, perform $m = m + 1$, move to stage 4.8.

5. Creation of a descendant $\tilde{P}$ population $\tilde{P}$ guided by the Bayesian network structure.

5.1. Generation Bayesian network $\tilde{G}$ with a defined topological sequence of nodes.

Integration of an additional element G_j into every

Bayesian network structure $G_j^{count} = |G_j^{in}|$, $j \in \overline{1, M \cdot E}$

element. $\tilde{G} = \emptyset$, set a $\{G_j \mid |G_j^{in}| = 0, j \in \overline{1, M \cdot E}\}$ subset

of elements in the empty S^G . Create each element q , $q \in \overline{1, M \cdot E}$, for G_j component:

- detaching S^G element from the $\tilde{G}_q$ and set it.
- when $|\tilde{G}_q^{out}| = 0$, move to stage h).
- $l = 1$.
- $j^* = \underset{j}{\text{equal}}(G_j^{num}, \tilde{g}_l^{out})$.
- $G_{j^*}^{count} = G_{j^*}^{count} - 1$.
- when $G_{j^*}^{count} \leq 0$, to detach element S^G from the G_{j^*} .
- when $l < |\tilde{G}_q^{out}|$, perform $l = l + 1$, move to stage d).
- $\tilde{G} = \tilde{G} \cup \{\tilde{G}_q\}$.

5.2. Creation each descendant population chromosome i , $i \in \overline{1, K}$, using PLS method.

- Generation each binary vector component j , $j \in \overline{1, M \cdot E}$. Child vector is initialized as set like $\hat{b}_i = (0, \dots, 0)$.

When $|\tilde{G}_j^{in}| = 0$, to determinate the probability

$$p_j = \frac{1}{\tilde{K}} \sum_{k=1}^{\tilde{K}} \tilde{b}_{kj}, \text{ to generate a binary vector}$$

$$\hat{b}_{ij} = \begin{cases} 1, & p_j < 0.5 \\ 0, & p_j \geq 0.5 \end{cases} \text{ and move to stage b).}$$

$$C = \{G_j^{num}\} \cup G_j^{in}.$$

Sets $Q_1, \dots, Q_{2^{|C|}}$ are generated, consisting of chromosomes $\tilde{b}_k$, in which the genes values with values from the C (genes values are $\tilde{b}_{k, c_1}, \dots, \tilde{b}_{k, c_{|C|}}$) represent a valid match, and the set Q_s value s for the chromosome $\tilde{b}_k$ is specified like $s = \sum_{e=1}^{|C|} 2^{|C|-e} \tilde{b}_{k, c_{E+1-e}}$, for chromosomes in the $Q_{2^{|S|-1}}, Q_{2^{|S|}}$ sets, values assigned to genes with values from C set (genes values are $\tilde{b}_{k, c_1}, \dots, \tilde{b}_{k, c_{|C|}}$) different exclusively for individual gene.

$$C1 = G_j^{in}.$$

$$\text{To compute } p_j = \frac{|Q_a|}{|Q_a| + |Q_b|},$$

$$a = \sum_{e=1}^{|C1|} 2^{|C1|-e} \hat{b}_{i, c1_{E+1-e}} + 2^{|C1|}, b = \sum_{e=1}^{|C1|} 2^{|C1|-e} \hat{b}_{i, c1_{E+1-e}},$$

for chromosomes in Q_a, Q_b sets, value of the genes with numbers in the C set (i.e., the genes values is

$\tilde{b}_{k, c_1}, \dots, \tilde{b}_{k, c_{|C|}}$) different exclusively for $\tilde{b}_{k, c_1}$ gene associated with the given vertex G_j^{num} .

To generate a binary vector

$$\hat{b}_{ij} = \begin{cases} 1, & p_j < 0.5 \\ 0, & p_j \geq 0.5 \end{cases}$$

- generating a real vector

$$\hat{x}_{ij} = x_j^{\min} + (x_j^{\max} - x_j^{\min}) \frac{\sum_{e=1}^E (2^{E-e} \cdot \hat{b}_{i, (j-1)E + E + 1 - e})}{2^E - 1},$$

$$j \in \overline{1, M}.$$

$$c) \hat{P} = \hat{P} \cup \{(\hat{b}_i, \hat{x}_i)\}.$$

6. Reduction.

6.1. Merging descendants $\hat{P}$ and population P

6.2. Ordering union by fitness function, $F(x_k) < F(x_{k+1})$.

6.3. Constructing a refined population P using the most optimal chromosomes K identified in the aggregated set.

7. To calculate the best chromosome

$$k^* = \arg \min_k F(x_k)$$

8. When $F(x_{k^*}) < F(x^*)$, perform $x^* = x_{k^*}$.

9. When $n < N$, then $n = n + 1$, move to stage 3.

The result is x^* .

The Markov Optimization Algorithm (MOA), incorporating simulated annealing, was developed to enhance search efficiency and solution quality. Originally introduced by Shakya and Santana [30], this method operates through a structured sequence of phases: reproduction, construction of a probabilistic Markov network, generation of offspring guided by the network and Gibbs sampling principles, and a final reduction phase. The underlying probabilistic graphical model utilized in this approach is illustrated in Figure 2.

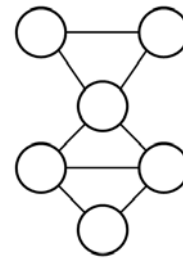


Figure 2 – Probabilistic structure of the Markov Optimization Algorithm [30]

The algorithmic workflow is delineated through the following sequence:

1. Initialization.

1.1. Determining the β , $T^{\max}$, α , and $0 < \beta < 1$, $T^{\max} > 0$, $\alpha > 1$.

1.2. Setting the $N_1, N_2, K, \tilde{K}, \hat{K}, A^{\max}$ connecting a Markov network subset of vertices with one vertex; $M, E, x_j^{\min}, x_j^{\max}, j \in \overline{1, M}$.

1.3. Determining the fitness function $F(x) \rightarrow \min_x$.

1.4. Generation each source population chromosome $k, k \in \overline{1, K}$,

a) generation each binary vector component $j, j \in \overline{1, M \cdot E}$

$$\lambda = U(0,1).$$

$$b_{kj} = \begin{cases} 1, & \lambda < 0.5 \\ 0, & \lambda \geq 0.5 \end{cases}$$

b) generation a real vector.

$$x_{kj} = x_j^{\min} + (x_j^{\max} - x_j^{\min}) \frac{\sum_{e=1}^E (2^{E-e} \cdot b_{k,(j-1)E+1-e})}{2^E - 1},$$

$j \in \overline{1, M}$.

c) when $(b_k, x_k) \notin \tilde{P}$, perform $P = P \cup \{(b_k, x_k)\}$

1.5. Finding the best chromosome $k^* = \arg \min_k F(x_k), x^* = x_{k^*}$.

2. Iteration count $n = 1$.

3. Replication and generation middle population $\tilde{P}$.

3.1. To sort P by fitness function, i.e. $F(x_k) < F(x_{k+1})$.

3.2. To generate $\tilde{P} = \{(\tilde{b}_k, \tilde{x}_k)\}$ population $\tilde{K}$ size from a P population based on a simulated annealing with combination of linearly ordered selection and random selection.

Calculation the choosing each i -th chromosome probability:

$$p(x_i) = \frac{1}{K} \exp(-1/T(n)) + \frac{1}{K} \left(a - (2a-2) \frac{i-1}{K-1} \right) (1 - \exp(-1/T(n)))$$

$$T(n) = \beta T(n-1), 0 < \beta < 1, T(0) = T_0, T_0 > 0.$$

This mechanism is applied at both the initial and terminal phases of the genetic algorithm's operation. In its early phase, the algorithm employs uniform random selection to enable broad exploration of the search space through stochastic sampling of chromosomes. In the final phase, a linearly ranked selection strategy is adopted, shifting the process toward targeted exploitation by retaining the currently best-performing individuals. This integrated approach operates without requiring modification of the fitness function.

4. To generate a Markov network G .

4.1. Adjustment of a component j , within the unconditional probability vector $j \in \overline{1, M \cdot E}$

a) a pair of chromosomes $\tilde{b}_k$ sets $Q0, Q1$ is constructed, where gene values g_j are consistent across members, and for these sets $Q0 \tilde{b}_{kj} = 0, Q1 \tilde{b}_{kj} = 1$.

$$b) p(g_j = 0) = \frac{|Q0|}{\tilde{K}}, p(g_j = 1) = \frac{|Q1|}{\tilde{K}}.$$

4.2. Modification of each element (i, j) , within the joint probability vector: $i \in \overline{1, M \cdot E - 2}, j \in \overline{2, M \cdot E - 1}$.

a) the sets $Q00, Q01, Q10, Q11$ are formed, containing the chromosomes $\tilde{b}_k$, where the j -th gene values and the i -th gene values coincide; for $Q00 \tilde{b}_{kj} = 0, \tilde{b}_{ki} = 0$, for $Q01 \tilde{b}_{kj} = 0, \tilde{b}_{ki} = 1$, for $Q10 \tilde{b}_{kj} = 1, \tilde{b}_{ki} = 0$, for $Q11 \tilde{b}_{kj} = 1, \tilde{b}_{ki} = 1$

$$b) p(g_i = 0, g_j = 0) = \frac{|Q00|}{\tilde{K}},$$

$$p(g_i = 1, g_j = 0) = \frac{|Q01|}{\tilde{K}},$$

$$p(g_i = 0, g_j = 1) = \frac{|Q10|}{\tilde{K}},$$

$$p(g_i = 1, g_j = 1) = \frac{|Q11|}{\tilde{K}}.$$

4.3. Generating of each joint information symmetric matrix element $(i, j), i \in \overline{1, M \cdot E - 2}, j \in \overline{2, M \cdot E - 1}$ when $i < j$, perform

$$b_{ij} = \sum_{g_i=0}^1 \sum_{g_j=0}^1 p(g_i, g_j) \ln \left(\frac{p(g_i, g_j)}{p(g_i)p(g_j)} \right), b_{ij} = b_{ji}.$$

4.4. Calculation of the joint information matrix average value:

$$a^{avg} = \frac{2}{(M \cdot E - 1)^2 - (M \cdot E - 1)} \sum_{i=1}^{M \cdot E - 2} \sum_{\substack{j=2 \\ j>i}}^{M \cdot E - 1} b_{ij}.$$

4.5. To expand the Markov network by adding new links (i, j) between nodes

$$i \in \overline{1, M \cdot E - 2}, j \in \overline{2, M \cdot E - 1},$$

when $i < j$ and $b_{ij} > a^{avg} \alpha$, to add an edge (i, j) to the Markov network.

4.6. Deleting worst edges for each vertex $i, i \in \overline{1, M \cdot E - 2}$

If a node i is connected to more than a specified number $A^{\max}$ of other nodes, retain only the top edges $A^{\max}$ with the highest values b_{ij} .

4.7. To determine adjacent elements for each vertex $j, j \in \overline{1, M \cdot E}$

To generate surroundings U_j from vertices linked by edges to a vertex j .

5. To generate each descendants' population chromosome k , $k \in \overline{1, K}$, from a population $\tilde{P}$ based on a Gibbs quantization and Markov network.

5.1. To initialize every entry j of the binary vector, $j \in \overline{1, M \cdot E}$

$$\lambda = U(0,1).$$

$$\hat{b}_{kj} = \begin{cases} 1, & \lambda < 0.5 \\ 0, & \lambda \geq 0.5. \end{cases}$$

5.2. During each iteration of Gibbs-based quantization l , $l \in \overline{1, N_2}$

a) random selection of the index corresponding to the modified gene, i.e., $j = \text{round}(1 + (M-1)U(0,1))$, $g_j = \hat{b}_{kj}$.

b) two groups of chromosomes $Q0, Q1$ are formed, where the gene values $\tilde{b}_m$ from the surrounding context U_j match those within the corresponding positions $\hat{b}_k$ of each chromosome, and for $Q0$ $\hat{b}_{kj} = 0$, for $Q1$ $\hat{b}_{kj} = 1$.

$$c) p(g_j = 0 | U_j) = \frac{|Q0|}{\tilde{K}}, p(g_j = 1 | U_j) = \frac{|Q1|}{\tilde{K}}.$$

$$d) \lambda = U(0,1).$$

e) changing of the j -th gene value of a binary vector:
 $b_{kj} =$

$$= \begin{cases} 0, & \lambda < \frac{\exp\left(\frac{p(g_j = 0)}{T(l)}\right)}{\exp\left(\frac{p(g_j = 0 | U_j)}{T(l)}\right) + \exp\left(\frac{p(g_j = 1 | U_j)}{T(l)}\right)} \\ 1, & \lambda \geq \frac{\exp\left(\frac{p(g_j = 0)}{T(l)}\right)}{\exp\left(\frac{p(g_j = 0 | U_j)}{T(l)}\right) + \exp\left(\frac{p(g_j = 1 | U_j)}{T(l)}\right)}. \end{cases}$$

$$f) T(l+1) = \beta T(l), \text{ where } T(1) = T^{\max}$$

5.3. To generate a real vector:

$$\hat{x}_{kj} = x_j^{\min} + (x_j^{\max} - x_j^{\min}) \frac{\sum_{e=1}^E (2^{E-e} \cdot \hat{b}_{k,(j-1)E+E+1-e})}{2^E - 1},$$

$$j \in \overline{1, M}.$$

$$5.4. \hat{P} = \hat{P} \cup \{(\hat{b}_k, \hat{x}_k)\}.$$

6. Reduction.

6.1. Integrate descendants $\hat{P}$ into the current population P .

6.2. To sort union by fitness function, i.e., $F(x_k) < F(x_{k+1})$.

6.3. Finding the K first (best) chromosomes in the union and generation a new population P .

$$7. \text{ Finding the best chromosome } k^* = \arg \min_k F(x_k).$$

$$8. \text{ When } F(x_{k^*}) < F(x^*), x^* = x_{k^*}.$$

$$9. \text{ When } n < N_1, \text{ perform } n = n + 1, \text{ move to stage 3.}$$

The result is x^* .

The extended compact genetic algorithm (ECGA) [31], enhanced with simulated annealing, was developed to improve optimization performance in complex search spaces. Originally introduced by Harik [31], ECGA operates through a structured sequence of stages: reproduction, construction of a probabilistic model based on the product of marginal limit distributions, offspring generation guided by this model, and a final reduction phase. The probabilistic graphical representation underlying this method is depicted in Figure 3.

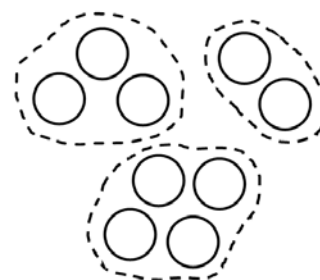


Figure 3 – The probabilistic model underlying the ECGA optimization approach [31]

The developed method unfolds through the following sequential steps:

1. Initialization.

1.1. To set N , K , $\tilde{K}$, $\hat{K}$, M , E , $x_j^{\min}, x_j^{\max}$, $j \in \overline{1, M}$.

1.2. Specification of the fitness evaluation function $F(x) \rightarrow \min_x$.

1.3. Generation each source population chromosome k , $k \in \overline{1, K}$

a) generation each binary vector j , $j \in \overline{1, M \cdot E}$

$$\lambda = U(0,1)$$

$$b_{kj} = \begin{cases} 1, & \lambda < 0.5 \\ 0, & \lambda \geq 0.5 \end{cases}$$

b) generation a real vector:

$$x_{kj} = x_j^{\min} + (x_j^{\max} - x_j^{\min}) \frac{\sum_{e=1}^E (2^{E-e} \cdot b_{k,(j-1)E+E+1-e})}{2^E - 1},$$

$$j \in \overline{1, M}.$$

c) when $(b_k, x_k) \notin \tilde{P}$, perform $P = P \cup \{(b_k, x_k)\}$.

1.4. Finding the best chromosome $k^* = \arg \min_k F(x_k)$, $x^* = x_{k^*}$.

2. Count of iteration $n = 1$.

3. Replication and generation middle population $\tilde{P}$.

3.1. Sort P according to goal function $F(x_k) < F(x_{k+1})$.

3.2. Population $\tilde{K}$ size is regenerated $\tilde{P} = \{(\tilde{b}_k, \tilde{x}_k)\}$ through a hybrid mechanism combining simulated annealing with both deterministic ranking and random selection techniques.

Calculation the choosing each i -th chromosome probability:

$$p(x_i) = \frac{1}{K} \exp(-1/T(n)) + \frac{1}{K} \left(a - (2a - 2) \frac{i-1}{K-1} \right) (1 - \exp(-1/T(n))),$$

$$T(n) = \beta T(n-1), \quad 0 < \beta < 1, \quad T(0) = T_0, \quad T_0 > 0.$$

4. Generation of a probabilistic model.

4.1. Generation of a disjoint sets (set of clusters) $S1$ containing one gene (vertex), i.e., $S1 = \{S1_j\}$, $S1_j = \{j\}$, $j \in \overline{1, M}$

4.2. Calculation combined complexity for each cluster i , $i \in \overline{1, |S1|}$ (used the MDL metric).

a) estimating computational intricacy $C1_i^{model} = \log_2 \tilde{K} (2^{|S1_i|} - 1)$.

b) chromosome subsets $Q_{i1}, \dots, Q_{i|S1|}$ are formed from the intermediate population, with gene values aggregated according to their cluster-specific frequencies $S1_i$ (i.e., the gene value $\tilde{b}_{k, S1_j}$) is equal to 1

c) probability computation $p1_{ij} = \frac{|Q_{ij}|}{\tilde{K}}$.

d) evaluation of population encoding efficiency:

$$C1_i^{pop} = \tilde{K} \sum_{j=1}^{2^{|S1_i|}} - p1_{ij} \log_2 p1_{ij}, \quad i \in \overline{1, |S1|}.$$

e) calculating the combined complexity (used the MDL metric) $C1_i = C1_i^{model} + C1_i^{pop}$.

4.3. Generating a set of two differ clusters unions $S2$.

a) $q = 1$, $S2 = \emptyset$

b) for each pair vertex (i, j) , $i \in \overline{1, |S1| - 1}$, $j \in \overline{1, |S1|}$, if $i < j$, then $S2_q = S1_i \cup S1_j$, $i_q = i$, $j_q = j$, $q = q + 1$.

4.4. calculation combined complexity for each combined cluster i , $i \in \overline{1, |S2|}$ (used the MDL metric).

a) computing model complexity $C2_i^{model} = \log_2 \tilde{K} (2^{|S2_i|} - 1)$.

b) population-derived chromosome sets $Q_{i1}, \dots, Q_{i|S2|}$ are formed, embedding gene values alongside their set-specific frequencies $S2_i$ (the gene value $b_{k, S2_{ij}}$) is equal to 1

c) probability assessment $p2_{ij} = \frac{|Q_{ij}|}{\tilde{K}}$.

d) evaluation of execution complexity

$$C2_i^{pop} = \tilde{K} \sum_{j=1}^{2^{|S2_i|}} - p2_{ij} \log_2 p2_{ij}.$$

e) estimation of total model complexity using the MDL principle $C2_i = C2_i^{model} + C2_i^{pop}$.

4.5. Finding the best cluster union

$$q^* = \arg \max_q (C1_{i_q} + C1_{j_q} - C2_q),$$

$$\Delta C^{best} = C1_{i_{q^*}} + C1_{j_{q^*}} - C2_{q^*}.$$

4.6. If $\Delta C^{best} > -1$, then $S1 = (S1 \cup \{S2_{q^*}\}) \setminus (\{S1_{i_{q^*}}\} \cup \{S1_{j_{q^*}}\})$, move to stage

4.2.

5. To generate each descendants' population chromosome k , $k \in \overline{1, K}$, from a population $\tilde{P}$ based on the generate marginal distributions model

a) generation of binary vectors for each cluster i , $i \in \overline{1, |S1|}$

Random initialization of a vector d populated with chromosome frequency values $\{1, \dots, \tilde{K}\}$

$$\tilde{b}_{ks1_{ij}} = \tilde{b}_{d_k s1_{ij}}, \quad k \in \overline{1, \tilde{K}}, \quad j \in \overline{1, S1_i}.$$

b) initialization of real-valued feature vectors

$$\hat{x}_{kj} = x_j^{\min} + (x_j^{\max} - x_j^{\min}) \frac{\sum_{e=1}^E (2^{E-e} \cdot \tilde{b}_{k, (j-1)E + e + 1 - e})}{2^E - 1},$$

$$j \in \overline{1, M}, \quad k \in \overline{1, K}$$

c) $\hat{P} = \{(\hat{b}_k, \hat{x}_k)\}$.

6. Reduction.

6.1. Merge newly generated descendants $\hat{P}$ with the current population P

6.2. Ranking the combined population based on fitness scores, i.e., $F(x_k) < F(x_{k+1})$.

6.3. Finding the K first (best) chromosomes in the union and generation a new population P .

7. Identifying the optimal individual based on fitness evaluation

$$k^* = \arg \min_k F(x_k).$$

8. When $F(x_{k^*}) < F(x^*)$, perform $x^* = x_{k^*}$.

9. When $n < N$, perform $n = n + 1$, move to stage 3.

The result is x^* .

4 EXPERIMENTS

Experimental simulations were conducted using the scikit-opt Python module, which provides a flexible framework for implementing metaheuristic optimization

algorithms. The Pandas library was employed for data preprocessing, including linear interpolation to address missing values and tabular input/output operations.

The study utilized a comprehensive database of economic indicators (https://unctadstat.unctad.org/wds/ReportFolders/reportFolders.aspx?sCS_ChosenLang=en) covering 159 countries over a 10-year period (2012–2022). The dataset includes both macroeconomic performance metrics and digital development indicators. Key parameters used in the optimization process were as follows: population size: $K=100$, maximum number of iterations: $N=100$, initial temperature: $T_0=106$, cooling coefficient: $\beta=0.94$, middle population size: $\tilde{K}=100$, descendants number: $\hat{K}=100$, edges maximum number $A^{\max}=2$, $a=1.7$.

5 RESULTS

The cooling behavior during annealing is modeled using a temperature decay equation $T(n)=\beta^n T_0$, graphically represented in Figure 4.

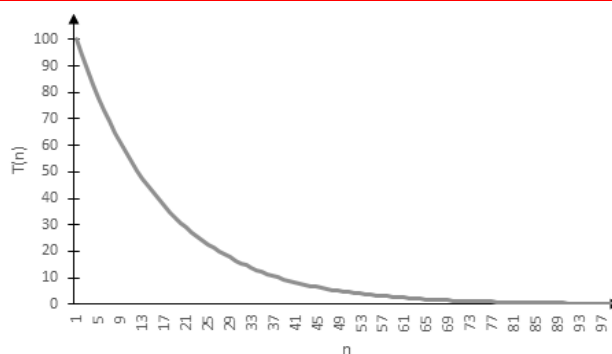


Figure 4 – Visualization of the annealing cooling strategy

The relationship between annealing temperature and iteration count, as illustrated in Figure 4, demonstrates a gradual decline in temperature as the number of iterations increases.

A comparative analysis of the proposed optimization techniques versus classical approaches is summarized in Table 1.

Table 1 – RMSE-Based evaluation of proposed and classical optimization algorithms

№	Method	Mean squared error evaluation of the proposed method	
		Proposed annealing-based algorithm	traditional technique
1	Bayesian optimization algorithm	0.04	0.08
2	Markovian optimization algorithm	0.03	0.07
3	Extended compact genetic algorithm	0.02	0.06

6 DISCUSSIONS

The parameter settings adopted in the proposed numerical optimization techniques ensure a dynamic mutation strategy: a high mutation probability is maintained during the initial iterations to promote broad exploration, while a gradual reduction in mutation probability during later stages facilitates convergence toward optimal solutions.

In contrast, classical probabilistic model-based methods typically overlook the influence of iteration count within the reproduction operator. This omission can lead to reduced solution accuracy, as evidenced by the comparative results presented in Table 1. The newly developed approaches effectively address these limitations by incorporating adaptive mechanisms that respond to the progression of the optimization process. Among the evaluated techniques, the extended compact genetic algorithm integrated with simulated annealing demonstrates the highest level of accuracy, as shown in Table 1.

CONCLUSIONS

The research addresses the challenge of low efficiency in parametric identification methods used for approximating economic indicators within intelligent economic computer systems. To enhance the performance of numerical optimization techniques, three novel methods were developed, each integrating probabilistic modeling with simulated annealing principles.

The first method, grounded in Bayesian optimization, comprises four stages: reproduction, construction of a Bayesian network, generation of offspring based on this network, and a reduction phase to refine results. The second method employs a Markov-based optimization approach. It includes reproduction, formation of a Markov network, offspring generation using Gibbs sampling within the network structure, and a final reduction step. The third method builds on the extended compact genetic algorithm. It involves reproduction, modeling through the product of limiting distributions, generation of offspring guided by this model, and a reduction phase.

These optimization strategies are designed to explore the full search space during early iterations and to focus the search direction in later stages. This dual-phase approach enhances accuracy without requiring transformation of the fitness function.

Applied relevance of the proposed methods. The proposed optimization techniques broaden the applicability of metaheuristic frameworks grounded in probabilistic modeling, offering enhanced capabilities for intelligent economic computing systems. By improving search accuracy and adaptability, these methods contribute to the development of more efficient and scalable decision-support tools within complex socioeconomic environments.

Future research will focus on extending the applicability of the proposed approach to a wider spectrum of

artificial intelligence tasks, including pattern recognition, predictive modeling, and adaptive control systems. This opens promising avenues for interdisciplinary integration and real-world deployment across diverse domains.

REFERENCES

1. Talbi E.-G. Metaheuristics: from design to implementation. New Jersey : Wiley & Sons, 2009, 618 p.
2. Yu X. Introduction to evolutionary algorithms. London, Springer-Verlag, 2010, 433 p.
3. Engelbrecht A. Computational Intelligence: an introduction. West Sussex, Wiley & Sons, 2007, 630 p.
4. Subbotin S. Diagnostic Rule Mining Based on Artificial Immune System for a Case of Uneven Distribution of Classes in Sample, *Communications*, 2016, No. 3, pp. 11–13.
5. Nakib A. Metaheuristics for Medicine and Biology. Berlin, Springer-Verlag, 2017, 211 p.
6. Glover F., Kochenberger G. A. Handbook of Metaheuristics. Dordrecht, Kluwer Academic Publishers, 2003, 570 p.
7. Shvachych G., Ivaschenko O., Busygin V., Fedorov Ye. Parallel computational algorithms in thermal processes in metallurgy and mining, *Naukovyi Visnyk Natsionalnoho Hirnychoho Universytetu*, 2018, No. 4, pp. 129–137.
8. Shlomchak G., Shvachych G., Moroz B., Fedorov E., Kozenkov D. Automated control of temperature regimes of alloyed steel products based on multi-processors computing systems, *Metalurgija*, 2019, Vol. 58, № 3–4, pp. 299–302.
9. Fedorov E., Lukashenko V., Patrushev V., Rudakov K., Mitsenko S. The method of intelligent image processing based on a three-channel purely convolutional neural network, *CEUR Workshop Proceedings*, 2018, Vol. 2255, pp. 336–351.
10. Yang X.-S. Nature-inspired Algorithms and Applied Optimization. Charm, Springer, 2018, 330 p.
11. Blum C. Hybrid Metaheuristics. Powerful Tools for Optimization. Charm, Springer, 2016, 57 p.
12. Martí R. Handbook of Heuristics. Charm, Springer, 2018, 1289 p.
13. Yang X.-S. Optimization Techniques and Applications with Examples. New Jersey, Wiley & Sons, 2018, 364 p.
14. Gendreau M. Handbook of Metaheuristics. New York, Springer, 2010, 640 p.
15. Bozorg-Haddad O. Meta-heuristic and Evolutionary Algorithms for Engineering Optimization. New Jersey, Wiley & Sons, 2017, 293 p.
16. Doerner K. F. Metaheuristics. Progress in Complex Systems Optimization. New York, Springer, 2007, 408 p.
17. de Castro L., von Zuben F. The Clonal Selection Algorithm with Engineering Applications, *Proceedings of the Genetic and Evolutionary Computation Conference “Workshop on Artificial Immune Systems and Their Applications”*, 2000, pp. 36–39.
18. Chopard B., Tomassini M. An Introduction to Metaheuristics for Optimization. New York, Springer, 2018, 230 p.
19. Radosavljević J. Metaheuristic Optimization in Power Engineering. New York. The Institution of Engineering and Technology, 2018, 536 p.
20. White J., Garrett S. Improved Pattern Recognition with Artificial Clonal Selection, *Lecture Notes in Computer Science*, 2003, Vol. 2787, pp. 181–193.
21. de Castro L. N., von Zuben F. J. Learning and optimization using clonal selection principle, *IEEE Transactions on Evolutionary Computation*, 2002, Vol. 6, pp. 239–251.
22. Babayigit B. A clonal selection algorithm for array pattern nulling by controlling the positions of selected elements, *Progress In Electromagnetics Research*, 2008, Vol. 6, pp. 257–266.
23. Brownlee J. Clever Algorithms: Nature-Inspired Programming Recipes. Melbourne, Brownlee, 2011, 436 p.
24. Alba E., Nakib A., Siarry P. Metaheuristics for Dynamic Optimization. Berlin, Springer-Verlag, 2013, 398 p.
25. Pinto C. Wasp Swarm Algorithm for Dynamic MAX-SAT Problems, *Proceedings of the 8th international conference on Adaptive and Natural Computing Algorithms*, 2007, Part I, pp. 350–357.
26. Kumar S., Mahapatra R. P. Analytical Analysis of Two-Warehouse Inventory Model Using Particle Swarm Optimization, *Proceedings of Congress on Intelligent Systems*, 2021, Vol. 2, pp. 215–226.
27. Du K.-L. Search and Optimization by Metaheuristics. Techniques and Algorithms Inspired by Nature. Charm, Springer, 2016, 434 p.
28. Sharma A., Sharma H., Khandelwal A., Sharma N. Designing Controller Parameter of Wind Turbine Emulator Using Artificial Bee Colony Algorithm, *Proceedings of Congress on Intelligent Systems*, 2020, pp. 143–151.
29. Nametala C., Faria W., Júnior B. On the performance of the Bayesian optimization algorithm with combined scenarios of search algorithms and scoring metrics, *Genetic Programming and Evolvable Machines*, 2022, No. 23, pp. 193–223.
30. Santana R., Bielza C., Larranaga P., Lozano J. A., Echegoyen C., Mendiburu A. Mateda-2.0: Estimation of Distribution Algorithms in MATLAB, *Journal of Statistical Software*, 2010, Vol. 35, pp. 1–30.
31. Nozari H., Aliahmadi A., Jafari-Eskandari M., Gh. Khaleghi An Extended Compact Genetic Algorithm for Milk Run Problem with Time Windows and Inventory Uncertainty, *International Journal of Applied Operational Research*, 2015. Vol. 5, № 2, pp. 35–48.

Received 10.06.2025.

Accepted 27.10.2025.

МЕТАЕВРИСТИЧНІ МЕТОДИ ПАРАМЕТРИЧНОЇ ІДЕНТИФІКАЦІЇ АПРОКСИМАЦІЙНОЇ МОДЕЛІ НА ОСНОВІ ІМОВІРНІСНИХ МОДЕЛЕЙ

Григор О. О. – д-р політ. наук, професор, ректор Черкаського державного технологічного університету, Черкаси, Україна.

Федоров Є. Є. – д-р техн. наук, професор, професор кафедри статистики та прикладної математики Черкаського державного технологічного університету, Черкаси, Україна.

Лещенко М. М. – канд. екон. наук, доцент, доцент кафедри міжнародної економіки та бізнесу Черкаського державного технологічного університету, Черкаси, Україна.

Рудаков К. С. – канд. техн. наук, доцент, доцент кафедри робототехніки та спеціалізованих комп'ютерних систем, Черкаського державного технологічного університету, Черкаси, Україна.

Сахно Т. А. – PhD, доцент кафедри міжнародної економіки та бізнесу Черкаського державного технологічного університету, Черкаси, Україна.

АНОТАЦІЯ

Актуальність. Для покращення якості аналізу економічних індикаторів та підвищення ефективності методів числової оптимізації було запропоновано методи, засновані на синтезі алгоритмів імовірнісних моделей та моделюванні відпаду (байєсівська оптимізація з моделюванням відпаду, марковська оптимізація з моделюванням відпаду та розширений компактний генетичний алгоритм з моделюванням відпаду). Ці методи покращують точність пошуку та не потребують трансформації функції пристосованості завдяки принципу організації дослідження всього простору пошуку на початкових ітераціях та коригування напрямку пошуку на фінальних.

Метою роботи є підвищення ефективності параметричної ідентифікації моделі апроксимації економічних показників в інтелектуальних фінансових комп'ютерних системах за допомогою метаевристичних методів, заснованих на імовірнісних моделях та моделюванні відпаду.

Методи. У дослідженні застосовано метаевристичні методи оптимізації на основі імовірнісних моделей та моделювання відпаду для покращення ефективності параметричної ідентифікації моделі апроксимації економічних показників. Запропоновано три підходи: байєсівська оптимізація з моделюванням відпаду, марковська оптимізація з моделюванням відпаду та розширений компактний генетичний алгоритм з моделюванням відпаду. Ці методи підвищують точність пошуку, досліджуючи весь простір пошуку на початкових ітераціях та коригуючи напрямок на фінальних. Байєсівська оптимізація використовує байєсівську мережу для структурованого пошуку та вдосконалення рішень. Марковська оптимізація інтегрує квантування Гіббса у марковську мережу для покращення точності пошуку. Розширений компактний генетичний алгоритм застосовує моделі граничних розподілів для генерації оптимальних рішень. Запропоновані методи усувають необхідність трансформації функції пристосованості, оптимізуючи обчислювальну ефективність.

Результати. Запропоновані методи оптимізації покращили точність параметричної ідентифікації в інтелектуальних фінансових комп'ютерних системах. Поєднання імовірнісних моделей та моделювання відпаду підвищило ефективність пошуку без необхідності трансформації функції пристосованості.

Висновки. Запропоновані методи розширюють застосування метаевристик в економічному моделюванні, покращуючи обчислювальну ефективність. Перспективами подальших досліджень є використання запропонованих методів для більш широкого класу задач машинного навчання.

КЛЮЧОВІ СЛОВА: байєсівський алгоритм оптимізації, марковський алгоритм оптимізації, розширений компактний генетичний алгоритм, моделювання відпаду, числова оптимізація, параметрична ідентифікація моделі апроксимації економічних показників.

ЛІТЕРАТУРА

1. Talbi E.-G. Metaheuristics: from design to implementation / E.-G. Talbi. – New Jersey : Wiley & Sons, 2009. – 618 p.
2. Yu X. Introduction to evolutionary algorithms / X. Yu. – London : Springer-Verlag, 2010. – 433 p.
3. Engelbrecht A. Computational Intelligence: an introduction / A. Engelbrecht. – West Sussex : Wiley & Sons, 2007. – 630 p.
4. Subbotin S. Diagnostic Rule Mining Based on Artificial Immune System for a Case of Uneven Distribution of Classes in Sample / S. Subbotin // Communications. – 2016. – No. 3. – P. 11–13.
5. Nakib A. Metaheuristics for Medicine and Biology. / A. Nakib. – Berlin : Springer-Verlag, 2017. – 211 p.
6. Glover F. Handbook of Metaheuristics. / F. Glover, G. A. Kochenberger. – Dordrecht : Kluwer Academic Publishers, 2003. – 570 p.
7. Parallel computational algorithms in thermal processes in metallurgy and mining / [G. Shvachykh, O. Ivaschenko, V. Busygin, Ye. Fedorov] // Naukovyi Visnyk Natsionalnoho Hirnychoho Universytetu. – 2018. – No. 4. – P. 129–137.
8. Automated control of temperature regimes of alloyed steel products based on multi-processors computing systems / [G. Shlomchak, G. Shvachykh, B. Moroz et al.] // Metalurgija. – 2019. – Vol. 58, № 3–4. – P. 299–302.
9. The method of intelligent image processing based on a three-channel purely convolutional neural network / [E. Fedorov, V. Lukashenko, V. Patrushev et al.] //

- CEUR Workshop Proceedings. – 2018. – Vol. 2255. – P. 336–351.
10. Yang X.-S. Nature-inspired Algorithms and Applied Optimization / X.-S. Yang. – Charm : Springer, 2018. – 330 p.
11. Blum C. Hybrid Metaheuristics. Powerful Tools for Optimization / C. Blum. – Charm : Springer, 2016. – 57 p.
12. Martí R. Handbook of Heuristics / R. Martí. – Charm : Springer, 2018. – 1289 p.
13. Yang X.-S. Optimization Techniques and Applications with Examples / X.-S. Yang. – New Jersey : Wiley & Sons, 2018. – 364 p.
14. Gendreau M. Handbook of Metaheuristics / M. Gendreau. – New York : Springer, 2010. – 640 p.
15. Bozorg-Haddad O. Meta-heuristic and Evolutionary Algorithms for Engineering Optimization / O. Bozorg-Haddad. – New Jersey : Wiley & Sons, 2017. – 293 p.
16. Doerner K. F. Metaheuristics. Progress in Complex Systems Optimization / K. F. Doerner. – New York : Springer, 2007. – 408 p.
17. de Castro L. The Clonal Selection Algorithm with Engineering Applications / L. de Castro, F. von Zuben // Proceedings of the Genetic and Evolutionary Computation Conference “Workshop on Artificial Immune Systems and Their Applications”. – 2000. – P. 36–39.
18. Chopard B. An Introduction to Metaheuristics for Optimization / B. Chopard, M. Tomassini. – New York : Springer, 2018. – 230 p.
19. Radosavljević J. Metaheuristic Optimization in Power Engineering / J. Radosavljević. – New York : The Institution of Engineering and Technology, 2018. – 536 p.
20. White J. Improved Pattern Recognition with Artificial Clonal Selection / J. White, S. Garrett // Lecture Notes in Computer Science. – 2003. – Vol. 2787. – P. 181–193.
21. de Castro L. N. Learning and optimization using clonal selection principle / L. N. de Castro, F. J. von Zuben // IEEE Transactions on Evolutionary Computation. – 2002. – Vol. 6. – P. 239–251.
22. Babayigit B. A clonal selection algorithm for array pattern nulling by controlling the positions of selected elements / B. Babayigit // Progress In Electromagnetics Research. – 2008. – Vol. 6. – P. 257–266.
23. Brownlee J. Clever Algorithms: Nature-Inspired Programming Recipes / J. Brownlee. – Melbourne : Brownlee, 2011. – 436 p.
24. Alba E. Metaheuristics for Dynamic Optimization / E. Alba, A. Nakib, P. Siarry. – Berlin : Springer-Verlag. – 2013. – 398 p.
25. Pinto C. Wasp Swarm Algorithm for Dynamic MAX-SAT Problems / C. Pinto // Proceedings of the 8th international conference on Adaptive and Natural Computing Algorithms. – 2007. – Part I. – P. 350–357.
26. Kumar S. Analytical Analysis of Two-Warehouse Inventory Model Using Particle Swarm Optimization / S. Kumar, R. P. Mahapatra // Proceedings of Congress on Intelligent Systems. – 2021. – Vol. 2. – P. 215–226.
27. Du K.-L. Search and Optimization by Metaheuristics. Techniques and Algorithms Inspired by Nature / K.-L. Du. – Charm : Springer, 2016. – 434 p.
28. Designing Controller Parameter of Wind Turbine Emulator Using Artificial Bee Colony Algorithm / [A. Sharma, H. Sharma, A. Khandelwal, N. Sharma] // Proceedings of Congress on Intelligent Systems. – 2020. – P. 143–151.
29. Nametala C. On the performance of the Bayesian optimization algorithm with combined scenarios of search algorithms and scoring metrics / C. Nametala, W. Faria, B. Júnior // Genetic Programming and Evolvable Machines. – 2022. – No. 23. – P. 193–223.
30. Mateda-2.0: Estimation of Distribution Algorithms in MATLAB / [R. Santana, C. Bielza, P. Larranaga et al.] // Journal of Statistical Software. – 2010. – Vol. 35. – P. 1–30.
31. An Extended Compact Genetic Algorithm for Milk Run Problem with Time Windows and Inventory Uncertainty / [H. Nozari, A. Aliahmadi, M. Jafari-Eskandari, Gh. Khaleghi] // International Journal of Applied Operational Research. – 2015. – Vol. 5, № 2. – P. 35–48.

FORMALIZED METHODOLOGY FOR COMPATIBILITY AND ADAPTATION OF REQUIREMENTS IN INTELLIGENT DIAGNOSTIC SYSTEMS

Komleva N. O. – PhD, Associate Professor, Head of Software Engineering Department, Odesa Polytechnic National University, Odesa, Ukraine.

Liubchenko V. V. – Dr. Sc., Professor of Software Engineering Department, Odesa Polytechnic National University, Odesa, Ukraine.

ABSTRACT

Context. Ensuring the consistency and adaptability of requirements in systems operating under dynamic conditions and limited resources is a pressing issue in modern requirements engineering, especially in intelligent diagnostic and decision-making environments. These systems must process conflicting, outdated, or ambiguous requirements while operating in environments characterized by high uncertainty and dynamic conditions.

Objective. This work introduces a formalized methodology for analyzing and managing the compatibility of system requirements. The proposed approach integrates logical consistency, functional interaction, resource feasibility, and priority alignment to support system stability and responsiveness.

Method. The methodology is implemented as a multi-level framework that incorporates formal representations of functional, non-functional, and data-related requirements. It employs scenario-based modeling, a set of compatibility assessment models, and a dynamic algorithm for integrating new requirements. The integration process includes compatibility checks, adaptive refinement, expert-based weighting, and real-time feedback. The methodology's applicability is demonstrated through a hypothetical intelligent medical diagnostic system.

Results. The proposed methodology enables systematic identification and resolution of requirement conflicts, ensuring consistent execution and effective prioritization under resource constraints. Scenario-driven modeling and the formalization of core requirements establish a foundation for adaptive system behavior and real-time decision-making.

Conclusions. The developed methodology, which includes models and algorithms, enhances the reliability of intelligent systems operating in critical contexts. Future work will focus on extending the framework by incorporating fuzzy logic, machine learning techniques, and developing software tools for automated compatibility analysis and adaptive requirements management.

KEYWORDS: methodology, software engineering, requirements engineering, requirements compatibility, scenario-based modeling, priority-based harmonization.

ABBREVIATIONS

AI is an artificial intelligence;
AHP is an Analytic Hierarchy Process;
BP is a blood pressure;
HR is a heart rate;
Hx is a patient's electronic medical history;
NLP is a natural language processing;
SpO₂ is a blood oxygen saturation;
T is a body temperature;
TOPSIS is a Technique for Order Preference by Similarity to Ideal Solution;
UML is a Unified Modeling Language.

NOMENCLATURE

Acc is a degree to which the data matches the actual values of the quality attributes;
attemptCounter is a current counter of adaptation iterations for a requirement;
C is a set of contexts;
c_i is a context for *i*-th scenario;
Com is an indicator of the presence of values for all required attributes;
Con is an indicator of the absence of contradictions among attribute values;
Cred is an evaluator of the authenticity and truthfulness of the data;
D is a set of data elements;
d is a particular data element;

L is a total available resource capacity;
MaxAttempts is a maximum number of allowed adaptation attempts;
P is a set of system parameters;
Q is a set of data quality characteristics;
q_i is an *i*-th data quality characteristic;
R is a set of system requirements;
R_{base} is a set of core (baseline) requirements;
R_{current} is an active set of consistent requirements;
R_f is a set of functional requirements;
R_{nf} is a set of non-functional requirements;
R_{new} is a set of new or updated requirements;
r_i is a particular requirement;
Rec is an evaluator of the freshness and relevance of the data for use;
S is a set of system scenarios;
s_i is an *i*-th scenario;
t is a system time;
w(r_i) is a priority weight for requirement *r_i*;
w_{limit} is a minimum acceptable weight threshold for attempting to integrate requirements;
w_{min} is a minimum weight threshold;
δ(s_i) is a criticality function for the *i*-th scenario;
ρ(r_i) is a resource consumption function;
ρ_{total} is a total resource load;
Φ is a set of preconditions for the system requirements;
φ_i is a precondition for the *i*-th requirement;

X is a set of constraint satisfaction evaluators;
 χ_i is a predicate function for the evaluation of the i -th constraint satisfaction;
 Ψ is a set of system actions;
 ψ_i is an i -th action of the system.

INTRODUCTION

As intelligent systems become increasingly complex, particularly in areas such as diagnostics and decision support, managing their requirements becomes more challenging. These systems operate in dynamic environments with limited computational resources and high degrees of uncertainty. In such environments, conventional methods of handling requirements often fall short. The effective and reliable functioning of such systems depends on their ability to manage evolving, sometimes conflicting, or ambiguous requirements in real-time.

Although many studies focus on classifying and ranking requirements, fewer explore how to resolve conflicts, manage resources, or adapt them over time. Despite substantial research, many methods still overlook how to resolve conflicts or adapt requirements under constraints [1]. In safety-critical domains such as medical diagnostics, transportation, and industrial automation, poorly coordinated or inconsistent requirements can lead to severe operational risks and failures [2].

The object of study is the process of managing system requirements under dynamic and uncertain conditions.

The subject of study is a formalized methodology designed to ensure the consistency and adaptability of requirements in intelligent diagnostic systems.

This work addresses critical limitations in existing methodologies by proposing a unified approach to functional, non-functional, and data-related requirements. The proposed solution supports real-time conflict detection, adaptive prioritization, and integration of new requirements through scenario-based modeling.

The purpose of the paper is to develop a formalized methodology for analyzing and managing the compatibility of system requirements in intelligent diagnostic systems. To achieve this goal, the following tasks are formulated:

- develop a multi-level framework for structuring requirements and operational scenarios;
- formalize logical, functional, and resource-based compatibility models;
- propose an adaptive algorithm for the integration of new requirements;
- validate the methodology through a practical example in the medical diagnostics domain.

The approach supports consistent, traceable, and context-aware requirements management across the system lifecycle. The methodology addresses key challenges faced by intelligent systems operating in dynamic and resource-constrained environments, such as resolving conflicts, adapting to new operational contexts, and maintaining stable system behavior. The proposed solution

fills existing research gaps by integrating formal compatibility analysis, scenario-based modeling, and real-time adaptability mechanisms into a unified and coherent framework. This enables adaptation to changing requirements without compromising key functions.

1 PROBLEM STATEMENT

Intelligent diagnostic systems must ensure consistent execution of requirements even under conditions of dynamic environments, limited resources, and the continuous inflow of new data. In this context, a formalized procedure is needed to verify, update, and reconcile system requirements, ensuring stable and predictable system behavior.

The system operates based on a set of requirements that includes both functional actions and quality attributes. During operation, the system may receive new or modified requirements that must be checked for compatibility with those already implemented. It is crucial to consider the limitations of available computational resources, as well as the criticality of specific usage scenarios.

As input data, we consider an existing set of system requirements, R , to which new or updated requirements R_{new} may be added. Each requirement $r_i \in R \cup R_{\text{new}}$ is characterized by a priority weight $w(r_i)$, which reflects its relative importance, and a resource estimate $p(r_i)$, representing the cost of implementing it. The total available resource capacity in the current operating context is denoted by a scalar value L .

The objective of the proposed approach is to construct a consistent set of requirements that the system can implement without internal conflicts and without exceeding the available resource limit. At the same time, the set should preserve high-priority requirements as much as possible and ensure the inclusion of functions marked as critical, based on expert assessments or scenario definitions.

A solution is considered valid if the selected set contains no logically inconsistent or mutually exclusive requirements, and the total estimated resource usage does not exceed L . To ensure operational feasibility, all included requirements must satisfy compatibility constraints across logical, functional, and resource-related dimensions. In scenarios where conflicts arise or the available resources are insufficient, the system may discard new or lower-priority requirements.

The defined problem setting serves as a basis for constructing a formal model for compatibility checking, which supports the incremental integration of new requirements and enables adaptability under dynamic operational conditions. This model is particularly relevant for critical domains, such as medical diagnostics, where inconsistencies or outdated requirements can lead to severe errors in decision-making. The proposed approach can be further extended using rule-based algorithms or formal reasoning techniques to enable automated conflict detection and intelligent requirement filtering.

2 REVIEW OF THE LITERATURE

Recent studies on requirements engineering have increasingly emphasized the integration of artificial intelligence (AI), optimization techniques, and natural language processing (NLP) to enhance the prioritization, coordination, and adaptation of requirements in dynamic software systems. Habiba et al. [3] conducted a systematic mapping study on AI-based systems, identifying persistent challenges in aligning functional and non-functional requirements. Wong, Wagner, and Treude [4] provided a classification of approaches to self-adaptive systems, underscoring the need for flexible requirement management under changing conditions.

Several researchers have explored prioritization techniques that incorporate inter-requirement dependencies, usability metrics, and AI-based enhancements. Tauqeer and Jamil [5] introduced a method that combines dependency modeling with usability assessment to improve ranking precision. Tufail et al. [6] compared prioritization strategies and demonstrated that method selection can significantly affect software quality. Abbas et al. [7] proposed a test-case prioritization approach based on requirement dependencies, which proved especially useful during later stages of development. Ahuja, Sujata, and Batra [8] applied genetic algorithms to optimize the selection of critical requirements, enhancing prioritization outcomes through evolutionary computation.

To strengthen formalization in prioritization, Leshob, Hadaya, and Renard [9] developed a goal-oriented requirements language. Hudaib et al. [10] evaluated various techniques, highlighting the importance of selecting methods tailored to project-specific factors. Tasneem et al. [11] emphasized the value of adaptive re-prioritization in Agile frameworks. Maulana et al. [12] analyzed optimization-based algorithms for large-scale projects, confirming the effectiveness of heuristic approaches.

Tariq et al. [13] developed a UML profile to support the early identification of critical requirements, improving lifecycle planning. Ashraf et al. [14] examined factors influencing conflict within development teams, stressing the importance of clearly defined and achievable expectations in requirement negotiation. Vijayakumar and Nethravathi [15] reviewed NLP-based tools for analyzing requirement specifications, highlighting their potential for automation. Similarly, Anwar and Bashir [16] surveyed AI-driven prioritization techniques and outlined the potential of hybrid models that combine multiple analytical strategies.

Recent work by Hujainah et al. [17] introduced SRPTackle, a semi-automated technique that merges user input with optimization methods to effectively handle large requirement sets. Ali et al. [18] explored prioritization methods in highly configurable systems and found a positive impact on overall software quality. Deshpande et al. [19] employed active learning and ontology-based modeling to extract requirement dependencies, enabling more informed decision-making. Saeeda et al. [20] proposed a framework for enhancing requirements elicitation

in Scrum environments, grounded in empirical insights from industrial projects.

Samer, Stettinger, and Felfernig [21] investigated the use of group recommender systems to facilitate collaborative prioritization, aiming to increase stakeholder consensus.

In summary, the reviewed literature reveals a clear trend toward automation, adaptability, and the use of intelligent techniques in requirements engineering. There is a growing emphasis on resolving inter-requirement conflicts, supporting decision-making under uncertainty, and optimizing requirement selection using AI, heuristics, and scenario-driven methods.

3 MATERIALS AND METHODS

This section outlines a formalized approach to representing system requirements, developing operational scenarios, and creating compatibility testing models and algorithms for the adaptive integration of new requirements within the dynamic environment of an intelligent diagnostic system.

System requirements are categorized into three main groups: functional, non-functional, and data requirements.

Functional requirements specify the expected behavioral logic of an intelligent system in response to defined input conditions. These requirements can be formalized using conditional-logical constructions of the form:

$$r_i \in R_f = (\varphi_i \Rightarrow \psi_i), \quad (1)$$

where $\varphi_i \in \Phi$ denotes a precondition expressed through observable features or parameter combinations, and $\psi_i \in \Psi$ represents the corresponding system action (e.g., hypothesis activation, state transition, or initiation of data processing).

Non-functional requirements impose constraints or define desired attributes of system operation, such as accuracy, response time, robustness to noise, adaptability, interpretability, and resource efficiency. These can be represented as predicates of the form:

$$r_i \in R_{nf} = \chi_i(p_k), p_k \in P, \quad (2)$$

where P is the set of system parameters, and $\chi_i \in X$ is a predicate function that evaluates constraint satisfaction.

Data Requirements. Data constitutes a critical resource in intelligent systems, forming the informational basis for decision-making, model training, and behavioral adaptation. Data requirements define the type, structure, and quality of information the system must acquire, store, process, and validate. These requirements have a significant impact on system performance and reliability.

Each data element used by the system is characterized by a set of attributes, or named features, that capture the properties of objects or events. Quality requirements apply directly to the values of these attributes. Let Q denote the set of data quality characteristics, defined as follows:

$$Q = \{Acc, Com, Con, Cred, Rec\}, \quad (3)$$

where *Acc* refers to the degree to which the data matches the actual values of the attributes, *Com* indicates whether values are present for all required attributes, *Con* represents the absence of contradictions among attribute values, *Cred* reflects the authenticity and truthfulness of the data, and *Rec* denotes the freshness and relevance of the data for its intended use. These characteristics are usually measured on a relative scale ranging from 0 to 1.

To enable formal verification of data quality, each characteristic $q_i \in Q$ is represented as a Boolean predicate applied to individual data elements. Let D be the set of data elements. For each $d \in D$, a predicate $q_i: D \rightarrow \{0,1\}$ is defined to verify compliance with a specific quality characteristic:

$$q_i: D \rightarrow \{0,1\} \quad q_i(d) = \begin{cases} 1, & \text{if } d \text{ satisfies requirement } q_i \\ 0, & \text{otherwise.} \end{cases} \quad (4)$$

The integral predicate describes the overall compliance with quality requirements:

$$Q(d) = \begin{cases} 1, & \text{if } \forall q_i \in Q, q_i(d)=1, \\ 0, & \text{otherwise.} \end{cases} \quad (5)$$

While predicates allow binary verification of compliance, practical applications often require more flexible, quantitative assessments. In cases of partial compliance or when data adaptation is required, it is beneficial to measure the degree of satisfaction with the requirement. For this purpose, each quality characteristic can be associated with a metric function that takes a value in the interval $[0, 1]$, reflecting the level of compliance with the respective requirements.

To ensure the consistency of requirements, it is necessary to formally structure the system's operation scenarios, which represent its modes of functioning under various contextual conditions.

The formation of a consistent set of requirements necessitates the explicit structuring of operation scenarios, which represent distinct modes of functioning for intelligent systems. Scenarios facilitate the specification of the sequence of system actions under defined conditions, reveal dependencies among requirements, and enable the evaluation of contextual influences on the relevance of requirements.

A system operation scenario is defined as a formalized representation of the sequence of actions, triggering conditions, and corresponding system responses associated with a specific operational context or goal-oriented mode. Each scenario encompasses a localized operational environment, along with its corresponding subset of relevant requirements.

Let C denotes the set of possible contexts (e.g., input conditions, resource availability, external events), and S represents the set of system operation scenarios. A requirement $r_i \in R$ can be represented as an ordered structure

that links a scenario, its activation context c_i , and the set of associated system actions $\psi_i \in \Psi$:

$$r_i = \langle s_i, c_i, \psi_i \rangle, \quad (6)$$

where $s_i \in S$ denotes the system scenario, $c_i \subseteq C$ defines the context under which the scenario is activated, $\psi_i \subseteq \Psi$ is the set of system actions to be performed within this scenario.

Among all scenarios $S = \{s_1, s_2, \dots, s_n\}$, a subset is identified as critical scenarios, those in which requirement violations are unacceptable due to their direct impact on system safety, stability, or goal achievement. Failures in such scenarios may lead to catastrophic consequences, including functional degradation, loss of control, or violation of essential operational constraints.

To formalize this, a Boolean criticality function is introduced:

$$\delta(s_i) \in \{0,1\}, \quad (7)$$

where $\delta(s_i) = 1$ indicates that scenario s_i is critical, $\delta(s_i) = 0$ denotes a non-critical or auxiliary scenario that supports extended functionality.

Scenario criticality is typically assessed using the following criteria:

- user expectations: the scenario delivers a primary outcome for which the system is designed;
- functional necessity: the scenario is indispensable for the system's core operation;
- safety or regulatory compliance: the scenario is mandated by standards or constraints;
- risk assessment: failure to execute the scenario may result in unacceptable operational risks;
- execution context: the criticality of a scenario may vary based on the system's current state or environment.

Thus, operation scenarios serve as a structural backbone for requirements modeling, defining the functional context within which system behavior unfolds. They establish formal connections between contextual conditions, expected system actions, and target operational states. Scenarios enable:

- the formation of a dynamic set of core (baseline) requirements;
- the verification of functional consistency within specific operating modes;
- the resolution of requirement conflicts and prioritization under resource constraints.

Furthermore, operation scenarios form the basis for constructing behavioral models of the system, which describe adaptive responses to dynamic external conditions. By incorporating scenario criticality, it becomes possible to identify and maintain a minimal subset of essential requirements that ensure continuous execution of key system functions in the most operationally significant situations.

To identify the minimally sufficient conditions for system stability and functionality, it is essential to formal-

ize a set of core requirements that reflect both universal needs and critical operational scenarios.

Core requirements represent the minimally sufficient set of conditions required to ensure the execution of a system's critically essential functions. These requirements are derived by combining two distinct subsets:

- universal requirements, which are common across all operational scenarios, and
- requirements that appear in at least one critical scenario.

Formally, the set of core (baseline) requirements R_{base} is defined as:

$$R_{base} = (\bigcap_{i=1}^n r_i) \cup (\bigcup_{\delta(s_i)=1} r_i), \quad (8)$$

where

- $\bigcap_{i=1}^n r_i$ denotes the intersection of all scenario-specific requirement sets, i.e., the set of universal requirements that are included in every system operation scenario, independent of criticality;
- $\bigcup_{\delta(s_i)=1} r_i$ represents the union of requirements associated with all critical scenarios, regardless of whether they are also present in other scenarios.

By treating universal requirements as a separate component, their independent monitoring becomes more manageable in practical scenarios. For example, when operational scenarios are modified or extended, this separation enables validation of whether the universality property of specific requirements remains intact.

This formulation ensures that both the system's global behavioral characteristics and its performance under critical conditions are taken into account. The core requirements thus serve as a foundation for:

- initial requirement alignment,
- subsequent compatibility verification of new or changing requirements, and
- decision-making in adaptive systems, particularly under resource-constrained conditions where trade-offs between requirements must be managed.

To ensure reliable system behavior, it is necessary to establish clear criteria for determining the compatibility of requirements across multiple dimensions, including logic, resources, functionality, and priorities.

Compatibility refers to the ability of two or more requirements to be executed concurrently under given contextual conditions, resource constraints, and the system's logical architecture. Identifying incompatible requirements is essential to prevent conflicts during system modeling, testing, and operation. The evaluation of requirement compatibility is conducted along the following key dimensions.

Logical consistency. Requirements must not contradict each other in their formal representations. For instance, if one requirement initiates an action under a condition ϕ , while another prohibits the same action under the same condition, a logical conflict exists. Such incon-

sistencies can be detected through formal logical analysis or the construction of requirement dependency graphs.

Resource feasibility. Logically consistent requirements may nonetheless be incompatible due to resource limitations, such as insufficient computational time, processing power, or data access. Compatibility in this context is assessed by comparing the aggregate resource demands of a requirement set to the available resource budget.

Functional interaction. Within a given scenario, requirements should be either functionally complementary or neutral. Conflicts arise when the fulfillment of one requirement modifies the system state in a way that renders another requirement infeasible or irrelevant. Functional compatibility is verified through system state transition analysis or impact modeling of requirements.

Priority consistency. In cases where requirements are mutually exclusive, a resolution must be guided by their relative priority. A priority function (e.g., weight coefficient or rank ordering) is applied to determine which requirement prevails under conflict. Compatibility is ensured if the system can resolve conflicts following this priority structure.

Importantly, incompatibility is often multi-dimensional, meaning that requirements may pass compatibility checks in one dimension but fail in another. Therefore, multi-stage analysis is recommended, with each stage evaluating compatibility in a distinct domain (logical, resource, functional, or priority-based) and considering the specific application context.

A crucial aspect of requirements management is maintaining consistency within the core set while accounting for its potential variability over time due to dynamic operational conditions and evolving system contexts.

Let R_{base} denote the set of core requirements, as defined in (8). We now formulate the consistency condition and a theorem on its temporal variability.

The following *statement* defines the condition under which the set R_{base} can be considered consistent.

The set R_{base} is said to be consistent if all pairs of requirements $r_i, r_j \in R_{base}$ satisfy the criteria of logical consistency, resource feasibility, functional interaction, and priority alignment.

If any of these criteria are violated, the structure of R_{base} must undergo preliminary alignment or adaptive modification.

The *corollary* to the above statement demonstrates the dynamic nature of R_{base} .

Since the criticality status of operational scenarios may evolve due to changes in environmental conditions, system configurations, or user priorities, the composition of R_{base} is also subject to change. Because the set changes over time, it needs to be revisited when the situation changes noticeably.

The following theorem formalizes the temporal variability of the core requirement set R_{base} under changing system conditions.

The set of core requirements R_{base} is temporally variable, i.e.,

$$R_{base} = R_{base}(t), \quad (9)$$

where t denotes system time. This implies that the composition of R_{base} may evolve due to changes in the set of scenarios S , the criticality function $\delta(s_i)$, and the structure or definition of the associated requirements.

Proof of the theorem.

Assume the contrary, i.e., R_{base} remains constant over time. Suppose the system operates in a dynamic environment, where external changes such as fluctuations in resource availability, threat levels, or technical parameters may occur. Moreover, user goals and priorities may shift over time. These factors naturally lead to the modification of scenarios S , with some scenarios being deprecated and others gaining critical status.

Since each scenario s_i is associated with a specific set of requirements r_i , any change in scenario composition or criticality status directly affects the union of requirements included in R_{base} . According to Equation (8), such changes invalidate the assumption that R_{base} is time-invariant.

Therefore, the assumption of constancy is false.

Various models can be employed to assess the compatibility of requirements, including those based on logical consistency, resource feasibility, functional interaction, and priority harmonization.

Logical analysis of condition-consequence rules is employed to detect structural incompatibilities among functional requirements specified in the form of conditional statements. A typical representation of a functional requirement $r_i \in R_f$ is as follows:

$$r_i = (\varphi_i \Rightarrow \psi_i), \quad (10)$$

where φ_i denotes the activation condition, ψ_i denotes the expected system response or action.

Under this representation, logical incompatibility may arise in several typical cases:

1. Identical Conditions with Conflicting Consequences. When $\varphi_i = \varphi_j$ but ψ_i and ψ_j are contradictory. For example:

- if temperature > 80°C $\Rightarrow$ activate_cooling
- if temperature > 80°C $\Rightarrow$ shut_down_system

2. Overlapping Conditions with Divergent Effects. If $\varphi_i \cap \varphi_j \neq \emptyset$, both conditions may be simultaneously satisfied and trigger requirements leading to different, possibly incompatible consequences. For example:

- if temperature > 70°C $\Rightarrow$ activate_alarm
- if temperature > 80°C $\Rightarrow$ stop_engine

3. Mutually Exclusive Consequences. Even when conditions are distinct, the resulting actions ψ_i and ψ_j may conflict. For example:

- $\psi_1 =$ activate_standby_mode
- $\psi_2 =$ activate_operational_mode

If these represent mutually exclusive system states, simultaneous activation leads to inconsistency.

A requirement dependency graph provides a structured means to model logical relationships among conditional requirements. Each node in the graph represents a requirement r_i , while directed edges $r_i \rightarrow r_j$ indicate a dependency, implying that:

- the consequence ψ_i of requirement r_i alters the condition φ_j of requirement r_j , or
- the actions ψ_i and ψ_j interact or contradict each other.

With the help of the dependency graph, four types of problematic situations can be identified.

1. Detection of Cyclic Dependencies. Cycles of the form $r_1 \rightarrow r_2 \rightarrow \dots \rightarrow r_n \rightarrow r_1$ indicate logical loops. These loops suggest that the execution of a requirement depends on a condition influenced by its own consequence. Such constructs can lead to:

- undefined or infinite execution behavior,
- situations where no requirement can be executed without violating preconditions.

2. Identification of Conflict Zones. A conflict zone is a subgraph where identical or overlapping conditions activate multiple requirements yet produce incompatible outcomes. These are especially critical in systems with parallel execution, where:

- conflicting requirements may compete for resources,
- simultaneous execution may produce inconsistent states.

3. Determination of Execution Order (Topological Sorting). If the dependency graph is acyclic, a topological sort can be applied to determine a valid sequence of requirement execution. This ensures:

- all prerequisites of a requirement are satisfied before its activation,
- consistency in execution logic.

For example: Check data availability $\rightarrow$ Perform computation $\rightarrow$ Output result.

4. Localization of Critical or Unstable Nodes. These are the nodes with:

- many incoming edges are condition-sensitive and may become unstable under minor context variations,
- many outgoing edges have high propagation impact and must be monitored for change management or testing prioritization.

By incorporating logical analysis and dependency graph modeling, system engineers can preemptively detect structural incompatibilities, ensure correct execution ordering, and manage the complexity of interdependent requirements in a dynamic system environment.

The resource feasibility model is employed to assess whether the cumulative consumption of system resources by active requirements remains within predefined limits. This assessment enables the detection of potential resource conflicts or system overloads before implementation or deployment.

Let $R = \{r_1, r_2, \dots, r_n\}$ denote the set of active requirements. The set R is said to be resource-feasible if the following inequality is satisfied:

$$\sum_{r_i \in R} \rho(r_i) \leq L, \quad (11)$$

where $\rho(r_i)$ is a resource consumption function that quantifies the amount of a specific system resource (e.g., CPU time, memory, bandwidth) required to fulfill requirement

r_i , L represents the resource constraint, i.e., the maximum available quantity of the resource under consideration.

This inequality ensures that the aggregate resource demand of all concurrently active requirements does not exceed the system's resource capacity. If the condition is violated, it implies that the system cannot safely support the simultaneous execution of all requirements in R . In such cases, one or more of the following strategies must be employed:

- Requirement prioritization and pruning. Eliminate or defer requirements with the lowest priority or least criticality.

- Requirement adaptation. Modify individual requirements to reduce their resource consumption, potentially through simplification, partial execution, or degradation of non-critical features.

- Resource reallocation or scaling. If feasible, increase the available resource capacity (e.g., dynamic scaling in cloud environments).

The model thus provides a quantitative basis for ensuring safe and efficient system operation under constrained conditions, especially in real-time, embedded, or resource-sensitive domains.

Functional analysis is employed to verify the consistency of requirements that operate within the scope of a single system scenario. This analysis is based on the formal representation of scenarios as triplets defined in expression (6). It involves examining whether the set of requirements within a given scenario exhibits contradictions in terms of activation conditions, expected effects, or execution order.

The following categories of conflicts are commonly identified through functional analysis:

- State conflict. A requirement alters the system state in a way that renders another requirement irrelevant or inapplicable. For example, the requirement “switch the system to standby mode” conflicts with “start diagnostics”, as standby mode may disable diagnostic operations.

- Conflict of execution conditions. Two requirements are activated under partially overlapping conditions but prescribe contradictory system actions. For example, one requirement triggers a shutdown when the system temperature exceeds 80°C, while another activates emergency cooling under the same or similar conditions.

- Effect inconsistency. Improper sequencing of requirement execution leads to erroneous system behavior. For example, attempting to read data from a sensor after the sensor has been powered down results in system failure or invalid outputs.

Scenarios identified as critical, as formalized in expression (7), require particular attention. Functional inconsistencies within such scenarios may compromise system safety, stability, or goal attainment. Therefore, ensuring the functional consistency of requirements in critical scenarios is a key step in the validation and reliability assurance process.

In scenarios where the simultaneous execution of all system requirements is infeasible due to logical conflicts or resource constraints, the system must selectively priori-

tize the most critical requirements. This is accomplished through priority-based harmonization, which utilizes weight coefficients to quantify the relative importance of each requirement.

Each requirement $r_i \in R$ is associated with a priority weight $w(r_i) \in [0,1]$, representing its contextual importance. These weights can be derived through expert judgment, analytical techniques, or data-driven methods, including historical analysis or dependency-based inference.

When a conflict between requirements arises, the system retains the requirement r_i for which $w(r_i) > w(r_j)$, ensuring that higher-priority requirements are preserved.

To resolve conflicts or address execution constraints, the system selects a subset of requirements based on their priorities. The following algorithmic strategies are commonly applied:

- Acceptability Threshold Strategy. A predefined minimum threshold, $w_{\min}$, is established. Only requirements satisfying $w(r_i) \geq w_{\min}$ are eligible for execution. This approach filters out low-priority requirements without requiring additional compatibility checks. The threshold may be static or dynamically adjusted based on the operational context.

- Greedy Sorting Strategy. All requirements are sorted in descending order of weight. Each requirement is sequentially evaluated for compatibility with the current active set. If compatible, it is included; otherwise, it is either discarded or deferred. This strategy ensures the maximum inclusion of high-priority requirements within constraints.

- Compromise Harmonization Strategy. Requirements with lower weights may be accepted, provided they do not conflict with those already selected. This approach enhances system flexibility by enabling the execution of non-critical yet beneficial requirements, particularly when strict prioritization is not mandatory.

In addition to basic strategies, more advanced or context-sensitive approaches may be applied:

- Multi-Criteria Ranking. Requirements are evaluated across multiple dimensions, such as benefit, risk, cost, and aggregated via AHP, TOPSIS, or similar methods.

- Heuristic and Evolutionary Algorithms. Techniques such as genetic algorithms or ant colony optimization are employed to identify near-optimal subsets of requirements in the presence of complex interdependencies and constraints.

- Contextual Harmonization Based on Scenarios. The priority weight $w(r_i)$ is interpreted relative to the specific scenario in which the requirement is applied. A requirement may have low priority in routine operation but become critical in emergency or failure scenarios. This adaptive approach ensures that system behavior remains aligned with environmental demands.

- Group Expert Harmonization. The importance of requirements is evaluated through a collective expert assessment, using aggregation methods such as the median or weighted voting to form a consensus-based priority structure.

– Stochastic Methods. Weights are modeled as random variables or probability distributions, allowing harmonization strategies to function under uncertainty and incomplete knowledge.

Priority-based harmonization provides a systematic framework for managing competing requirements under limitations. By combining weighted importance with compatibility analysis, the system can adaptively optimize its functionality, achieving an effective balance between goal satisfaction, resource efficiency, and operational safety.

The integration of new requirements into a software system is a critical task that demands rigorous checks for compatibility, relevance, and priority. To preserve system consistency and prevent operational conflicts, this process must be governed by structured algorithmic mechanisms that support informed decision-making.

To support structured integration of new requirements, a baseline algorithm is proposed that ensures compatibility validation, prioritization, and bounded adaptation.

At system initialization, the active set of consistent requirements is defined as:

$$R_{\text{current}} := R_{\text{base}}$$

where R_{base} represents the baseline functionality, guaranteed to be conflict-free and executable.

Let R_{new} denote the set of new requirements to be considered for integration; $w(r_i) \in [0,1]$ be the priority weight of requirement r_i ; w_{limit} be the minimum acceptable weight threshold for attempting integration; MaxAttempts

be the maximum number of adaptation attempts allowed; *attemptCounter* be the current counter of adaptation iterations for a requirement.

The integration process proceeds as follows:

1. Sorting. Sort all requirements in R_{new} in descending order by their weights $w(r_i)$.
2. Sequential Processing. For each requirement $r_i \in R_{\text{new}}$:
 - Perform a compatibility check with the current set R_{current} :
 - If compatible, add r_i directly to R_{current} ;
 - If incompatible, continue to Step 3.
3. Priority Evaluation
 - If $w(r_i) < w_{\text{limit}}$, reject r_i without adaptation;
 - If $w(r_i) \geq w_{\text{limit}}$, proceed to adaptation.
4. Adaptation Loop
 - Initialize *attemptCounter* := 0;
 - While r_i remains incompatible and *attemptCounter* < *MaxAttempts*:
 - o Attempt to refine or modify r_i ;
 - o Re-evaluate compatibility with R_{current} ;
 - o Increment *attemptCounter*.
5. Post-Adaptation Decision
 - If the modified r_i becomes compatible, add it to R_{current} ;
 - Otherwise, permanently reject r_i .

Figure 1 illustrates the integration flow, depicting the logical transitions between evaluation stages, compatibility checks, and adaptation outcomes.

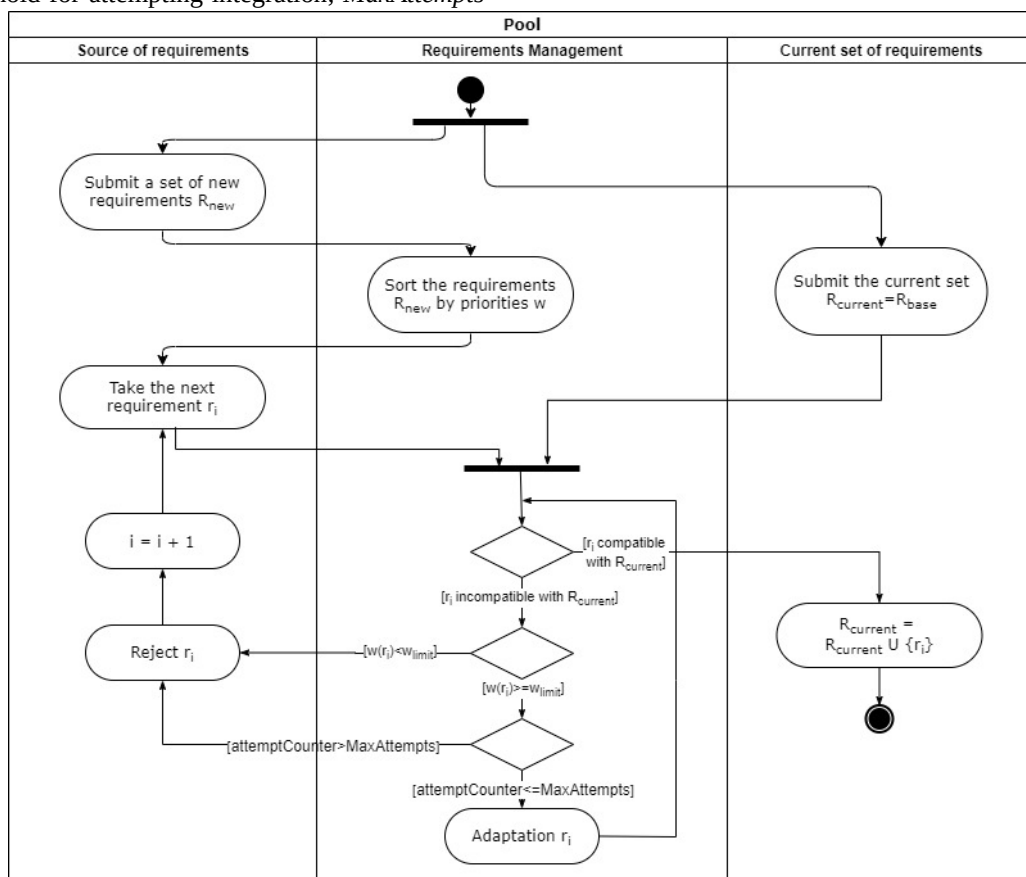


Figure 1 – Basic requirement integration algorithm

We introduce a structured algorithm to align new requirements with the current system configuration. It enables incremental evolution of the requirement set while ensuring that consistency is maintained through compatibility validation and bounded adaptation attempts. Unlike expert- or context-driven methods, this model assumes a static decision logic, which is suitable for baseline harmonization.

During system operation, runtime feedback can be collected regarding the failure patterns of specific requirements, e.g., those that consistently trigger conflicts, violate constraints, or fail due to environmental conditions. This feedback may be used to adjust the priority weights $w(r_i)$ dynamically.

Such adjustments support the development of an adaptive prioritization mechanism, where:

- frequently violated or inapplicable requirements may have their weight reduced;
- persistently successful or mission-critical requirements may have their weight increased;
- requirements with declining utility can be excluded from critical scenarios.

By leveraging operational feedback, the system gradually evolves into a context-aware and conflict-resilient behavior, thereby improving its ability to reconcile incoming requirements while maintaining stability and resource efficiency.

4 EXPERIMENTS

To demonstrate the applicability of the proposed requirement alignment and adaptation model, we present a hypothetical use case of an intelligent diagnostic system designed for real-time medical monitoring in an intensive care unit. The system continuously evaluates patient physiological parameters and generates diagnostic actions to enable timely intervention in critical clinical situations.

The system acquires real-time data from a network of physiological sensors, which measure the following parameters:

- body temperature (T);
- heart rate (HR);
- blood oxygen saturation (SpO₂);
- blood pressure (BP).

Additionally, the system utilizes information from the patient's electronic medical history (Hx) for contextual analysis.

The system supports the execution of three essential diagnostic scenarios:

- s_1 : Detection of hyperthermia,
- s_2 : Recognition of hypoxia,
- s_3 : Shock prediction and warning.

Each scenario is operationalized through a corresponding set of functional requirements, structured according to the representation model introduced in (6).

Table 1 outlines six functional requirements associated with scenarios. Each requirement includes the activation context, the prescribed system action, and the relevant scenario.

Table 1 – System Requirements and Scenario Association

Requirement	Context	System actions	Scenario
r_1	$T > 38^\circ\text{C}$	Activate cooling	s_1
r_2	$\text{SpO}_2 < 90\%$	Trigger alarm	s_2
r_3	$\text{BP} < 90/60$	Check heart rate	s_3
r_4	$\text{HR} > 110$	Raise suspicion of tachycardia	s_3
r_5	Hx contains "stroke"	Increase the priority of SpO ₂ check	s_2
r_6	$T > 38^\circ\text{C} \wedge \text{SpO}_2 < 85\%$	Activate complex response	$s_1 \cap s_2$

Figure 2 presents a dependency graph illustrating the logical interdependencies and potential conflicts among the system requirements.

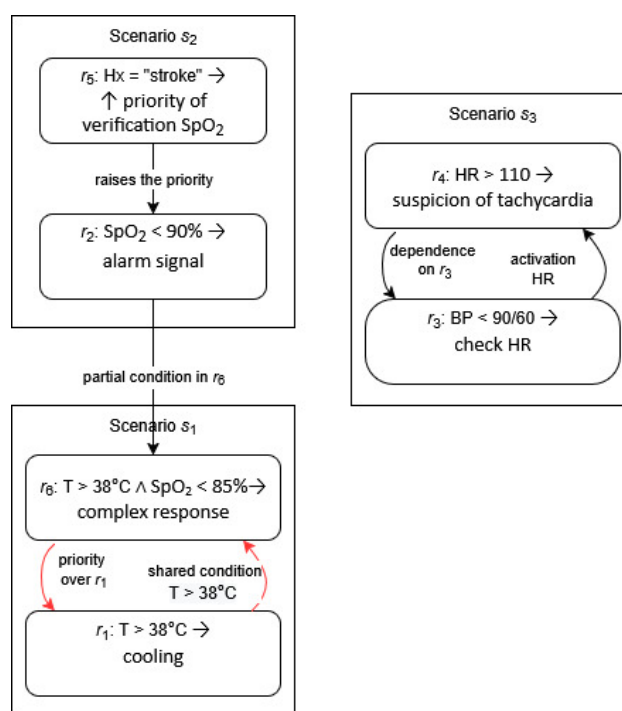


Figure 2 – Graph of dependencies between system requirements with conflict indications

In scenario s_3 , a cyclic interdependence exists between r_3 and r_4 : a drop in blood pressure (r_3) triggers HR monitoring, which, if elevated (r_4), indicates a risk of tachycardia, potentially reinforcing the check.

In scenario s_2 , requirement r_5 increases the priority of r_2 when the patient's history includes a diagnosis of stroke. Additionally, r_2 is logically a subset of r_6 , which activates a complex response in the presence of combined hyperthermia and severe hypoxia.

A notable conflict arises between r_1 and r_6 . Both are activated by the same temperature threshold ($T > 38^\circ\text{C}$), but prescribe different actions: cooling (r_1) and a complex response (r_6), which may be mutually incompatible. This conflict, indicated in red in the graph, necessitates priority-based harmonization.

5 RESULTS

To assess system feasibility under operational constraints, Table 2 provides the estimated resource consumption, $\rho(r_i)$, and priority weights, $w(r_i)$, for each requirement. Let the maximum allowable resource load be limited to 100 units.

Table 2 – Requirement Resource Usage and Priority

Requirement	Action description	$\rho(r_i)$	$w(r_i)$
r_1	Activate cooling	20	0.7
r_2	Trigger alarm	10	0.9
r_3	Check heart rate	15	0.6
r_4	Raise suspicion of tachycardia	20	0.5
r_5	Increase the priority of SpO ₂ check	10	0.8
r_6	Activate complex response	50	1.0

As shown in the table, the highest priority is assigned to response r_6 , while the lowest belongs to r_4 . Resource usage varies from 10 to 50 units, enabling the evaluation of acceptable requirement combinations under the 100-unit constraint.

Feasibility analysis indicates that the execution of r_6 in combination with any two additional requirements remains within the acceptable load. However, a configuration involving r_1 , r_6 , r_4 , and r_5 reaches the upper resource limit. If all six requirements are simultaneously triggered, the total resource load reaches 125 units, exceeding the system's capacity.

The system must harmonize requirements based on their relative priorities to resolve conflicts under constraints. In the case of a conflict between r_1 and r_6 , the higher priority of r_6 ($w = 1.0$) prevails over r_1 ($w = 0.7$), and r_6 is selected.

Assume that the system detects the following conditions concurrently:

- $T > 38^\circ\text{C}$,
- $\text{SpO}_2 < 90\%$,
- $\text{BP} < 90/60$,
- $\text{HR} > 110$.

This scenario activates the requirements: r_1 , r_2 , r_3 , r_4 , and r_6 . The corresponding total resource load is:

$$\rho_{\text{total}} = 20(r_1) + 10(r_2) + 15(r_3) + 20(r_4) + 50(r_6) = 115 \text{ units.}$$

Since this exceeds the 100-unit limit, a greedy algorithm is applied. Requirements are sorted by descending priority: $r_6(1.0)$, $r_2(0.9)$, $r_1(0.7)$, $r_3(0.6)$, $r_4(0.5)$.

The system sequentially includes requirements until the cumulative load reaches the limit:

- include $r_6 \rightarrow \rho_{\text{total}} = 50$;
- include $r_2 \rightarrow \rho_{\text{total}} = 60$;
- include $r_1 \rightarrow \rho_{\text{total}} = 80$;
- include $r_3 \rightarrow \rho_{\text{total}} = 95$;
- r_4 is excluded (would exceed the limit).

Thus, the system executes r_6 , r_2 , r_1 , and r_3 , while r_4 is dropped due to insufficient resources.

These results are obtained via simulation-based evaluation of requirement interaction models within the developed formal framework.

Thus, the model enables a quantitative assessment of requirement compatibility and supports decision-making regarding inclusion/exclusion based on priorities and available resources.

6 DISCUSSIONS

The proposed method demonstrates that adaptive harmonization, grounded in compatibility, prioritization, and resource constraints, supports reliable, context-aware system behavior. Even under conflicting conditions and limited capacity, the system prioritizes mission-critical functionality, ensuring continuity of care and system consistency.

These findings demonstrate that the proposed methodology significantly enhances the reliability of intelligent diagnostic systems operating in real-time, resource-constrained environments. By applying formalized compatibility checks and dynamic priority-based filtering, the system effectively balances operational feasibility with functional adequacy.

Unlike conventional approaches that rely on static prioritization or fixed thresholds, the proposed model introduces structured multi-factor evaluation. For instance, while rule-based engines are commonly used in medical expert systems, they often lack built-in mechanisms to resolve requirement conflicts under resource constraints. Our method enhances these capabilities by integrating logical consistency, functional dependencies, and resource feasibility into the modeling process.

Unlike approaches that mainly focus on requirement classification and selection, our framework directly addresses internal inconsistencies and contextual variability. Moreover, the integration of scenario-based reasoning enables the adaptive inclusion of critical functionality, even in rapidly changing conditions, the factor that has been underrepresented in prior models.

A key advantage of the proposed model is its applicability across a wide range of mission-critical domains. While the use case involves medical diagnostics, the methodology is equally applicable to other mission-critical domains, such as industrial monitoring, autonomous systems, and cyber-physical infrastructures.

Nonetheless, the current implementation has several limitations that warrant attention. The current implementation assumes that input parameters, such as priority weights and resource estimates, are externally defined, which may introduce subjectivity or inconsistency in real-world deployment. Additionally, the conflict resolution algorithm used in the simulation is greedy and may not always produce globally optimal configurations.

Further research is required to explore the integration of machine learning techniques for the automated estimation of requirement parameters, as well as advanced optimization heuristics to manage trade-offs among competing requirements better.

CONCLUSIONS

This paper presents a systematic methodology for ensuring the compatibility and adaptability of requirements

in intelligent diagnostic systems operating under dynamic environments and limited resources. The proposed approach addresses critical shortcomings of conventional techniques, which often emphasize prioritization but insufficiently consider logical consistency, resource constraints, and context-awareness in requirement harmonization.

The key contributions of this work are as follows:

- formal structure that distinguishes functional, non-functional, and data-related requirements, treating data-specific constraints as a distinct and formally defined category;
- scenario-driven framework for determining the criticality of requirements and constructing a dynamic baseline set of core requirements;
- multi-dimensional compatibility analysis model that integrates logical consistency, resource feasibility, functional interaction, and priority-based reasoning;
- adaptive integration algorithm that incorporates compatibility checks, expert-informed prioritization, and feedback-driven weight adjustments.

A practical evaluation using a hypothetical intelligent medical diagnostic system demonstrated the effectiveness of the proposed models in detecting requirement conflicts, improving system consistency, and supporting stable behavior under operational constraints.

Future research will focus on developing a prototype tool for automated compatibility analysis and adaptive requirements management. Additionally, the framework will be expanded to incorporate fuzzy logic, machine learning techniques, and scenario-based multi-criteria optimization, thereby supporting advanced real-time decision-making.

REFERENCES

1. Radwan A. M., Abdel-Fattah M. A., Mohamed W. AI-Driven Prioritization Techniques of Requirements in Agile Methodologies: A Systematic Literature Review, *International Journal of Advanced Computer Science and Applications*, 2024, Vol. 15, No. 9, pp. 812–823. DOI: 10.14569/IJACSA.2024.0150983.
2. Heyn H.-M., Knauss E., Muhammad A. P. et al. Requirement Engineering Challenges for AI-intense Systems Development, *arXiv*, 2021. DOI: 10.48550/arXiv.2103.10270.
3. Habiba U., Haug M., Bogner J. et al. How mature is requirements engineering for AI-based systems? A systematic mapping study on practices, challenges, and future research directions, *Requirements Engineering*, 2024, Vol. 29, pp. 567–600. DOI: 10.1007/s00766-024-00432-3.
4. Wong T., Wagner M., Treude C. Self-adaptive systems: A systematic literature review across categories and domains, *Information and Software Technology*, 2022, Vol. 148, 106934. DOI: 10.1016/j.infsof.2022.106934.
5. Tauqeer M., Jamil H. Requirements Prioritization – Modeling Through Dependency and Usability with Fusion of Artificial Intelligence Technique, *International Journal of Innovative Science and Technology*, 2025, Vol. 7, No. 1, pp. 146–158.
6. Tufail H., Qasim I., Masoo M. F. et al. Towards the Selection of Optimum Requirements Prioritization Technique: A Comparative Analysis, *IEEE 5th International Conference on Information Management : proceedings*. Cambridge, UK, 2019, pp. 227–231. DOI: 10.1109/INFOMAN.2019.8714709.
7. Abbas M., Inayat I., Saadatmand M. et al. Requirements Dependencies-Based Test Case Prioritization for Extra-Functional Properties, *IEEE 12th International Conference on Software Testing, Verification and Validation : proceedings*. Xi'an, China, 2019, pp. 159–163. DOI: 10.1109/ICSTW.2019.00045.
8. Ahuja H., Sujata, Batra U. Performance Enhancement in Requirement Prioritization by Using Least-Squares-Based Random Genetic Algorithm, *Innovations in Computational Intelligence*, 2018, Vol. 713, pp. 251–263. DOI: 10.1007/978-981-10-4555-4_17.
9. Leshob A., Hadaya P., Renard L. Software Requirements Prioritization with the Goal-Oriented Requirement Language, *Lecture Notes in Data Engineering and Communications Technologies*, 2020, Vol. 41, pp. 187–198. DOI: 10.1007/978-3-030-34986-8_13.
10. Hudaib A., Masadeh R., Qasem M. H. et al. Requirements Prioritization Techniques Comparison, *Modern Applied Science*, 2018, Vol. 12, No. 2, pp. 62–80. DOI: 10.5539/mas.v12n2p62.
11. Tasneem N., Zulzalil H. B., Hassan S. Enhancing Agile Software Development: A Systematic Literature Review of Requirement Prioritization and Reprioritization Techniques, *IEEE Access*, 2025, Vol. 13, pp. 32993–33034. DOI: 10.1109/access.2025.3539357.
12. Maulana M. Z. N., Siahaan D., Saikhu A. et al. Optimization Algorithm for Prioritizing Software Requirements: A Comparative Study, *7th International Seminar on Research of Information Technology and Intelligent Systems : proceedings*. Yogyakarta, 2024, pp. 284–289. DOI: 10.1109/ISRITI64779.2024.10963649.
13. Tariq A., Azam F., Anwar M. W. et al. A UML Profile for Prediction of Significant Software Requirements, *IEEE 10th Annual Conference on Information Technology, Electronics and Mobile Communication : proceedings*. Vancouver, 2019, pp. 979–984. DOI: 10.1109/IEMCON.2019.8936227.
14. Ashraf M., Tubaishat A., Al-Obeidat F. et al. Managerial Conflict Among the Software Development Team, *Lecture Notes in Networks and Systems*, 2022, Vol. 350, pp. 331–341. DOI: 10.1007/978-981-16-7618-5_29.
15. Vijayakumar S., Nethravathi P. S. Use of Natural Language Processing in Software Requirements Prioritization – A Systematic Literature Review, *International Journal of Applied Engineering and Management Letters*, 2021, Vol. 5, No. 2, pp. 152–174. DOI: 10.47992/IJAEM.2581.7000.0110.
16. Anwar R., Bashir M. B. A Systematic Literature Review of AI-Based Software Requirements Prioritization Techniques, *IEEE Access*, 2023, Vol. 11, pp. 143815–143860. DOI: 10.1109/ACCESS.2023.3343252.
17. Hujainah F., Abu Bakar R. B., Nasser A. B. et al. SRPTackle: A semi-automated requirements prioritisation technique for scalable requirements of software system projects, *Information and Software Technology*, 2021, Vol. 131, 106501. DOI: 10.1016/j.infsof.2020.106501.
18. Ali A., Hafeez Y., Hussain S. et al. Role of Requirement Prioritization Technique to Improve the Quality of Highly-Configurable Systems, *IEEE Access*, 2020, Vol. 8, pp. 27549–27573. DOI: 10.1109/ACCESS.2020.2971382.
19. Deshpande G., Motger Q., Palomares C. et al. Requirements Dependency Extraction by Integrating Active Learning with Ontology-Based Retrieval, *IEEE 28th International Confer-*

- ence on Requirements Engineering : proceedings. Zurich, 2020, pp. 78–89. DOI: 10.1109/RE48521.2020.00020.
20. Saeeda H., Dong J., Wang Y. et al. A Proposed Framework for Improved Software Requirements Elicitation Process in SCRUM: Implementation by a Real-Life Norway-Based IT Project, *Journal of Software: Evolution and Process*, 2020, Vol. 32, Issue 7, e2247. DOI: 10.1002/smr.2247.
21. Samer R., Stettinger M., Felfernig A. Group Recommender User Interfaces for Improving Requirements Prioritization, *28th ACM Conference on User Modeling, Adaptation and Personalization : proceedings*. New York, 2020, pp. 221–229. DOI: 10.1145/3340631.3394851.

Received 07.07.2025.
Accepted 16.10.2025.

УДК004.421.2

ФОРМАЛІЗОВАНА МЕТОДОЛОГІЯ УЗГОДЖЕННЯ ТА АДАПТАЦІЇ ВИМОГ В ІНТЕЛЕКТУАЛЬНИХ ДІАГНОСТИЧНИХ СИСТЕМАХ

Комлева Н. О. – канд. техн. наук, доцент, завідувач кафедри програмної інженерії Національного університету «Одеська політехніка», Одеса, Україна.

Любченко В. В. – д-р техн. наук, професор кафедри програмної інженерії Національного університету «Одеська політехніка», Одеса, Україна.

АНОТАЦІЯ

Актуальність. Забезпечення узгодженості та адаптивності вимог у системах, що функціонують в умовах динамічного середовища та обмежених ресурсів, є актуальною проблемою сучасної інженерії вимог, особливо в інтелектуальних діагностичних та експертно-рішальних системах. Такі системи мають обробляти суперечливі, застарілі або неоднозначні вимоги під час роботи в складних умовах.

Мета роботи – розробка формалізованої методології аналізу та керування сумісністю вимог в інтелектуальних діагностичних системах. Запропонований підхід інтегрує логічну узгодженість, функціональну взаємодію, ресурсну здійсненність та узгодження пріоритетів для забезпечення стабільності та адаптивності системи.

Метод. Методологія реалізована як багаторівнева структура, що включає формальні представлення функціональних, нефункціональних та вимог, пов'язаних з обробкою даних. Вона використовує моделювання на основі сценаріїв, набір моделей оцінки сумісності та динамічний алгоритм інтеграції нових вимог. Процес інтеграції включає перевірку сумісності, адаптивне уточнення, експертне зважування та зворотний зв'язок у реальному часі. Застосовність методології демонструється на гіпотетичному прикладі інтелектуальної медичної діагностичної системи.

Результати. Запропонована методологія дозволяє систематично виявляти та усувати конфлікти між вимогами, забезпечуючи узгоджене виконання та ефективну пріоритизацію в умовах обмежених ресурсів. Сценарне моделювання та формалізація базових вимог створюють основу для адаптивної поведінки системи та прийняття рішень у реальному часі.

Висновки. Розроблена методологія, яка включає моделі та алгоритми, підвищує надійність інтелектуальних систем, що працюють у критичних умовах. У майбутніх дослідженнях передбачається розширення структури шляхом впровадження нечіткої логіки, методів машинного навчання та розробки програмних засобів для автоматизованого аналізу сумісності та адаптивного керування вимогами.

КЛЮЧОВІ СЛОВА: методологія, програмна інженерія, інженерія вимог, сумісність вимог, моделювання на основі сценаріїв, узгодження на основі пріоритетів.

ЛІТЕРАТУРА

1. Radwan A. M. AI-Driven Prioritization Techniques of Requirements in Agile Methodologies: A Systematic Literature Review / A. M. Radwan, M. A. Abdel-Fattah, W. Mohamed // *International Journal of Advanced Computer Science and Applications*. – 2024. – Vol. 15, No. 9. – P. 812–823. DOI: 10.14569/IJACSA.2024.0150983.
2. Requirement Engineering Challenges for AI-intense Systems Development / [H.-M. Heyn, E. Knauss, A. P. Muhammad et al.] – Access mode: <https://arxiv.org/abs/2103.10270>. DOI: 10.48550/arXiv.2103.10270.
3. How mature is requirements engineering for AI-based systems? A systematic mapping study on practices, challenges, and future research directions / [U. Habiba, M. Haug, J. Bogner et al.] // *Requirements Engineering*. – 2024. – Vol. 29. – P. 567–600. DOI: 10.1007/s00766-024-00432-3.
4. Wong T. Self-adaptive systems: A systematic literature review across categories and domains / T. Wong, M. Wagner, C. Treude // *Information and Software Technology*. – 2022. – Vol. 148. – P. 106934. DOI: 10.1016/j.infsof.2022.106934.
5. Tauqeer M. Requirements Prioritization – Modeling Through Dependency and Usability with Fusion of Artificial Intelligence Technique / M. Tauqeer, H. Jamil // *International Journal of Innovative Science and Technology*. – 2025. – Vol. 7, No. 1. – P. 146–158.
6. Towards the Selection of Optimum Requirements Prioritization Technique: A Comparative Analysis / [H. Tufail, I. Qasim, M. F. Masoo et al.] // *Information Management : 5th International Conference, Cambridge, UK, 24–27 March 2019 : proceedings*. – Cambridge: IEEE, 2019. – P. 227–231. DOI: 10.1109/INFOMAN.2019.8714709.
7. Requirements Dependencies-Based Test Case Prioritization for Extra-Functional Properties / [M. Abbas, I. Inayat, M. Saadatmand et al.] // *Software Testing, Verification and Validation Workshops : IEEE 12th International Conference, Xi'an, China, 22–23 April 2019 : proceedings*. – Los Alamitos: IEEE, 2019. – P. 159–163. DOI: 10.1109/ICSTW.2019.00045.
8. Ahuja H. Performance Enhancement in Requirement Prioritization by Using Least-Squares-Based Random Genetic Algorithm / H. Ahuja, Sujata, U. Batra // *Innovations in Computational Intelligence*. – 2018. – Vol. 713. – P. 251–263. DOI: 10.1007/978-981-10-4555-4_17.
9. Leshob A. Software Requirements Prioritization with the Goal-Oriented Requirement Language / A. Leshob,

- P. Hadaya, L. Renard // *Lecture Notes in Data Engineering and Communications Technologies*. – 2020. – Vol. 41. – P. 187–198. DOI: 10.1007/978-3-030-34986-8_13.
10. Requirements Prioritization Techniques Comparison / [A. Hudaib, R. Masadeh, M. H. Qasem et al.] // *Modern Applied Science*. – 2018. – Vol. 12, No. 2. – P. 62–80. DOI: 10.5539/mas.v12n2p62.
11. Tasneem N. Enhancing Agile Software Development: A Systematic Literature Review of Requirement Prioritization and Reprioritization Techniques / N. Tasneem, H. B. Zulzalil, S. Hassan // *IEEE Access*. – 2025. – Vol. 13. – P. 32993–33034. DOI: 10.1109/access.2025.3539357.
12. Optimization Algorithm for Prioritizing Software Requirements: A Comparative Study / [M. Z. N. Maulana, D. Siahaan, A. Saikhu et al.] // *Research of Information Technology and Intelligent Systems : 7th International Seminar, Yogyakarta, Indonesia, 11 December 2024 : proceedings*. – Yogyakarta: IEEE, 2024. – P. 284–289. DOI: 10.1109/ISRITI64779.2024.10963649.
13. A UML Profile for Prediction of Significant Software Requirements / [A. Tariq, F. Azam, M. W. Anwar et al.] // *Information Technology, Electronics and Mobile Communication : IEEE 10th Annual Conference, Vancouver, British Columbia, Canada, 17–19 October 2019 : proceedings*. – Vancouver: IEEE, 2019. – P. 979–984. DOI: 10.1109/IEMCON.2019.8936227.
14. Managerial Conflict Among the Software Development Team / [M. Ashraf, A. Tubaishat, F. Al-Obeidat et al.] // *Lecture Notes in Networks and Systems*. – 2022. – Vol. 350. – P. 331–341. DOI: 10.1007/978-981-16-7618-5_29.
15. Vijayakumar S. Use of Natural Language Processing in Software Requirements Prioritization – A Systematic Literature Review / S. Vijayakumar, P. S. Nethravathi // *International Journal of Applied Engineering and Management Letters*. – 2021. – Vol. 5, No. 2. – P. 152–174. DOI: 10.47992/IJAEML.2581.7000.0110.
16. Anwar R. A Systematic Literature Review of AI-Based Software Requirements Prioritization Techniques / R. Anwar, M. B. Bashir // *IEEE Access*. – 2023. – Vol. 11. – P. 143815–143860. DOI: 10.1109/ACCESS.2023.3343252.
17. SRPTackle: A semi-automated requirements prioritisation technique for scalable requirements of software system projects / [F. Hujainah, R. B. Abu Bakar, A. B. Nasser et al.] // *Information and Software Technology*. – 2021. – Vol. 131. – P. 106501. DOI: 10.1016/j.infsof.2020.106501.
18. Role of Requirement Prioritization Technique to Improve the Quality of Highly-Configurable Systems / [A. Ali, Y. Hafeez, S. Hussain et al.] // *IEEE Access*. – 2020. – Vol. 8. – P. 27549–27573. DOI: 10.1109/ACCESS.2020.2971382.
19. Requirements Dependency Extraction by Integrating Active Learning with Ontology-Based Retrieval / [G. Deshpande, Q. Motger, C. Palomares et al.] // *Requirements Engineering : IEEE 28th International Conference, Zurich, Switzerland, 31 August – 4 September 2020 : proceedings*. – Los Alamitos: IEEE, 2020. – P. 78–89. DOI: 10.1109/RE48521.2020.00020.
20. A Proposed Framework for Improved Software Requirements Elicitation Process in SCRUM: Implementation by a Real-Life Norway-Based IT Project / [H. Saeeda, J. Dong, Y. Wang et al.] // *Journal of Software: Evolution and Process*. – 2020. – Vol. 32, Issue 7. – e2247. DOI: 10.1002/smr.2247.
21. Samer R. Group Recommender User Interfaces for Improving Requirements Prioritization / R. Samer, M. Stettinger, A. Felfernig // *User Modeling, Adaptation and Personalization : 28th ACM Conference, Genoa, Italy, 14–17 July 2020 : proceedings*. – New York : ACM, 2020. – P. 221–229. DOI: 10.1145/3340631.3394851.

RESEARCH ON A HYBRID LSTM-CNN-ATTENTION MODEL FOR TEXT-BASED WEB CONTENT CLASSIFICATION

Kuz M. V. – Dr. Sc., Professor of the Department of Information Technology Vasyl Stefanyk Precarpathian National University, Ukraine.

Lazarovych I. M. – PhD, Associate Professor of the Department of Information Technology Vasyl Stefanyk Precarpathian National University, Ukraine.

Kozlenko M. I. – PhD, Associate Professor of the Department of Information Technology Vasyl Stefanyk Precarpathian National University, Ukraine.

Pikuliak M. V. – PhD, Associate Professor of the Department of Information Technology Vasyl Stefanyk Precarpathian National University, Ukraine.

Kvasniuk A. D. – Graduate student, Vasyl Stefanyk Precarpathian National University, Ukraine.

ABSTRACT

Context. Text-based web content classification plays a pivotal role in various natural language processing (NLP) tasks, including fake news detection, spam filtering, content categorization, and automated moderation. As the scale and complexity of textual data on the web continue to grow, traditional classification approaches – especially those relying on manual feature engineering or shallow learning techniques – struggle to capture the nuanced semantic relationships and structural variability of modern web content. These limitations result in reduced adaptability and poor generalization performance on real-world data. Therefore, there is a clear need for advanced models that can simultaneously learn local linguistic patterns and understand the broader contextual meaning of web text.

Method. This study presents a hybrid deep learning architecture that integrates Long Short-Term Memory (LSTM) networks, Convolutional Neural Networks (CNN), and an Attention mechanism to enhance the classification of web content based on text. Pre-trained GloVe embeddings are used to represent words as dense vectors that preserve semantic similarity. The CNN layer extracts local n -gram patterns and lexical features, while the LSTM layer models long-range dependencies and sequential structure. The integrated Attention mechanism enables the model to focus selectively on the most informative parts of the input sequence. The model was evaluated using the dataset, which consists of over 10,000 HTML-based web pages annotated as legitimate or fake. A 5-fold cross-validation setup was used to assess the robustness and generalizability of the proposed solution.

Results. Experimental results show that the hybrid LSTM-CNN-Attention model achieved outstanding performance, with an accuracy of 0.98, precision of 0.94, recall of 0.92, and F1-score of 0.93. These results surpass the performance of baseline models based solely on CNNs, LSTMs, or transformer-based classifiers such as BERT. The combination of neural network components enabled the model to effectively capture both fine-grained text structures and broader semantic context. Furthermore, the use of GloVe embeddings provided an efficient and effective representation of textual data, making the model suitable for integration into systems with real-time or near-real-time requirements.

Conclusions. The proposed hybrid architecture demonstrates high effectiveness in text-based web content classification, particularly in tasks requiring both syntactic feature extraction and semantic interpretation. By combining convolutional, recurrent, and attention-based mechanisms, the model addresses the limitations of individual architectures and achieves improved generalization. These findings support the broader use of hybrid deep learning approaches in NLP applications, especially where complex, unstructured textual data must be processed and classified with high reliability.

KEYWORDS: web content classification, LSTM-CNN-Attention, deep learning, natural language processing, GloVe embeddings, text classification, hybrid model, sequence modeling.

ABBREVIATIONS

CNN is a Convolutional Neural Network;

LSTM is a Long Short-Term Memory;

GloVe is a set of Global Vectors for Word Representation;

AUC is an Area Under the Curve.

NOMENCLATURE

D is a set of documents (e.g., HTML content of web pages);

f is a classification function mapping documents or features to predicted classes;

R^m is a feature space after text preprocessing;

θ is a set of trainable model parameters;

$L(\theta)$ is a loss function measuring prediction error;

θ^* is an optimal parameter set minimizing the loss function;

True Positive (TP) is a number of web pages that the model correctly classified as belonging to the target category (e.g., fake news or spam);

False Positive (FP) is a number of web pages that the model incorrectly classified as belonging to the target category when they do not;

False Negative (FN) is a number of web pages that the model incorrectly classified as not belonging to the target category when they actually do;

True Negative (TN) is a number of web pages that the model correctly classified as not belonging to the target category;

TPR is a True Positive Rate;

FPR is a False Positive Rate, which are plotted on the ROC curve;

$x^{(s)}$ is a s -th input text sample;

$w^{(t)}$ is a t -th word in the sequence;

E is a pre-trained embedding matrix;
 V is a vocabulary size;
 d is a dimensionality of the word embeddings;
 e_t is an embedding vector of the word w_t ;
 $X^{(s)}$ is a matrix of embeddings for sample s ;
 f (in CNN layer) is a convolutional filter;
 k is a kernel width of the filter;
 $X_{i:i+k-1}$ is a sliding window input over the embeddings;
 b is a bias term in CNN;
 $\langle \cdot, \cdot \rangle$ is an element-wise product followed by summation;
 c_i is a feature after convolution and activation;
 C is a feature map;
 $\hat{c}$ is a result of max-pooling;
 x_t is an input vector at time step t in the LSTM;
 h_t is a cell state at time step t ;
 $\sigma(\cdot)$ is a sigmoid activation function;
 $\circ$ is an element-wise (Hadamard) product;
 i_t, f_t, o_t are input, forget, and output gates in the LSTM;
 $\tilde{c}_t$ is a candidate cell state;
 W_i, W_f, W_o, W_c are weight matrices for the input;
 U_i, U_f, U_o, U_c are recurrent weight matrices;
 b_i, b_f, b_o, b_c are bias vectors for each gate;
 H is a matrix of LSTM hidden states;
 u is a context vector in the attention mechanism;
 α_t is an attention weight for time step t ;
 c (in attention) is a context vector, a weighted sum of hidden states;
 $\hat{s}$ is a predicted probability distribution over classes;
 W_c is a weight matrix in the classification layer;
 b_c is a bias term in the classification layer.

INTRODUCTION

The explosive growth of web-based textual data has led to increasing demand for intelligent systems capable of automatically classifying web content. Tasks such as fake news detection, spam filtering, opinion mining, and automated content moderation rely heavily on accurate and scalable classification methods. A major challenge in this domain is the diversity and complexity of natural language used across different types of web content, which often includes both structured and unstructured text embedded in HTML pages.

Traditional classification systems used to detect such deceptive content often rely on manually designed features, blacklists, or rule-based heuristics [1]. While these approaches can be effective in specific, known scenarios, they typically struggle to generalize to new or sophisticated cases, especially as the structure and language of web content continue to evolve. Their limited adaptability

and dependence on static features reduce their effectiveness in large-scale or dynamic environments.

In contrast, modern deep learning approaches [2] – particularly those incorporating Convolutional Neural Networks (CNN), Long Short-Term Memory (LSTM) networks, and Attention mechanisms – offer a powerful alternative. These models can learn meaningful patterns directly from raw textual input, capturing both local structures and long-range dependencies. They are well suited for web content classification tasks that require the extraction of complex semantic and syntactic features from natural language.

The object of study is the process of classifying text-based web content using a hybrid deep learning model that integrates CNN, LSTM, and Attention components. The model is trained on labeled HTML-based web pages that include both legitimate and deceptive content, with the goal of automatically assigning them to the correct class.

The subject of study is the development of a hybrid neural architecture that combines the strengths of CNNs, LSTMs, and Attention mechanisms for improved classification of web-based textual data.

The purpose of the work is to enhance the effectiveness and generalization capabilities of web content classification models by leveraging advanced deep learning techniques. The aim is to improve both feature extraction and sequence modeling, ultimately leading to a more robust and accurate system for identifying deceptive or misleading textual web content.

1 PROBLEM STATEMENT

The classification of web content based on textual information is a fundamental task in natural language processing. It plays a critical role in applications such as topic categorization [3], content moderation [4], spam detection, and personalized content delivery. Given the large volume and unstructured nature of online data, there is a growing need for intelligent models that can automatically analyze and classify web content with high accuracy.

The web content classification task can be formulated as a supervised multi-class or binary classification problem over a set of documents. Let:

- $D = \{d_1, d_1, \dots, d_n\}$ as a set of documents (HTML content of web pages);
- where each $d_i \in D$ has an associated label $y_i \in \{0, 1\}$, with 0 representing a standard (legitimate) page and 1 representing a target class of interest.

The objective is to find a classification function (1):

$$f : D \rightarrow \{0, 1\}. \quad (1)$$

Since HTML documents are textual in nature, they are typically preprocessed and transformed into numerical representations before being used in classification models.

Therefore, the problem can be equivalently defined in the feature space as (2):

$$f : R^m \rightarrow \{0,1\}. \quad (2)$$

The goal is to optimize the classification function's parameters θ in such a way that a chosen loss function $L(\theta)$ is minimized (3):

$$\theta^* = \arg \min_{\theta} L(\theta). \quad (3)$$

2 REVIEW OF THE LITERATURE

Initial methods for web content classification were rule-based or relied on classic machine learning models, such as Naïve Bayes [5], Support Vector Machines (SVM) [5], and Decision Trees [5]. These models worked by extracting features from web pages, such as URL characteristics and HTML content. Despite their utility, these models had several limitations. They lacked contextual awareness of the text, and their performance degraded when handling large datasets or sophisticated content structures.

For example, Naïve Bayes and SVMs were constrained by the simplicity of the features they utilized, often failing to capture the complex structure of web pages. Moreover, these models required extensive manual tuning and were prone to misclassifying web pages that did not exhibit obvious, rule-based features.

With the advent of natural language processing (NLP) techniques, newer methods based on word embeddings, such as TF-IDF [6], Word2Vec [6], and GloVe, began to emerge. These methods improved traditional feature extraction by capturing the semantic meaning of words and their contextual relationships [7]. However, while TF-IDF represented a significant improvement, it lacked the ability to account for word order, which was crucial for understanding the structure of the content. Word2Vec enhanced this by embedding words in continuous vector spaces that preserved their meanings in context. Despite this, Word2Vec still struggled with understanding the global context of the text, making it less effective for long-form web pages or complex content.

GloVe addressed some of these challenges by factorizing the word co-occurrence matrix to capture global word relationships. However, as noted in the study, GloVe's use of pre-trained word vectors allowed for faster training and reduced computational complexity compared to models like BERT, which required significant resources for fine-tuning [8].

Recent advances in deep learning have significantly impacted web content classification. Convolutional Neural Networks (CNNs) [9, 10] and Long Short-Term Memory (LSTM) networks [11] have been successfully applied to text classification tasks due to their ability to model both local and sequential dependencies, respectively. CNNs are particularly effective for detecting local patterns, such as identifying suspicious phrases or structural

features in web pages. However, CNNs struggle with capturing long-term dependencies in textual data.

In contrast, LSTMs excel at modeling sequential data and are better at preserving the long-term context in web content. While LSTMs are more effective at handling longer sequences, they can still encounter difficulties in modeling very long-term dependencies and require significant training time.

The introduction of transformer-based models, such as BERT [12], marked a new era in text classification. BERT, with its attention mechanism, is capable of understanding the context of words within a sentence by considering the relationships between all words in the text. This makes it highly effective for understanding complex, context-dependent content. However, transformer models are computationally expensive and require fine-tuning on domain-specific data to achieve optimal results.

Although BERT shows superior performance in tasks like web content classification, its high computational cost has prompted researchers to explore lighter alternatives. For instance, the hybrid CNN-LSTM-Attention model proposed in this study offers a good balance between performance and computational efficiency. This model combines the strengths of CNNs and LSTMs, with the addition of an Attention mechanism that highlights the most critical parts of the text, improving both the interpretability and accuracy of the model.

3 MATERIALS AND METHODS

For web content classification in this study, a dataset containing web pages in HTML code format, labeled as either “phishing” or “legitimate”, was used. This dataset [13] is publicly available and was originally presented as an SQL archive. To facilitate its usage, the data was extracted, opened in the HeidiSQL environment, and exported to the CSV format, which is more convenient for processing in Python.

The dataset consists of three tables, but only the websites table was used for this research. Two key columns were selected from it:

- htmlContent, which contains the HTML code of the web pages;
- isPhish, which indicates whether a page contains deceptive web content or legitimate content.

After preprocessing, which includes text cleaning, lemmatization, bigram generation, and tokenization, the Phishload dataset contains 10.373 records, of which 9.198 are deceptive web pages and 1.176 are legitimate ones. This distribution highlights a significant class imbalance, which may affect model training, as the majority of samples belong to one class.

Fig. 1 illustrates the class distribution in this dataset, showing the dominance of deceptive web pages over legitimate ones, which necessitates the use of specialized balancing techniques.

Fig. 2 illustrates the distribution of HTML code length in the Phishload dataset, allowing for an assessment of the differences between deceptive and legitimate web pages.

Text preprocessing is a crucial step in machine learning and natural language processing tasks [14]. It enables the transformation of textual data into a more structured and analyzable format, significantly improving the efficiency of classification algorithms. This is particularly relevant for harmful content detection, as fraudulent messages often contain word variations, grammatical distortions, and excessive information, complicating analysis.

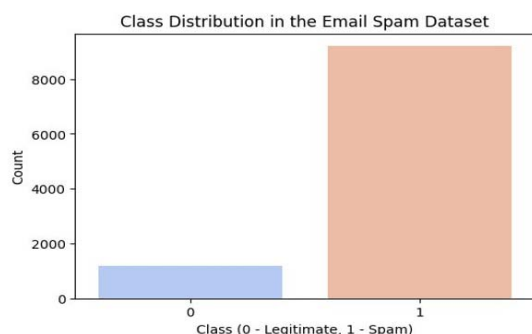


Figure 1 – The class distribution in the HTML dataset

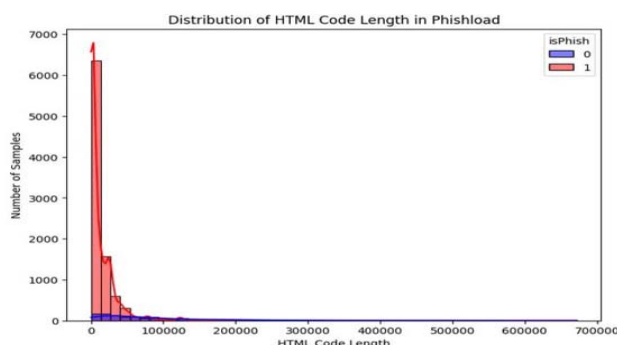


Figure 2 – The distribution of HTML code length in the dataset

This study examines key text preprocessing methods used for analyzing and classifying web content. The main methods include lemmatization, tokenization, stop-word removal, and bigram formation. These techniques enhance the quality of textual data, positively impacting the accuracy of classification models.

Tokenization splits text into smaller units (tokens). Deceptive web pages may use obfuscation (e.g., 'p@ssw0rd') to evade detection. NLTK [15] ensures precise segmentation. Lemmatization reduces words to their base form (e.g., 'running' → 'run') to minimize redundancy. spaCy was used for effective context-aware processing.

Stop-word removal eliminates common words ('the', 'is', etc.) to retain only relevant terms. Harmful web content often includes manipulative phrases ('urgent', 'click here'), which were filtered using NLTK.

Bigrams capture meaningful word pairs ("credit_card") to enhance context understanding. The generate_ngrams function created bigrams for improved text analysis.

Applying tokenization, lemmatization, stop-word removal, and bigram formation significantly improves the quality of textual data before passing it into a machine learning model. This, in turn, enhances the efficiency of classification algorithms, allowing for more accurate detection of harmful content.

Text vectorization [16] is a crucial step in NLP tasks, as textual data must be transformed into a format that machine learning algorithms can process.

Fig. 3 illustrates the sequence of key text preprocessing steps before feeding data into a machine learning model.

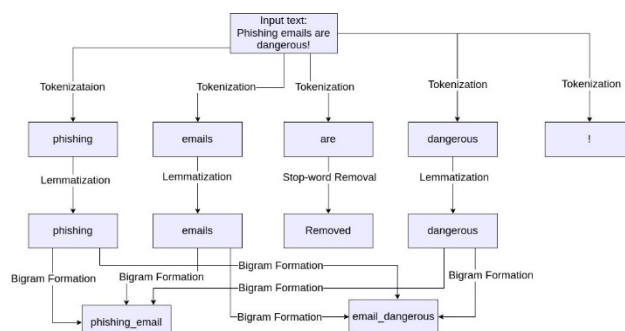


Figure 3 – General process of text preprocessing for web content classification

In this study, a comparative analysis of key vectorization approaches was conducted using several evaluation metrics to assess their effectiveness in the context of web content classification. Specifically, the models were compared based on the following criteria:

Precision and recall. These metrics were used to evaluate the effectiveness of each vectorization method in correctly identifying harmful content. Precision and recall are defined as (4, 5):

$$\text{Precision} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Positives}}. \quad (4)$$

$$\text{Recall} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Negatives}}. \quad (5)$$

F1 Score. The F1 score was calculated as (6):

$$F1 = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}. \quad (6)$$

Accuracy. The overall classification accuracy was computed as (7):

$$\text{Accuracy} = \frac{\text{True Positives} + \text{True Negatives}}{\text{Total Samples}}. \quad (7)$$

In addition to the above metrics, AUC was specifically utilized to further evaluate and compare the overall classification performance of the models, providing insight into their ability to distinguish between relevant and irrelevant web content across different classification thresholds.

The AUC score reflects the model's ability to separate positive (target category) and negative (non-target category) classes over all possible classification thresholds. It is computed as the area under the ROC curve (Receiver Operating Characteristic curve), which plots the true positive rate against the false positive rate, as (8):

$$AUC = \int_0^1 TRP(t) d(FPR(t)). \quad (8)$$

The following methods were employed for text vectorization:

- TF-IDF – a classical method that evaluates term importance in a document.
- Word2Vec – a neural network-based method that generates vector representations based on context.
- GloVe – a statistical method that captures word co-occurrences within a corpus.
- BERT – a transformer-based approach for contextual text understanding.

TF-IDF evaluates term frequency and is known for its simplicity and interpretability; however, it does not consider the context of words. Word2Vec trains a neural network based on word context, allowing it to preserve semantic relationships, but it requires a large corpus for effective training. GloVe factorizes a word co-occurrence matrix [17], making it well-suited for capturing word analogies, yet it demands significant computational resources. BERT, which utilizes transformers, achieves high accuracy in complex tasks but comes with a high computational cost.

A hybrid approach is proposed that combines CNN, LSTM, and an Attention mechanism, leveraging the strengths of each architecture. CNN effectively extracts local patterns in the text, allowing it to identify potentially suspicious phrases and structures. LSTM retains long-term context, which is crucial for analyzing sequences and understanding the overall meaning of a message. The Attention mechanism [18–19] further enhances the model by focusing on the most important words in the text, strengthening relevant features for classification.

The architecture of the hybrid model consists of several key stages. First, the input text is processed through a vectorization layer using GloVe. Next, CNN detects local patterns and key textual features, after which LSTM processes the obtained representations to preserve sequential context. The Attention mechanism identifies the most significant words that are critical for classification, and the final output is passed to a fully connected layer for decision-making.

The combination of local and global text analysis results in a more accurate model for web content classification. Integrating CNN and LSTM with the Attention mechanism enhances performance compared to using each method individually. This approach improves the model's ability to capture subtle linguistic patterns and semantic context, making it more adaptable and effective for handling diverse and complex textual data in real-world applications.

To formalize the operation of the proposed CNN-LSTM-Attention architecture, a mathematical description of each component is presented below.

Let $x^{(s)} = [w_1, w_2, \dots, w_T]$ denote the s -th text sample, consisting of T words.

To convert words into dense numerical representations, a pre-trained word embedding matrix $E \in R^{V \times d}$ is used [20]. The resulting embedding sequence is (9):

$$X^{(s)} = [e_1, e_2, \dots, e_T] \in R^{T \times d}, \quad e_t = E(w_t). \quad (9)$$

The convolutional layer is used to capture local patterns in the text [21]. Let $f \in R^{k \times d}$ denote a convolutional filter of width k . The convolution operation over a window $X_{i:i+k-1}$ is defined as (10):

$$c_i = \text{ReLU}(\langle f, X_{i:i+k-1} \rangle) + b. \quad (10)$$

A feature map is formed as (11):

$$C = [c_1, c_2, \dots, c_{T-k+1}]. \quad (11)$$

To reduce dimensionality, a max-pooling operation is applied (12):

$$\hat{c} = \max(C). \quad (12)$$

The Long Short-Term Memory (LSTM) layer models long-range dependencies in the sequence [22]. The input is represented as (13):

$$X = [x_1, \dots, x_T]. \quad (13)$$

For each time step t , the LSTM computes (14):

$$\begin{aligned} i_t &= \sigma(W_i x_t + U_i h_{t-1} + b_i), \\ f_t &= \sigma(W_f x_t + U_f h_{t-1} + b_f), \\ o_t &= \sigma(W_o x_t + U_o h_{t-1} + b_o), \\ \tilde{c}_t &= \tanh(W_c x_t + U_c h_{t-1} + b_o), \\ c_t &= f_t \circ c_{t-1} + i_t \circ \tilde{c}_t, \\ h_t &= o_t \circ \tanh(c_t). \end{aligned} \quad (14)$$

To focus on the most informative parts of the sequence, a soft attention mechanism is applied. Let $H = [h_1, h_2, \dots, h_T]$ be the sequence of LSTM hidden states [23]. The attention weight α_t for each time step is computed as (15):

$$\alpha_t = \frac{\exp(h_t^T u)}{\sum_{j=1}^T \exp(h_j^T u)}. \quad (15)$$

The context vector c , which summarizes the weighted hidden states, is calculated as (16):

$$c = \sum_{t=1}^T \alpha_t \times h_t. \quad (16)$$

The resulting context vector c , is passed through a fully connected layer with a softmax activation to obtain class probabilities (17):

$$\hat{s} = \text{softmax}(W_c c + b_c), \quad \hat{s} \in R^2. \quad (17)$$

The model is trained by minimizing the cross-entropy loss function [24] (18):

$$L = -\sum_{s=1}^N y^{(s)} \times \log(\hat{s}^{(s)}). \quad (18)$$

Fig. 4 illustrates the architecture of the CNN, LSTM, and Attention mechanism.

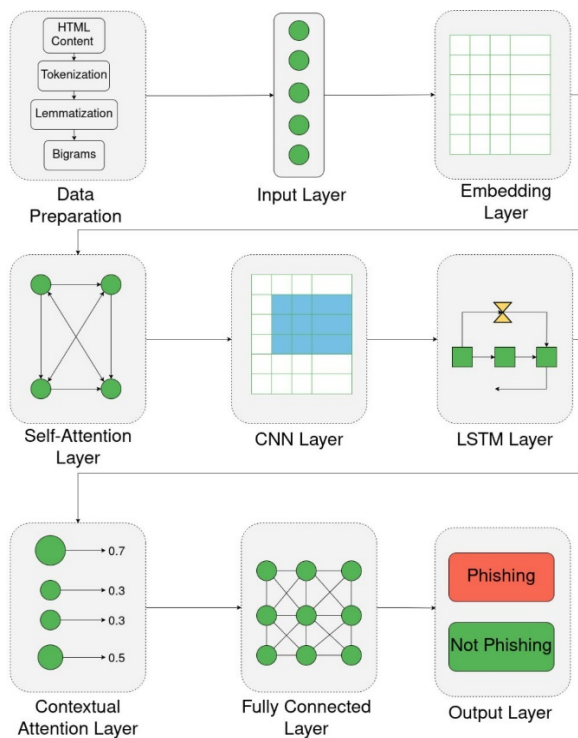


Figure 4 – Architecture of a Text-Based Web Content Classification Model Using CNN, LSTM, and Attention

At the initial stage, the model receives input data in the form of an HTML document, which has been preprocessed into a sequence of 200 tokens. This ensures a uniform input length, which is a prerequisite for further neural processing. Tokenization typically retains only the most frequent elements (e.g., top-5000).

Each input token is mapped into a 100-dimensional vector using pre-trained GloVe embeddings. This enables the model to preserve semantic relationships between words before training even begins. The parameter trainable=False ensures that the embeddings remain unchanged during training, promoting stability and generalization.

Once the tokens are embedded, convolutional operations are applied to extract local patterns in the HTML sequence. A 1D convolution with 128 filters of size 5 captures characteristic fragments of code (e.g., tag patterns or suspicious sequences of attributes). MaxPooling with a pool size of 5 reduces the dimensionality while retaining the most salient features.

To model long-range dependencies in HTML documents, the architecture employs a multi-head self-attention mechanism. This allows the model to attend to key structural elements, regardless of their position in the document. For example, a <form> tag at the beginning may be closely related to a <submit> button at the end. Functionally, this component is described as Multi-HeadAttention(num_heads=4, key_dim=64).

To retain sequential context, the model incorporates a Long Short-Term Memory (LSTM) network. LSTM captures both positional and structural dependencies in HTML, which is especially critical for processing nested tags. The return_sequences=True setting ensures that the entire sequence is preserved for the next attention mechanism.

On top of the LSTM outputs, a trainable contextual attention mechanism is applied. This layer computes the context vector by focusing on the most relevant parts of the sequence, effectively summarizing the input into a weighted representation for final classification.

To mitigate overfitting, the model applies a Dropout layer with a dropout rate of 30%, randomly deactivating neurons during training. This improves the robustness and generalization ability of the model.

The final layer is a dense [25] classifier with a softmax activation function, which produces a probability distribution over the two output classes: relevant (1) or irrelevant (0) web content.

4 EXPERIMENTS

For deceptive content classification, an experiment was conducted to assess model performance. K-Fold cross-validation [26] with $K = 5$ was applied, splitting the dataset into 80% for training and 20% for testing. The hyperparameters set are as follows: embedding dimension is 100, LSTM units are 128, optimizer is adam, learning rate is 0.001, batch size is 32, and the number of epochs is 20. This approach ensures a more objective evaluation,

reduces dependency on specific dataset partitions, and improves the reliability of the results.

Among the four text vectorization methods (BERT, GloVe, Word2Vec, TF-IDF), the GloVe method was selected for the following reasons:

Stable High Accuracy. The model with GloVe quickly achieves consistently high accuracy on both training and validation data (Fig. 5). This indicates that GloVe effectively extracts key semantic features of the text.

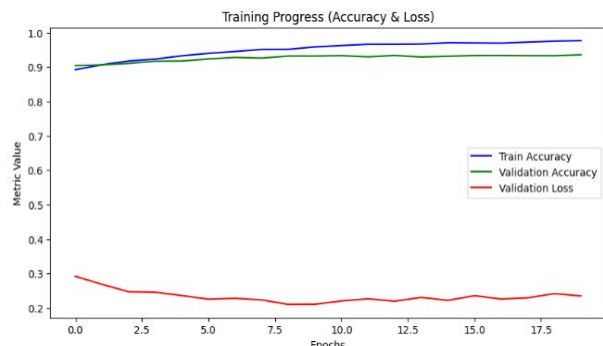


Figure 5 – The training process: changes in accuracy and loss

Fig. 6–7 present the results after performing K-Fold cross-validation for CNN, LSTM, and the Attention mechanism.



Figure 6 – Training process after performing K-Fold cross-validation for CNN, LSTM, and the Attention mechanism: changes in accuracy and loss

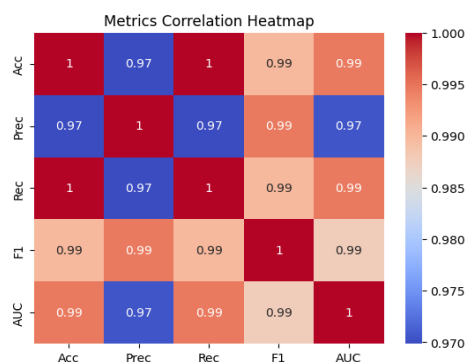


Figure 7 – Correlation analysis between Accuracy, Precision, Recall, F1, and AUC after performing K-Fold cross-validation for CNN, LSTM, and the Attention mechanism

The presented results demonstrate that the model maintains consistently high accuracy on both training and validation data, effectively distinguishing between deceptive and legitimate messages. This is also confirmed by the balance of the Accuracy, Precision, Recall, F1-score, and AUC metrics, indicating the model's stability without bias toward any particular class.

5 RESULTS

The performance metrics presented in tables were obtained using the dataset [13], which contains web pages in HTML code format labeled as either “phishing” or “legitimate”.

From Table 1, we can observe the performance of different methods in harmful content classification based on accuracy, precision, recall, and F1-score. TF-IDF shows the lowest performance across all metrics, with an accuracy of 0.88 and an F1-score of 0.61, indicating its limitations in capturing contextual relationships. Word2Vec and GloVe outperform TF-IDF, achieving 0.93 and 0.94 accuracy, respectively. Both methods maintain high precision (0.87 and 0.88) and recall (0.87 and 0.86), resulting in an F1-score of 0.87 for both. This suggests that these word embedding techniques effectively capture semantic relationships. BERT achieves an accuracy of 0.92, comparable to Word2Vec and GloVe. However, while its precision (0.88) is on par with GloVe, it has a lower recall (0.79), leading to a slightly lower F1-score (0.81). This could indicate that BERT's performance is affected by the dataset size or training constraints. Overall, GloVe provides the highest accuracy (0.94) and maintains a strong balance between precision and recall, making it the best-performing method in this comparison.

From Table 2, the results indicate that the model achieves high accuracy and strong generalization ability. The addition of the attention mechanism improves the extraction of key textual features, enhancing classification quality. Fig. 7 illustrates the balance between accuracy, precision, recall, F1-score, and AUC, which is essential for deceptive content detection. The combination of CNN and LSTM effectively captures both local and global dependencies in text, reducing the gap between training and validation accuracy (Fig. 6). While the model requires more resources than the baseline LSTM, it is significantly faster than BERT, making it an optimal choice for harmful content classification.

Table 1 – Comparison of text preprocessing methods for risky content analysis

Method	Accuracy	Precision	Recall	F1-score
TF-IDF	0.88	0.59	0.65	0.61
Word2Vec	0.93	0.87	0.87	0.87
GloVe	0.94	0.88	0.86	0.87
BERT	0.92	0.88	0.79	0.81

Table 2 – Metrics of the trained models

Method	Accuracy	Precision	Recall	F1-score
CNN	0.95	0.91	0.89	0.90
LSTM	0.96	0.92	0.88	0.89
BERT	0.95	0.91	0.87	0.88
CNN + LSTM	0.97	0.93	0.90	0.91
CNN + LSTM + ATTENTION	0.98	0.94	0.92	0.93

6 DISCUSSION

The results of the experiments clearly show that the proposed hybrid CNN-LSTM-Attention model is highly effective for text-based web content classification. The model demonstrates significant improvements in accuracy and overall classification performance compared to traditional methods and other deep learning-based approaches. The integration of convolutional layers for capturing local text patterns, LSTM for modeling sequential dependencies, and the attention mechanism for emphasizing important features results in a robust and accurate model. Additionally, the model's strong generalization across different dataset partitions makes it well-suited for practical applications involving large-scale and diverse web content.

The results also indicate that the use of pre-trained GloVe embeddings significantly improves the performance of the model by providing rich semantic representations of the text. The use of K-Fold cross-validation ensures that the evaluation is not biased towards specific splits, and the consistent performance across folds highlights the model's robustness.

Additionally, the comparative analysis of different text vectorization techniques (TF-IDF, Word2Vec, GloVe, BERT) provides insights into their strengths and weaknesses. While TF-IDF offers interpretability, it fails to capture contextual meanings, leading to lower performance. Word2Vec and GloVe achieve better results by capturing semantic relationships, with GloVe outperforming the others due to its ability to model word co-occurrences effectively. BERT, although highly effective in understanding context, requires significant computational resources and may not always generalize well with limited training data.

The experimental results indicate that GloVe provides the highest accuracy (0.94) and maintains a strong balance between precision and recall, making it the best-performing vectorization method in this study. However, despite its effectiveness, further research could explore fine-tuned transformer-based models to enhance text-based web content classification.

Furthermore, the importance of text preprocessing techniques such as tokenization, lemmatization, stop-word removal, and bigram formation has been demonstrated. These methods significantly contribute to improv-

ing model performance by refining the textual input and enhancing its quality before vectorization.

The integration of the attention mechanism in the CNN-LSTM model further improves classification performance by identifying key words that contribute to the semantic understanding of web content. The addition of this mechanism allows the model to focus on the most relevant parts of the input text, reducing noise and enhancing decision-making accuracy.

Finally, while the proposed model achieves high accuracy and robustness, it is important to consider computational efficiency. The hybrid approach offers a balanced trade-off between accuracy and processing speed, making it suitable for real-time or near-real-time content classification systems. Future work may focus on optimizing the model for lower computational costs while maintaining high classification capabilities, as well as testing its effectiveness on larger and more diverse datasets.

In conclusion, the hybrid CNN-LSTM-Attention model presents a promising approach to tackling the challenges of text-based web content classification, achieving high accuracy and robustness. Further advancements could explore additional feature engineering techniques, alternative embedding methods, or ensemble learning strategies to enhance overall classification performance.

CONCLUSIONS

A hybrid approach combining CNN, LSTM, and Attention is proposed to improve the accuracy of text-based web content classification. The optimized model demonstrates consistently high performance, achieving strong results on both training and validation datasets. The Attention mechanism plays a crucial role in enhancing feature identification, which refines the model's classification capabilities. The balanced correlation of key evaluation metrics such as Accuracy, Precision, Recall, F1-score, and AUC indicates the model's stability and helps minimize class bias. Additionally, the model excels at generalization by effectively handling both local and global text dependencies, which helps reduce the gap between training and validation accuracy. While the model requires more computational resources than LSTM, it is still significantly faster than BERT, maintaining high recognition quality. Overall, the proposed hybrid model improves deceptive content detection accuracy by approxi-

mately 0.19%–0.37% compared to other contemporary approaches, such as those described in [5], [6], [9], [11], and [12], further demonstrating its effectiveness in real-world scenarios.

The scientific novelty of the proposed approach lies in the integration of a hybrid CNN-LSTM-Attention architecture for text-based web content classification. This model combines the strengths of Convolutional Neural Networks (CNN) for capturing local linguistic patterns, Long Short-Term Memory (LSTM) for modeling long-range sequential dependencies, and the Attention mechanism for identifying and emphasizing the most informative features of the text. This combination enhances the model's ability to accurately classify diverse and complex web-based textual content. The incorporation of pre-trained GloVe word embeddings further improves semantic understanding by providing dense vector representations that capture word meaning. Additionally, the use of K-Fold cross-validation ensures robust evaluation, reducing the risk of overfitting and improving generalization. This novel hybrid architecture represents a significant advancement in the field of web content classification.

The practical significance of the proposed model is evident in its ability to address a critical issue in web content classification—accurately identifying and categorizing various types of web content. With its high accuracy and efficiency, the hybrid CNN-LSTM-Attention model can be directly applied to real-world text-based classification tasks, including sorting relevant and irrelevant web pages, classifying content into predefined categories, and improving content moderation. The model's robust generalization ensures that it can effectively adapt to new content types and domain-specific text without being overfitted to specific data partitions. Additionally, the integration of the Attention mechanism allows the model to focus on the most critical parts of the text, improving interpretability and making the classification process more transparent. These features make the model suitable for use in real-time content classification systems.

Prospects for further research are focused on optimizing the model for better computational efficiency, especially in resource-constrained environments, expanding its capabilities to handle multimodal data inputs like images and videos, adapting the model for cross-language and cross-cultural content classification, and investigating its resilience against adversarial attacks while integrating it with other content-based systems to enhance its robustness and effectiveness.

REFERENCES

1. Evans K., Abuadbba A., Wu T., Moore K., Ahmed M., Pogrebna G., Nepal S., Johnstone M. RAIDER: Reinforcement-aided Spear Phishing Detector, *Network and System Security*, 2022, Vol. 13787, pp. 23–59. DOI: https://doi.org/10.1007/978-3-031-23020-2_2.
2. Kozlenko M., Lazarovych I., Tkachuk V., Kuz M. Deep Learning Demodulation of Amplitude Noise Shift Keying Spread Spectrum Signals, *IEEE International Conference on Problems of Infocommunications. Science and Technology (PIC S&T)*. Kharkiv, Ukraine, 2020, pp. 717–720. DOI: <https://doi.org/10.1109/PICST51311.2020.9468063>.
3. Dhar A., Mukherjee H., Roy K., Santosh K.C., Dash N.S. Hybrid approach for text categorization: A case study with Bangla news article, *Journal of Information Science*, 2023, Vol. 49(3), pp. 762–777. DOI: <https://doi.org/10.1177/01655515211027770>.
4. Vörös T., Bergeron S.P., Berlin K. Web Content Filtering through knowledge distillation of Large Language Models [Electronic resource]. Access mode: <https://doi.org/10.48550/arXiv.2305.05027>.
5. Verma P., Goyal A., Gigras Y. Email phishing: Text classification using natural language processing, *Computer Science and Information Technologies*, 2020, Vol. 1(1), pp. 1–12. DOI: <https://doi.org/10.11591/csit.v1i1.p1-12>.
6. Somesha M., Alwyn R. Pais. Classification of Phishing Email Using Word Embedding and Machine Learning Techniques, *Journal of Cyber Security and Mobility*, 2022, Vol. 11(3), pp. 279–320. DOI: <https://doi.org/10.13052/jcsm2245-1439.1131>.
7. Asudani D. S., Nagwani N. K. & Singh P. Impact of word embedding models on text analytics in deep learning environment: a review, *Artificial Intelligence Review*, 2023 Vol. 56, pp. 10345–10425. DOI: <https://doi.org/10.1007/s10462-023-10419-1>.
8. Petridis C. Text Classification: Neural Networks VS Machine Learning Models VS Pre-trained Models [Electronic resource]. Access mode: <https://doi.org/10.48550/arXiv.2412.21022xiv.org/abs/2412.21022>.
9. Gurumurthy M., Chitra K. Email Phishing Detection Model using CNN Model, *Journal of Innovation and Technology*, 2024, Vol. 2024(43), pp. 1–8. DOI: <https://doi.org/10.61453/joit.v2024no43>.
10. Kozlenko M., Lazarovych I., Tkachuk V., Vialkova V. Software Demodulation of Weak Radio Signals Using Convolutional Neural Network, *IEEE 7th International Conference on Energy Smart Systems (ESS)*. Kyiv, Ukraine, 2020, pp. 339–342. DOI: <https://doi.org/10.1109/ESS50319.2020.9160035>.
11. Benavides-Astudillo E., Fuertes W., Sanchez-Gordon S., Nuñez-Agurto D., Rodríguez-Galán G. A Phishing-Attack-Detection Model Using Natural Language Processing and Deep Learning, *Applied Sciences*, 2023, Vol. 13(9), P. 5275. DOI: <https://doi.org/10.3390/app13095275>.
12. Maurer M.E. Phishload [Electronic resource]. Access mode: <https://www.medien.ifi.lmu.de/team/max.maurer/files/phishload>.
13. Tawil A.A., Almazaydeh L., Qawasmeh D., Qawasmeh B., Alshinwan M., Elleithy K. Comparative Analysis of Machine Learning Algorithms for Email Phishing Detection Using TF-IDF, Word2Vec, and BERT, *Computers, Materials & Continua*, 2024, Vol. 81(2), pp. 3395–3412. DOI: <https://doi.org/10.32604/cmc.2024.057279>.
14. Camacho-Collados J., Pilehvar M.T. On the Role of Text Pre-processing in Neural Network Architectures: An Evaluation Study on Text Categorization and Sentiment Analysis [Electronic resource]. Access mode: <https://doi.org/10.48550/arXiv.1707.01780>.
15. Joshi A., Lloyd L., Westin P., Seethapathy S. Using Lexical Features for Malicious URL Detection – A Machine Learning Approach [Electronic resource]. Access mode: <https://doi.org/10.48550/arXiv.1910.06277org/abs/1910.06277>.
16. Giorgi J., Nitski O., Wang B., Bader G. DeCLUTR: Deep Contrastive Learning for Unsupervised Textual Representations [Electronic resource]. Access mode: <https://doi.org/10.48550/arXiv.2006.03659>.
17. Priya R., Sujatha S. W2V and Glove Embedding based Sentiment Analysis of Text Messages, *International Journal of Advanced Research in Science, Communication and Technology*, 2022, Vol. 2(1), pp. 229–234. DOI: <https://doi.org/10.48175/IJARSCT-7677>.

18. Li S., Chen L., Song C., Liu X. Text Classification Based on Knowledge Graphs and Improved Attention Mechanism [Electronic resource]. Access mode: <https://doi.org/10.48550/arXiv.2401.03591>.
19. Metzner C., Gao S., Hermannova D., Lima-Walton E. Attention Mechanisms in Clinical Text Classification: A Comparative Evaluation, *IEEE Journal of Biomedical and Health Informatics*, 2024, Vol. 28(4), pp. 2247–2258. DOI: <https://doi.org/10.1109/JBHI.2024.3355951>.
20. Vallebuena A., Handan-Nader C., Manning C. D., Ho D. E. Statistical Uncertainty in Word Embeddings: GloVe-V [Electronic resource]. Access mode: <https://doi.org/10.48550/arXiv.2406.12165>.
21. Jarrahi A., Mousa R., Safari L. SLCNN: Sentence-Level Convolutional Neural Network for Text Classification [Electronic resource]. Access mode: <https://doi.org/10.48550/arXiv.2301.11696>.
22. Zhang R., Lee H., Radev D.R. Dependency Sensitive Convolutional Neural Networks for Modeling Sentences and Documents [Electronic resource]. Access mode: <https://doi.org/10.18653/v1/N16-1177>.
23. Kang L., He S., Long F., Wang M. Bilingual attention based neural machine translation, *Applied Intelligence*, 2022, Vol. 53(4), pp. 4302–4315. DOI: <https://doi.org/10.1007/s10489-022-03556>.
24. Terven J., Cordova-Esparza D. M., Ramirez-Pedraza A., Chavez-Urbiola E.A. Loss Functions and Metrics in Deep Learning [Electronic resource]. Access mode: <https://doi.org/10.48550/arXiv.2307.02694>.
25. Kozlenko M., Vialkova V. Software Defined Demodulation of Multiple Frequency Shift Keying with Dense Neural Network for Weak Signal Communications, *IEEE 15th International Conference on Advanced Trends in Radioelectronics, Telecommunications and Computer Engineering (TCSET)*. Lviv-Slavske, Ukraine, 2020, pp. 590–595. DOI: <https://doi.org/10.1109/TCSET49122.2020.235501>.
26. Dutschmann T.M., Kinzel L., ter Laak A., et al. Large-scale evaluation of k -fold cross-validation ensembles for uncertainty estimation, *Journal of Cheminformatics*, 2023, Vol. 15(1), P. 49. DOI: <https://doi.org/10.1186/s13321-023-00709-9>.

Received 20.05.2025.

Accepted 06.10.2025.

УДК 004.056.5

ДОСЛІДЖЕННЯ ГІБРИДНОЇ МОДЕЛІ LSTM-CNN-ATTENTION ДЛЯ КЛАСИФІКАЦІЇ ВЕБ-КОНТЕНТУ НА ОСНОВІ ТЕКСТУ

Кузь М. В. – д-р техн. наук, професор кафедри інформаційних технологій, Прикарпатський національний університет імені В. Стефаника, Україна.

Лазарович І. М. – канд. техн. наук, доцент кафедри інформаційних технологій, Прикарпатський національний університет імені В. Стефаника, Україна.

Козленко М. І. – канд. техн. наук, доцент кафедри інформаційних технологій, Прикарпатський національний університет імені В. Стефаника, Україна.

Пікуляк М. В. – канд. техн. наук, доцент кафедри інформаційних технологій, Прикарпатський національний університет імені В. Стефаника, Україна.

Кваснюк А. Д. – магістр, Прикарпатський національний університет імені В. Стефаника, Україна.

АНОТАЦІЯ

Актуальність. Класифікація веб-контенту на основі тексту відіграє ключову роль у різних завданнях обробки природної мови (NLP), включаючи виявлення фейкових новин, фільтрацію спаму, категоризацію контенту та автоматизовану модерацию. Оскільки обсяг і складність текстових даних в Інтернеті продовжують зростати, традиційні підходи до класифікації – особливо ті, що спираються на ручне створення ознак або поверхневі методи навчання – мають труднощі в уловлюванні тонких семантичних зв'язків і структурної мінливості сучасного веб-контенту. Ці обмеження призводять до зниження адаптивності та поганої здатності до узагальнення на реальних даних. Тому існує чітка потреба в удосконаленні моделей, які можуть одночасно навчатися локальним мовним патернам і розуміти ширший контекстуальний зміст веб-тексту.

Мета роботи – підвищення точності та узагальнювальних властивостей моделей класифікації веб-контенту на основі тексту шляхом використання передових технік глибокого навчання. Завданням є покращення витягування локальних та глобальних ознак тексту та навчання послідовностей, що дозволить створити більш ефективну та точну модель для класифікації веб-сторінок з урахуванням їх змісту та контексту.

Метод. Це дослідження представляє гібридну архітектуру глибокого навчання, яка інтегрує мережі Long Short-Term Memory (LSTM), згорткові нейронні мережі (CNN) та механізм уваги для покращення класифікації веб-контенту на основі тексту. Для подання слів використовуються попередньо навчений вектор GloVe, який зберігає семантичну подібність. Згорткова мережа (CNN) видобуває локальні патерни n -грам і лексичні ознаки, в той час як LSTM моделює довготривалі залежності та послідовну структуру. Інтегрований механізм уваги дозволяє моделі вибірково фокусуватися на найважливіших частинах вхідної послідовності. Модель була оцінена за допомогою датасету, що складається з понад 10 000 веб-сторінок на основі HTML, позначених як легітимні або фейкові. Для оцінки стійкості та узагальненості запропонованого рішення використовувалася 5-кратна крос-валідація.

Результати. Експериментальні результати показують, що гібридна модель LSTM-CNN-Attention досягла відмінних результатів, з точністю 0,98, точністю (precision) 0,94, відзивом (recall) 0,92 і F1-мірою 0,93. Ці результати перевершують ефективність базових моделей, що спираються лише на CNN, LSTM або трансформерні класифікатори, такі як BERT. Поєднання компонентів нейронних мереж дозволило моделі ефективно захоплювати як дрібні текстові структури, так і ширший семантичний контекст. Крім того, використання векторів GloVe надало ефективне та дієве подання текстових даних, роблячи модель придатною для інтеграції в системи з вимогами до реального часу або майже реального часу.

Висновки. Запропонована гібридна архітектура демонструє високу ефективність у класифікації веб-контенту на основі тексту, особливо в завданнях, що вимагають одночасного видобутку синтаксичних ознак та семантичної інтерпретації. Поєднуючи згорткові, рекурентні та засновані на увазі механізми, модель долає обмеження окремих архітектур і досягає покращеного узагальнення. Ці висновки підтримують більш широке використання гібридних підходів глибокого навчання в додатках NLP, особливо там, де потрібно обробляти та класифікувати складні, неструктуровані текстові дані з високою надійністю.

КЛЮЧОВІ СЛОВА: класифікація веб-контенту, LSTM-CNN-Attention, глибоке навчання, обробка природної мови, вектори GloVe, класифікація тексту, гібридна модель, моделювання послідовностей.

ЛІТЕРАТУРА

1. Evans K. RAIDER: Reinforcement-aided Spear Phishing Detector / [K. Evans, A. Abuadba, T. Wu et al.] // *Network and System Security*. – 2022. – Vol. 13787. – P. 23–59. DOI: https://doi.org/10.1007/978-3-031-23020-2_2.
2. Deep Learning Demodulation of Amplitude Noise Shift Keying Spread Spectrum Signals / [M. Kozlenko, I. Lazarovych, V. Tkachuk, M. Kuz] // *IEEE International Conference on Problems of Infocommunications. Science and Technology (PIC S&T)*. – Kharkiv, Ukraine, 2020. – P. 717–720. DOI: <https://doi.org/10.1109/PICST51311.2020.9468063>.
3. Hybrid approach for text categorization: A case study with Bangla news article / [A. Dhar, H. Mukherjee, K. Roy et al.] // *Journal of Information Science*. – 2023. – Vol. 49(3). – P. 762–777. DOI: <https://doi.org/10.1177/01655515211027770>.
4. Vörös T. Web Content Filtering through knowledge distillation of Large Language Models [Електронний ресурс] / T. Vörös, S. P. Bergeron, K. Berlin. – Режим доступу: <https://doi.org/10.48550/arXiv.2305.05027>.
5. Verma P. Email phishing: Text classification using natural language processing / P. Verma, A. Goyal, Y. Gigras // *Computer Science and Information Technologies*. – 2020. – Vol. 1(1). – P. 1–12. DOI: <https://doi.org/10.11591/csit.v1i1.p1-12>.
6. Somesha M. Classification of Phishing Email Using Word Embedding and Machine Learning Techniques / M. Somesha, R. Alwyn Pais // *Journal of Cyber Security and Mobility*. – 2022. – Vol. 11(3). – P. 279–320. DOI: <https://doi.org/10.13052/jcsm2245-1439.1131>.
7. Asudani D. S. Impact of word embedding models on text analytics in deep learning environment: a review / D. S. Asudani, N. K. Nagwani, P. Singh // *Artificial Intelligence Review*. – 2023. – Vol. 56. – P. 10345–10425. DOI: <https://doi.org/10.1007/s10462-023-10419-1>.
8. Petridis C. Text Classification: Neural Networks VS Machine Learning Models VS Pre-trained Models [Електронний ресурс] / C. Petridis. – Режим доступу: <https://doi.org/10.48550/arXiv.2412.21022>.
9. Gurumurthy M. Email Phishing Detection Model using CNN Model / M. Gurumurthy, K. Chitra // *Journal of Innovation and Technology*. – 2024. – Vol. 2024(43). – P. 1–8. DOI: <https://doi.org/10.61453/joit.v2024no43>.
10. Software Demodulation of Weak Radio Signals Using Convolutional Neural Network / [M. Kozlenko, I. Lazarovych, V. Tkachuk, V. Vialkova] // *IEEE 7th International Conference on Energy Smart Systems (ESS)*. – Kyiv, Ukraine, 2020. – P. 339–342. DOI: <https://doi.org/10.1109/ESS50319.2020.9160035>.
11. A Phishing-Attack-Detection Model Using Natural Language Processing and Deep Learning / [E. Benavides-Astudillo, W. Fuertes, S. Sanchez-Gordon et al.] // *Applied Sciences*. – 2023. – Vol. 13(9). – P. 5275. DOI: <https://doi.org/10.3390/app13095275>.
12. Maurer M. E. Phishload [Електронний ресурс] / M. E. Maurer. – Режим доступу: <https://www.medien.fim.uni.de/team/max.maurer/files/phishload>.
13. Comparative Analysis of Machine Learning Algorithms for Email Phishing Detection Using TF-IDF, Word2Vec, and BERT / [A. A. Tawil, L. Almazaydeh, D. Qawasmeh et al.] // *Computers, Materials & Continua*. – 2024. – Vol. 81(2). – P. 3395–3412. DOI: <https://doi.org/10.32604/cmc.2024.057279>.
14. Camacho-Collados J. On the Role of Text Preprocessing in Neural Network Architectures: An Evaluation Study on Text Categorization and Sentiment Analysis [Електронний ресурс] / J. Camacho-Collados, M. T. Pilehvar. – Режим доступу: <https://doi.org/10.48550/arXiv.1707.01780>.
15. Joshi A. Using Lexical Features for Malicious URL Detection – A Machine Learning Approach [Електронний ресурс] / A. Joshi, L. Lloyd, P. Westin, S. Seethapathy. – Режим доступу: <https://doi.org/10.48550/arXiv.1910.06277>.
16. DeCLUTR: Deep Contrastive Learning for Unsupervised Textual Representations [Електронний ресурс] / [J. Giorgi, O. Nit-ski, B. Wang, G. Bader]. – Режим доступу: <https://doi.org/10.48550/arXiv.2006.03659>.
17. Priya R. W2V and Glove Embedding based Sentiment Analysis of Text Messages / R. Priya, S. Sujatha // *International Journal of Advanced Research in Science, Communication and Technology*. – 2022. – Vol. 2(1). – P. 229–234. DOI: <https://doi.org/10.48175/IJARSC-7677>.
18. Text Classification Based on Knowledge Graphs and Improved Attention Mechanism [Електронний ресурс] / [S. Li, L. Chen, C. Song, X. Liu]. – Режим доступу: <https://doi.org/10.48550/arXiv.2401.03591>.
19. Attention Mechanisms in Clinical Text Classification: A Comparative Evaluation / [C. Metzner, S. Gao, D. Herrmannova, E. Lima-Walton] // *IEEE Journal of Biomedical and Health Informatics*. – 2024. – Vol. 28(4). – P. 2247–2258. DOI: <https://doi.org/10.1109/JBHI.2024.3355951>.
20. Statistical Uncertainty in Word Embeddings: GloVe-V [Електронний ресурс] / [A. Vallebuena, C. Handan-Nader, C. D. Manning, D. E. Ho]. – Режим доступу: <https://doi.org/10.48550/arXiv.2406.12165>.
21. Jarrahi A. SLCNN: Sentence-Level Convolutional Neural Network for Text Classification [Електронний ресурс] / A. Jarrahi, R. Mousa, L. Safari. – Режим доступу: <https://doi.org/10.48550/arXiv.2301.11696>.
22. Zhang R. Dependency Sensitive Convolutional Neural Networks for Modeling Sentences and Documents [Електронний ресурс] / R. Zhang, H. Lee, D. R. Radev. – Режим доступу: <https://doi.org/10.18653/v1/N16-1177>.
23. Kang L. Bilingual attention based neural machine translation / L. Kang, S. He, F. Long, M. Wang // *Applied Intelligence*. – 2022. – Vol. 53(4). – P. 4302–4315. DOI: <https://doi.org/10.1007/s10489-022-0356>.
24. Loss Functions and Metrics in Deep Learning [Електронний ресурс] / [J. Terven, D. M. Cordova-Esparza, A. Ramirez-Pedraza, E. A. Chavez-Urbiola]. – Режим доступу: <https://doi.org/10.48550/arXiv.2307.02694>.
25. Kozlenko M. Software Defined Demodulation of Multiple Frequency Shift Keying with Dense Neural Network for Weak Signal Communications / M. Kozlenko, V. Vialkova // *IEEE 15th International Conference on Advanced Trends in Radioelectronics, Telecommunications and Computer Engineering (TCSET)*. – Lviv-Slavske, Ukraine, 2020. – P. 590–595. DOI: <https://doi.org/10.1109/TCSET49122.2020.235501>.
26. Large-scale evaluation of k-fold cross-validation ensembles for uncertainty estimation / [T. M. Dutschmann, L. Kinzel, A. ter Laak, et al.] // *Journal of Cheminformatics*. – 2023. – Vol. 15(1). – P. 49. DOI: <https://doi.org/10.1186/s13321-023-00709-9>.

CLASSIFICATION OF HISTOLOGICAL IMAGES BASED ON CONVOLUTIONAL NEURAL NETWORKS

Liashchynskyi P. B. – Post-graduate student of the Department of Computer Engineering, West Ukrainian National University, Ternopil, Ukraine.

ABSTRACT

Context. Automated classification of histological images is of great importance for speeding up and improving the accuracy of diagnostics in medicine. Taking into account the complexity and high variability of histological structures, the use of deep learning and convolutional neural networks, in particular, is a promising direction in solving this problem.

The object of study is the process of classifying histological images using convolutional neural networks to determine and optimize the architecture with the highest accuracy rate.

Objective. The aim of this work is to develop an effective approach to histological image classification using convolutional neural networks that provides high accuracy through step-by-step optimization of the architecture and application of data expansion methods using affine transformations and synthesis based on diffusion models.

Method. The study includes four main stages. The first stage involves a comparative analysis of 12 well-known convolutional neural network architectures on a basic histological dataset. The second and third stages involve comparative analysis on extended data, including affine transformations and synthetic images generated by the diffusion model, respectively. The final, fourth stage involves a neuroevolutionary search for the optimal architectural cell. Once it is found, it is integrated into the model architecture, where for each layer a choice is made between certain blocks and the found cell. This approach allows to automatically form the optimal sequence of blocks in the model, which ensures the highest classification accuracy.

Results. The proposed approach improved the accuracy of histological image classification compared to the initial architectures. The addition of synthetic images to the training set provided an increase in model performance. The search for the optimal cell and its integration into the model with further optimization demonstrated an additional improvement in classification quality, increasing the accuracy to 94.9, 96.1, and 99.8% on each dataset, respectively.

Conclusions. The proposed approach allows achieving high accuracy of histological image classification through a step-by-step process that includes the use of classical convolutional neural network architectures, generation of synthetic data, and search for optimal architectural and hyperparametric configurations.

A software module for classifying histological images has been developed that can be used in an automatic diagnostic system.

KEYWORDS: convolutional neural networks, classification, optimization, histology.

ABBREVIATIONS

AI is an artificial intelligence;
CAD is a computer aided diagnosis;
CNN is a convolutional neural network;
CPU is a central processing unit;
CUDA is a compute unified device architecture;
DCNN is a deep convolutional neural network;
FTU is a functional tissue unit.
GB is a gigabyte;
GPU is a graphics processing unit;
LSTM is a long short-term memory;
NAS is a neural architecture search;
NNI is a neural network intelligence;
RAM is a random-access memory;
ROC is a receiver-operating characteristic;
UML is an unified modeling language;
VGG is a visual geometry group.

NOMENCLATURE

Acc is an accuracy metric;
B is a set of available blocks;
b_i is an available *i*-th block for neural network architecture;
C is a number of color channels;
Cell is a found optimized cell;
D is a dataset of histological images;
D is a number of nodes in the cell;

FN is a false negative, represents the number of FTU pixels being misclassified as background;

FN is a false positive, represents the number background pixels being misclassified as FTUs (due to misalignment);

H is a height of the image in pixels;

M is a model of neural network;

*M** is a model of neural network which showed the highest quality of classification;

M⁺ is a developed and optimized model of neural network;

N is a number of images in the dataset;

n is a number of output classes;

O is a set of available operations;

p is a number of known CNN models;

R is a real numbers;

TN is a true negative, represents the number of background pixels that have been properly classified as background;

TP is a true positive, represents the number of FTU pixels that have been properly classified as FTU;

v is a number of layers in CNN;

W is a width of the image in pixels;

w_i is a width of the *i*-th block;

x_i is a *i*-th histology image;

Y is a set of all possible classes labels;

y_j is a *j*-th class label of histological image.

INTRODUCTION

Computer image recognition plays an important role in modern medicine. This process is key in diagnostics, including automated analysis of biomedical images.

Due to the rapid development of deep learning methods, it is now possible to process complex visual data contained in medical images of various types.

One of them is histological images, which are images of tissues obtained by microscopy after preliminary staining of samples.

These images are widely used in cancer diagnostics and help to assess the morphological structure of tissues and detect pathological changes [1–5]. However, the accuracy of the diagnosis in manual analysis largely depends on the experience of the specialist. This creates prerequisites for the implementation of automated solutions using neural networks.

Significant progress in the field of AI has led to the active implementation of CNNs in biomedical image analysis [6–12]. However, existing approaches often do not take into account the structural specificity of different tissue types, which reduces the accuracy of models in clinical settings. In addition, the impact of different convolutional neural network architectures on classification accuracy in the context of a limited number of annotated images is not sufficiently studied.

Thus, conducting experiments using well-known CNN architectures for histological image classification and further developing an optimized architecture to improve accuracy is a relevant and important task.

The object of study is the processes classification of histological images based on convolutional neural networks.

The subject of study are models, and tools for classifying histological images based on convolutional neural networks.

The purpose of the work is to improve the classification accuracy by optimizing the neural network architecture.

1 PROBLEM STATEMENT

Suppose we have a dataset of histological images consisting of N examples :

$$D = \{(x_i, y_j)\}_{i=1, j=1}^{N,3}$$

Each image has a format: $x_i \in R^{H \times W \times C}$, with $y_j \in Y$ and $Y = \{1, 2, 3\}$.

We need to build a classification model that can be described as follows:

$$M : R^{H \times W \times C} \rightarrow Y.$$

That is, the model takes an image as an input and returns the label of the class to which this image belongs.

The problem of finding M^* can be solved by conducting experiments on a set of known CNN architectures:

$$\{M_1, M_2, \dots, M_p\}.$$

The accuracy metric is described by the formula:

$$Acc = \frac{TP + TN}{TP + TN + FN + FP}.$$

The problem of finding an optimized version of M^+ can be solved in two steps: finding the optimal cell and building the macroarchitecture.

The process of finding a cell is to find the set of operations and the number of nodes that maximize the accuracy on the validation set:

$$Cell = \arg \max_d Acc(O, D).$$

Having found the optimal cell, we form the architecture at the macro level as a sequence of blocks. That is, the architecture is defined as an ordered sequence of $B = [b_1, b_2, \dots, b_n]$.

The task is to find the optimal sequence of blocks and their number that maximizes the accuracy on the validation dataset:

$$M^+ = \arg \max_v Acc(Cell, D).$$

2 REVIEW OF THE LITERATURE

Image classification using convolutional neural networks is a fairly popular method. This technology has become especially popular in medical diagnostics.

Paper [13] describes the classification of benign and malignant subtypes of histopathological visualization of breast cancer using hybrid transfer learning based on CNN-LSTM.

The authors of [14] describe the analysis of histological images using hybrid methods for the early detection of multiclass breast cancer based on Fusion CNN and Handcrafted functions.

In [15], the authors propose a method to increase the robustness of CNNs to scanner and image variability (brightness, color, etc.) based on immunohistochemical images for epithelial segmentation in cervical histology.

The authors of [16] propose a new type of CNN based on a statistical filter. The paper describes the classification of histological tissue.

Paper [17] describes a distance regression and classification network for segmenting nuclei instances and classifying the type in histological images.

In [18], the authors propose a new classification method based on an extended decoding network based on the traditional U-shaped structure. The proposed method demonstrates good accuracy, reaching 0.92.

Paper [19] describes the use of CNNs for celiac disease classification. The authors also describe the use of

the same model for diagnosing carcinoma and Crohn's disease. The network showed good results in classifying 5 different diagnoses.

The authors of [20] conducted a detailed comparative analysis of different CNN models (VGG-16, AlexNet, ResNet-50) for classifying histological images to detect breast cancer. The BreakHis dataset was used for the test.

Paper [21] describes the use of CNNs with VGG19 architecture for the detection of oral squamous cell carcinoma.

The authors of [22] describe a trial model based on deep learning for four-class histological classification of colon tumors based on narrow band imaging.

Paper [23] describes colorectal cancer detection based on convolutional neural networks and a ranking algorithm.

The authors of [24] propose the FabNet CNN model for classifying histological images by accumulating layers closer to each other for further fusion of semantic and spatial features.

In [25], the authors used hybrid optimization using a deep convolutional neural network (DCNN) to identify and classify breast cancer types into different subclasses.

The authors of [26] describe the improvement of breast cancer grade classification using an integrated Swin Transformer.

Paper [27] describes CVITS-Net, a new CNN model for classifying histological images. The model achieves an accuracy of 96.06%.

Paper [28] describes the classification of histopathological images of lung cancer using transfer learning based on the EfficientNetB7 convolutional neural network model.

The authors of [29] improved the detection and classification of breast cancer by adjusting hyperparameters.

In [30], the authors describe the process of classifying histopathological images of breast cancer using CNNs and the VGG-19 model.

The authors of [31] propose a CNN architecture for classifying histopathological images of kidneys. The authors tested and compared the proposed architecture with seven state-of-the-art architectures.

Study [32] proposed a CNN-based classification of histopathological images of breast cancer. The authors used transfer learning to build four pre-trained models. Out of the four pre-trained models, DensNet201 achieved an accuracy of 91.37% and a sensitivity level of 100% at 200x magnification.

3 MATERIALS AND METHODS

A convolutional neural network is one of the most effective deep learning tools for image processing. Its architecture is designed to automatically detect characteristic features in input images using a cascade of specialized layers.

A CNN consists of input and output layers, and also has several hidden layers. CNNs usually consist of convo-

lutional (responsible for detecting local patterns such as edges, textures, or more complex morphological features of tissues), subsampling (reduce the size of features while retaining the most important information, which helps to reduce computational complexity), normalization layers (stabilize the learning process and speed up model convergence), and fully connected layers (perform the final interpretation of features and classification based on the generalized information from the previous layers).

The general architecture of a convolutional neural network is shown in Fig. 1. The illustration shows the sequential passage of an image through the layers of the model – from the initial filtering to the final class prediction.

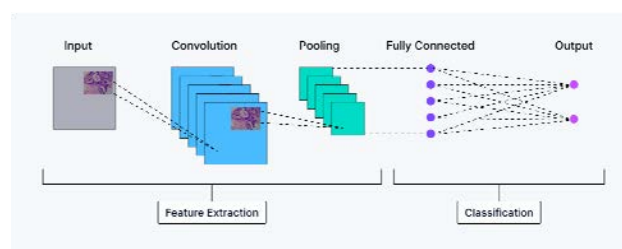


Figure 1 – General CNN architecture

Fig. 2 shows the general process of automatic image classification using neural networks.

This diagram shows the full cycle of an artificial intelligence system for the task of classifying histological images, illustrating the effectiveness of CNNs in medical computer vision tasks.

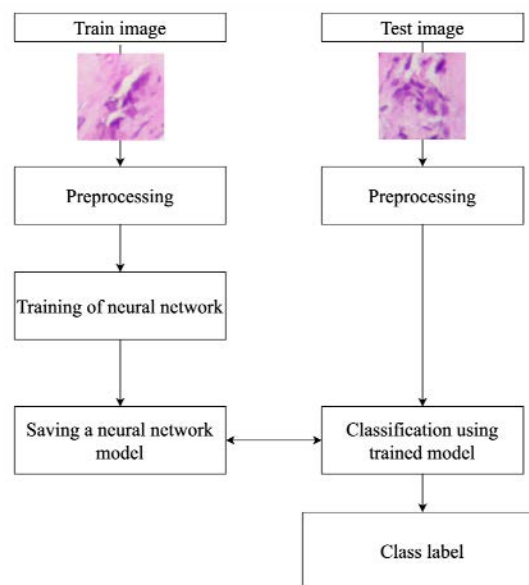


Figure 2 – General scheme of the image classification process using neural networks

As you can see from Fig. 2, the train image needs to be pre-processed (noise reduction or brightness equalization, resizing, etc.). The neural network is trained on images, and then its model is saved. As a result of

training, the model is able to classify images. To do this, we submit the Test Image as an input, pre-process it, and apply the already trained model for classification, after which we get the result – the label of the class to which the image belongs.

This process needs to be repeated several times until we go through all the following architectures that have

been selected for the experiments: AlexNet, VGG-16, ResNet 18, Wide ResNet, ResNeXt, GoogLeNet, Inception V3, EfficientNet, EfficientNetV2, MobileNet V2, MobileNet V3, ConvNeXt.

The general scheme of the study and the process of finding an optimized architecture is shown in Fig. 3.

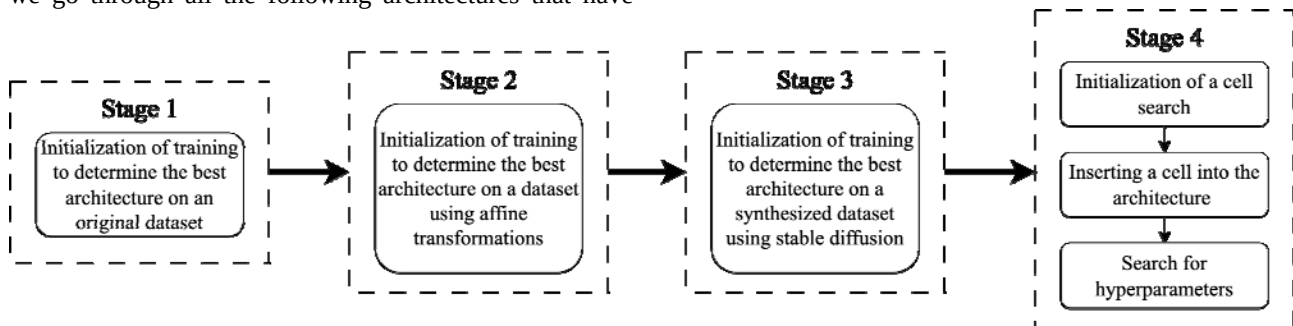


Figure 3 – General scheme of the process of research and search for an optimized architecture

As shown in Fig. 3, the experimental methodology is divided into several sequential stages (Stage 1–4), each of which performs an important function in the research process.

Stage 1 involves initial testing of various neural network architectures on the original histological image dataset.

The purpose of this stage is to preliminarily evaluate the performance of the models without additional transformations of the input data. This allows you to determine the baseline level of accuracy against which the results of the following stages can be compared.

In Stage 2 and Stage 3, similar experiments are performed, but on extended versions of the dataset. At these stages, affine transformations are applied to the images:

- affine distortions (scaling, rotation, mirroring), that increase the variability of the sample and allow the model to better generalize information;
- diffusion models – modern generative methods used to synthesize additional samples to expand the data.

Stage 4 focuses on architecture search and optimization. First, Neural Architecture Search is performed to find the optimal structure of the computing cell (for this purpose, the method developed and proposed by the authors in [33] was used), which is then integrated into the overall model as a building block. After that, the hyperparametric optimization of the updated architecture is performed, which allows to achieve maximum classification accuracy by adjusting the parameters.

To find the optimal network architecture, we used the regularized evolution algorithm, which is a variant of the evolutionary algorithm used to find the best neural network architecture in the model space. This method, which was presented in [34], has the following features:

1) population – a fixed number of architectures are stored;

2) parent selection – the best architecture is selected from a random subset of the population based on performance;

3) mutation – a random change is applied to the selected architecture (e.g., changing the layer, number of filters, etc.);

4) replacement – a new architecture is added to the population, and the oldest architecture is removed (this step is called “regularization” because it limits the accumulation of old solutions).

Evolutionary search with regularization is an effective NAS method that balances between using the best architectures and exploring new options. The pseudocode of the algorithm is shown in Fig. 4.

```

1 archive ← ∅ // Stores all evaluated architectures
2 queue ← empty FIFO structure // Current population of models
3
4 // Initial random population generation
5 while |queue| < MAX_POPULATION do
6     individual.arch ← CREATE_RANDOM_ARCH()
7     individual.score ← TRAIN_AND_ASSESS(individual.arch)
8     add individual to end of queue
9     add individual to archive
10 end while
11
12 // Evolutionary search loop
13 while |archive| < MAX_EVALUATIONS do
14     pool ← ∅ // Candidates for parent selection
15
16     // Sample parent candidates randomly
17     while |pool| < POOL_SIZE do
18         candidate ← random element from queue
19         add candidate to pool
20     end while
21
22     // Select best-performing parent
23     parent ← element in pool with highest score
24
25     // Generate and evaluate a mutated child
26     offspring.arch ← MUTATE_ARCHITECTURE(parent.arch)
27     offspring.score ← TRAIN_AND_ASSESS(offspring.arch)
28
29     // Add offspring to population and archive
30     add offspring to end of queue
31     add offspring to archive
32
33     // Remove oldest individual from population
34     remove first element from queue
35 end while
36
37 // Return the best discovered architecture
38 return element in archive with highest score
    
```

Figure 4 – Pseudocode of the regularized evolution algorithm

To search for the optimal cell, several available operations are predefined (Fig. 5), which are randomly selected in a certain sequence until the highest accuracy is achieved.

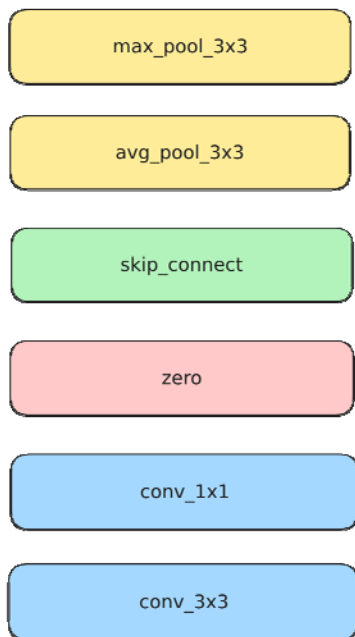


Figure 5 – Available operations pool

Let us consider each of them in more detail [35].

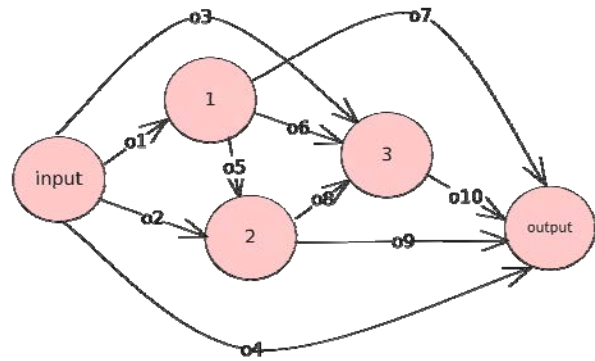
1) Pool. The pooling layer is used to reduce the dimensionality of the input data. At this stage, feature maps are reduced after convolution layers. Popular types of pooling are: average, maximum, and global. Only two types of pooling are used in the cell operations: average with a 3×3 kernel and maximum with a 3×3 kernel.

2) Skip connect. The input data is not transformed after the operation is applied. This operation can be considered as a function that accepts data and immediately returns it. This operation allows the output of a node to be directly linked to the input of another node. This operation helps to avoid gradient damping. This operation is also called the identity function.

3) Zero. The zero operation replaces the output of a node with a matrix of zeros. This operation can be used to remove a node from the cell architecture and effectively simplify the architecture. Applying this operation between two nodes means that there is no connection between them.

4) Conv. The input data is multiplied with the convolution kernel and then the result is summed. The output is a new matrix. This operation reflects the standard convolution function in neural networks.

The cell architecture is encoded according to the example shown in Fig. 6.



XXXX-XXX-XX-X = o1 o2 o3 o4-o5 o6 o7-o8 o9-o10

Figure 6 – An example of coding cell architecture

The letter X represents a digit from 0 to 5, i.e. $o = \{0, 1, 2, 3, 4, 5\}$. These digits correspond to the indices of the elements in the set of available operations (see Fig. 4). The first three digits in the code represent operations that connect the input node with nodes 1, 2, 3, and output, respectively (o1, o2, o3, o4). The next three digits encode the operations between nodes 1, 2, 3 (o5, o6) and output (o7). The next two digits encode the operations between nodes 2, 3, and output (o8, o9). The last digit encodes the operation between node 3 and output (o10).

To find the optimal network architecture, the choice is made from among the available blocks, including the optimal cell found.

The evolutionary algorithm searches for the sequence with the highest accuracy.

4 EXPERIMENTS

The experimental studies were implemented using the Python 3 programming language, which is a standard in the field of machine learning and scientific computing. PyTorch was chosen as the main library for building, training, and testing neural networks, as it is one of the most popular open source frameworks that provides flexibility in model design and a user-friendly interface for working with GPUs.

All experiments were performed in the Google Colab cloud environment, which allows you to use computing resources directly from your browser. To ensure high performance, we used an NVIDIA A100 GPU with 40 GB of memory and 83.5 GB of RAM, which allowed us to effectively train large neural networks and search for optimal architectures.

To implement the automatic search for neural network architectures, we used the NNI framework, which provides powerful tools for the experimental study of neural architectures, including support for popular NAS methods.

The software implementation is divided into two main modules: search.py and optimize.py.

The search.py script is responsible for automatically searching through the space of possible architectures, generating them, and initially evaluating them according to a given metric. It also implements the logic for

interacting with the search strategy and saving the results of each experiment.

The `optimize.py` script is used in the last step – optimizing the best architecture found. At this stage, the hyperparameters are fine-tuned, the model is retrained, and the final evaluation of its quality is performed.

The structure of the modules and the logic of building experiments is implemented using object-oriented programming. Fig. 7 shows a class diagram that illustrates the key components of the `search.py` module and their interrelationships.

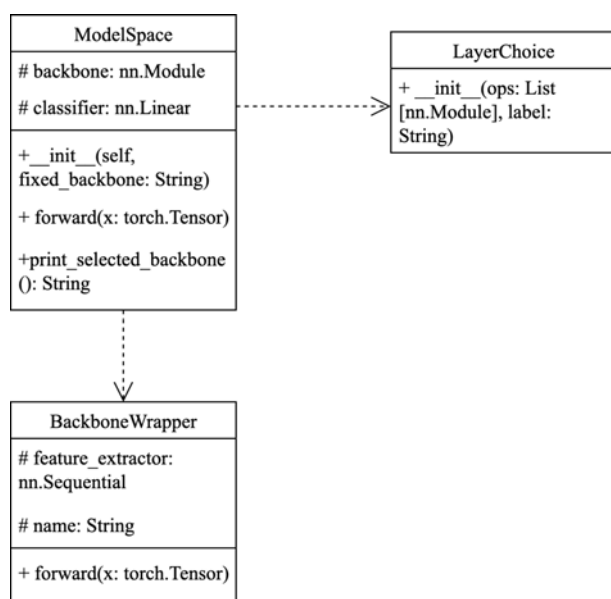


Figure 7 – UML diagram of architecture search classes

The `BackboneWrapper` class serves as a wrapper for each architecture to unify the interface. It is used to describe models from the PyTorch library.

The `ModelSpace` class describes the implementation of the model selected by `LayerChoice`. In addition, through the `fixed_backbone` parameter, it can be passed an architecture from predefined ones. This is useful when you need to test a specific model.

After performing the search and obtaining the architecture with the highest accuracy rate, we can use the same script to conduct experiments on a sample of data with affine distortion and a synthesized sample of histological images.

The dataset used in the study consists of histological images with three categories of labels: G1, G2, and G3, which correspond to different degrees of tissue differentiation. Each image has a resolution of 4096×3286 pixels, which provides high detail for analysis.

Fig. 8 illustrates examples of histological images for each of the classes, which allows you to visually assess the difference between them and emphasizes the importance of applying powerful deep learning methods for their effective classification.

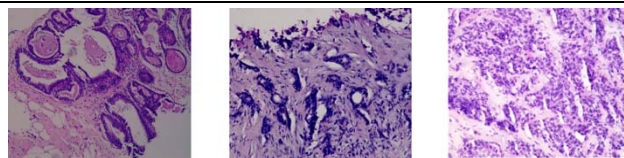


Figure 8 – Example of histological images for each class

To conduct the experiments, all images were previously resized to a uniform size of 512×512 pixels. This reduction in resolution allowed us to strike a balance between preserving enough morphological detail for analysis and reducing the computational load on the model during training.

To improve the generalizability of the model, a number of affine distortions were applied. The images were processed using the Torchvision library, which is part of the PyTorch stack. During these distortions, random horizontal or vertical mirroring, random rotation by 90 or 180 degrees, random crop with scaling, rotation without edge artifacts, and image shift without edge artifacts were applied to the images.

Each image underwent the described transformations to create two augmented copies, which were stored in the corresponding directory along with the original. Thus, for each image, a total of three variants are created (1 original and 2 fragmented).

To conduct computer experiments and generate images, we used the Stable Diffusion Web UI from Jarvis Labs with an A6000 Ampere GPU (CUDA 12.3) with 48 GB of video memory, 32 GB of RAM, and 7 CPUs.

This configuration ensures high performance for AI-generated images, allowing to effectively use the capabilities of the Stable Diffusion model to generate high-quality results.

The model `v1-5-pruned-emaonly.safetensors` was used for training and generation.

The diffusion models were trained for 33 epochs with the number of steps being 10000. The total training time was 45 minutes. One step took 0.27 seconds and one epoch took 81 seconds.

Images were generated on 5 prints saved at steps 2000, 4000, 6000, 8000, and 10000. Each print generated 1000 images, for a total of 5000 images.

The generation time was 2 hours 37 minutes and 41 seconds for G1, 2 hours 19 minutes and 58 seconds for G2, and 2 hours 38 minutes and 12 seconds for G3.

Fig. 9 shows examples of synthesized images generated by the diffusion models.

In the Stage 1, Stage 2, and Stage 3, each model was trained for 25 epochs, and the number of possible architectures was 12. The batch size for each architecture was the same, with 64 images.

For the Stage 4, the number of optimization cycles was 20, the population size was 30, and the number of epochs was 25.



Figure 9 – Example of synthesized histological images

As a result, the architectures with the highest accuracy coefficient were identified, which made it possible to assess the impact of data preparation on model accuracy, find the optimal cell, and optimize the neural network model to improve classification accuracy.

5 RESULTS

As a result of the experiments conducted in the first stage (Stage 1), the results of which are shown in Table 1, the RegNet architecture demonstrated the highest efficiency. All the results presented are sorted in descending order of the main accuracy metric, which allows us to clearly identify the best architectures among the models considered.

Table 1 – Results of experiments on an original dataset

№	Name	Acc, %
1	RegNet	90,7
2	Inception V3	90,2
3	MobileNet V2	90
4	Wideresnet	88,3
5	ResNeXt	87,7
6	EfficientNet	84,3
7	VGG16	84,2
8	ResNet	84
9	GoogLeNet	82
10	AlexNet	81,9
11	ConvNeXt	80
12	MobileNet V3	36,7

The further experiments conducted in the second (Stage 2) and third (Stage 3) stages are detailed in Table 2–3, respectively. Similarly to the first stage, the results are also sorted in descending order for easy analysis and comparison.

Table 2 – Results of experiments on a dataset using affine transformations

№	Name	Acc, %
1	RegNet	91,1
2	Wideresnet	90,9
3	ResNeXt	90,84
4	MobileNet V2	90,8
5	Inception V3	90,7
6	EfficientNet	89,6
7	ResNet	89,1
8	GoogLeNet	87,3
9	AlexNet	86
10	ConvNeXt	84
11	VGG16	82,9
12	MobileNet V3	34,4

Table 3 – Results of experiments on a synthesized dataset

№	Name	Acc, %
1	RegNet	96,6
2	ResNeXt	94,3
3	Wideresnet	94
4	Inception V3	93,1
5	MobileNet V2	92,9
6	EfficientNet	92,5
7	ResNet	92,1
8	GoogLeNet	92
9	AlexNet	91,9
10	ConvNeXt	88,5
11	VGG16	86,5
12	MobileNet V3	38,4

Analyzing the results presented in Table 1–3, we can confidently state that the RegNet architecture consistently demonstrated the highest classification accuracy on each of the studied datasets. Accordingly, RegNet was chosen as the base model for further work on optimizing and improving the architecture.

The general structure of RegNet is shown in Fig. 10, and its detailed structure is shown in Table 4.

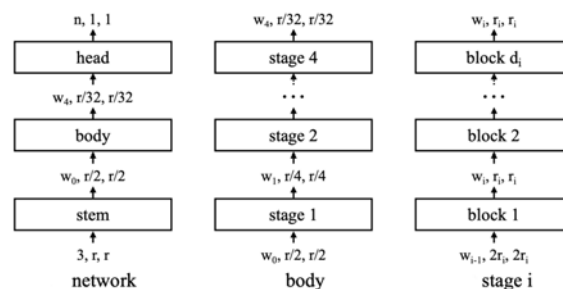


Figure 10 – General RegNet structure

Each network consists of a stem (stride-two 3×3 conv with $w_0 = 32$ output channels), followed by the network body that performs the bulk of the computation, and then a head (average pooling followed by a fully connected layer) that predicts n output classes.

The network body is composed of a sequence of stages that operate at progressively reduced resolution. Each stage consists of a sequence of identical blocks, except the first block which uses stride-two conv. While the general structure is simple, the total number of possible network configurations is vast [36].

Table 4 – Detailed RegNet structure

Layer name	Layer type	Params	Output shape
Stem	Stride-two 3×3 conv	Kernel = 3, stride = 2	32×112×112
Stage 1	Sequence of X Blocks	Stride = 2 (for the first block)	24×56×56
Stage 2		Stride = 2 (for the first block)	56×28×28
Stage 3	Sequence of X Blocks	Stride = 2 (for the first block)	152×14×14
Stage 4		Stride = 2 (for the first block)	368×7×7
Head	AvgPool + Fully Connected Layer		3
Block structure			

Table 5 shows the comparative results of classification accuracy obtained for the original RegNet architecture and optimized version developed in this study. The results are presented for different types of datasets, which allows us to assess the versatility and adaptability of the optimized model. In particular, the optimized architecture demonstrates an improvement in classification accuracy, which indicates the success of the applied improvement methods. This underscores the effectiveness of the chosen optimization approach, which allows us to improve the model's performance without significantly increasing its complexity.

Table 5 – Comparison between the original version of RegNet and the developed version

Name	Dataset	Acc, %
RegNet	Original	90,7
Developed		94,9
RegNet	Affines	91,1
Developed		96,1
RegNet	Synthesized	95,6
Developed		99,8

The architecture of the CustomCell shown in Fig. 11 consists of six nodes that are interconnected by a complex network of edges. Each edge corresponds to a specific operation that transforms the input features and is selected from a set of basic operations, including avg_pool_3x3,

skip_connect, zero, conv_1x1, conv_3x3, max_pool_3x3 in a different sequence.

The structure of the cell is organized in such a way that each node aggregates information from several previous nodes using selected operations, which provides flexibility and multi-level feature integration. Thanks to this topology, the cell can balance between preserving the original information through pass-through paths and its gradual transformation through convolution and pooling.

As a result, the resulting cell has an individual composition of operations and connections that is maximally adapted to solve a specific task of classifying histological images.

The cell architecture is coded as follows: 23141–3232–334–21–2.

The architecture of the developed neural network is shown in Fig. 12 and consists of 6 layers.

The structure of the developed neural network is shown in Table 6.

The neural network has a simple and efficient structure with an initial convolutional layer, a series of optimized cells (CustomCell), convolutional and pooling operations, and a final classifier. The network has 16 channels across the entire depth, with dimensionality reduction occurring only once (Layer 2). The main computing unit is an optimized CustomCell.

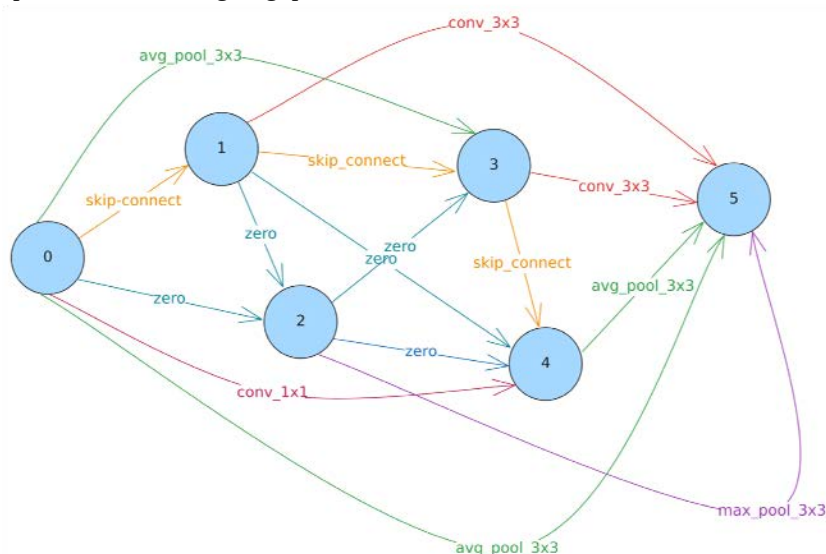


Figure 11 – Found cell architecture

The ROC curve of the optimized architecture is shown in Fig. 13. The result shows the classification accuracy on the dataset synthesized by diffusion models. The accuracy reached 99.8%.

6 DISCUSSION

As a result of the study, a new optimized architecture for histological image classification was developed, which demonstrates improved performance compared to known architectures. To quantify the performance of the models, we used the Accuracy metric, which is the key metric in classification tasks.

Tables 1–3 illustrate the results of comparisons among various known architectures, allowing to analyze their performance on test datasets.

In turn, Table 5 presents the results of the experiments conducted in the fourth stage (Stage 4), where the accuracy of the classical RegNet architecture and the developed optimized model on different types of histological datasets is compared.

In contrast to the traditional RegNet architecture, the optimized model, the details of which are shown in Fig. 12 and Table 6, is characterized by much greater flexibility. This flexibility is manifested in the ability to dynamically select types of operations, adjust the network depth, and vary its topology. This approach allows us to adapt the model to the specifics of a particular dataset, which in turn leads to an increase in classification accuracy on each of the datasets under consideration.

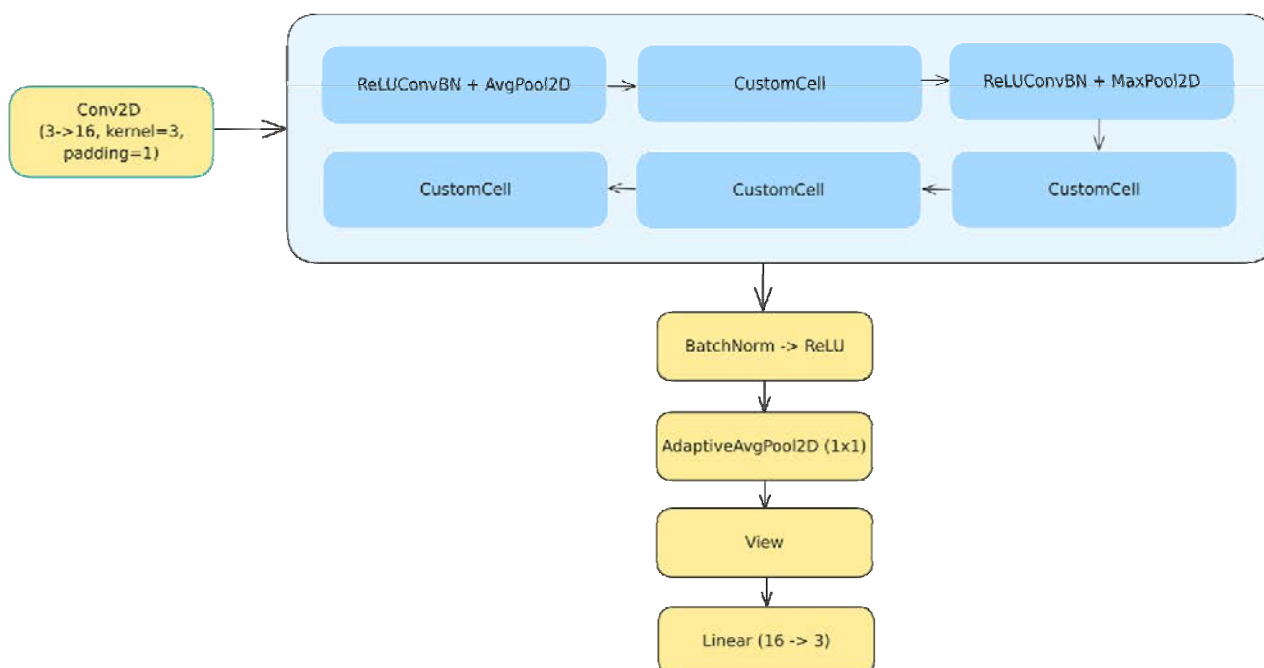


Figure 12 – Developed CNN architecture

Table 6 – Developed CNN structure

Layer name	Layer type	Params	Output shape
Stem	Conv2D	Kernel = 3, stride = 1, padding = 1	16×512×512
Layer 0	ReLUConvBN + AvgPool2D	Kernel = 3, stride = 1, padding = 1	16×512×512
Layer 1	CustomCell	Kernel = 3, stride = 1, padding = 1, Nodes = 6	16×512×512
Layer 2	ReLUConvBN + MaxPool2D	Kernel = 3, stride = 2, padding = 1	16×256×256
Layer 3	CustomCell	Kernel = 3, stride = 1, padding = 1, Nodes = 6	16×256×256
Layer 4	CustomCell	Kernel = 3, stride = 1, padding = 1, Nodes = 6	16×256×256
Layer 5	CustomCell	Kernel = 3, stride = 1, padding = 1, Nodes = 6	16×256×256
Last act	BatchNorm + ReLU		16×256×256
Global pooling	AdaptiveAvgPool2D		16×1×1
Flatten	View		16
Classifier	Linear		3

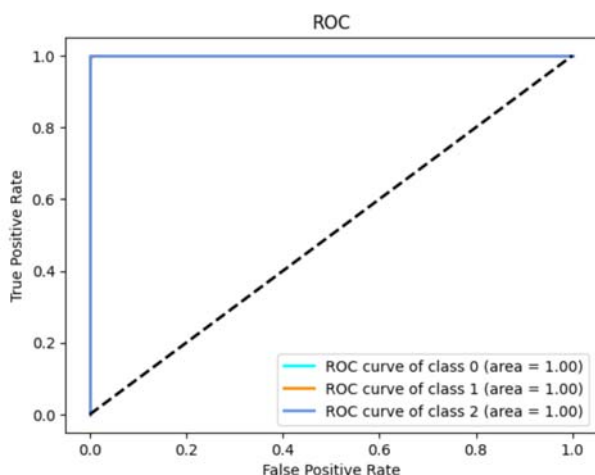


Figure 13 – ROC-curve of a developed architecture

In addition, the optimized architecture has lower memory requirements because it is more compact than the original RegNet. It is important to note that this model stores spatial information in intermediate layers more efficiently, which is critical for histological image classification tasks where cell and tissue structure patterns are highly significant.

In this study, the experiments were not conducted with high-resolution images due to limited hardware resources and a considerable training time.

The maximum image size used was 512×512 pixels. To work with higher resolutions, it is necessary to increase the amount of available GPU memory, which is a promising area for future research.

It should also be noted that the experimental base was limited to histological images of breast cancer. Accordingly, one of the key areas of further work is to expand the scope of the model by adapting it to the classification of histological images of other organs and working with higher resolution data. This opens up additional opportunities for implementing the developed architecture in clinical practice and research.

CONCLUSIONS

The paper solves the scientific and applied problem of improving the accuracy of histological image classification by gradually optimizing the architecture of a convolutional neural network and improving the learning process by using synthetic and transformed data.

For the first time, a consistent approach to building a classification model has been implemented, which includes:

- 1) training and comparison of 12 known CNN architectures on different types of histological datasets to determine the model with the highest accuracy, which allowed us to assess the impact of different data preparation options on classification results;
- 2) integrating the cell found by the method described in [35] into the CNN model and further optimizing it to improve accuracy.

The scientific novelty is to improve the approach to building efficient CNN architectures for histological image classification:

- for the first time, the effectiveness of using synthetic histological images generated by the diffusion model for training a classifier was investigated;
- the process of step-by-step comparison, training, and optimization has been improved, which has led to a significant increase in accuracy.

The practical significance is the software implementation of a module that can be integrated into an automatic diagnostic system to improve the quality of medical research [37].

Prospects for further research is the development of a CAD system. The system uses neural networks to segment, classify, and generate biomedical images [38, 39].

ACKNOWLEDGEMENTS

The authors of the article express their gratitude to the Department of Pathological Anatomy with a Sectional Course and Forensic Medicine of I. Ya. Horbachevsky Ternopil National Medical University (Ternopil, Ukraine) for providing histopathological images of breast cancer and personally to Doctor of Medical Sciences, Professor, and Head of the Department Petro Romanovych Selskyi.

REFERENCES

1. Underwood J. C. E. More than meets the eye: the changing face of histopathology, *Histopathology*, 2016, Vol. 70, № 1, pp. 4–9. DOI: 10.1111/his.13047.
2. Schafer et al. K. A. Use of Severity Grades to Characterize Histopathologic Changes, *Toxicologic Pathology*, 2018, Vol. 46, № 3, pp. 256–265. DOI: 10.1177/0192623318761348.
3. Ho C., Rodig S. J. Immunohistochemical markers in lymphoid malignancies: Protein correlates of molecular alterations, *Seminars in Diagnostic Pathology*, 2015, Vol. 32, № 5, pp. 381–391. DOI: 10.1053/j.semdp.2015.02.016.
4. Li Z. et al. Large-scale retrieval for medical image analytics: A comprehensive review, *Medical Image Analysis*, 2018, Vol. 43, pp. 66–84. DOI: 10.1016/j.media.2017.09.007.
5. Zeiser FE. et al. Breast cancer intelligent analysis of histopathological data: A systematic review, *Applied Soft Computing*, 2021, Vol. 113, P. 107886. DOI: 10.1016/j.asoc.2021.107886.
6. Houssein E. H. et al. Deep and machine learning techniques for medical imaging-based breast cancer: A comprehensive review, *Expert Systems with Applications*, 2020, Vol. 167, P. 114161. DOI: 10.1016/j.eswa.2020.114161.
7. He W. et al. A review: The detection of cancer cells in histopathology based on machine vision, *Computers in Biology and Medicine*, 2022, Vol. 146, P. 105636. DOI: 10.1016/j.combiomed.2022.105636.
8. Komura D., Ochi M., Ishikawa S. Machine learning methods for histopathological image analysis: Updates in 2024, *Computational and Structural Biotechnology Journal*. – 2024, Vol. 27, pp. 383–400. DOI: 10.1016/j.csbj.2024.12.033.
9. Abdelsamea M. M. et al. A survey on artificial intelligence in histopathology image analysis, *WIREs Data Mining and Knowledge Discovery*, 2022, Vol. 12, № 6, P. e1474. DOI: 10.1002/widm.1474.

10. Yan T. et al. Convolutional Neural Network with Parallel Convolution Scale Attention Module and ResCBAM for Breast Histology Image Classification, *Heliyon*, 2024, Vol. 10, № 10, P. e30889. DOI: 10.1016/j.heliyon.2024.e30889.
11. Rafiq A. et al. Detection and Classification of Histopathological Breast Images Using a Fusion of CNN Frameworks, *Diagnostics*, 2023, Vol. 13, № 10, P. 1700. DOI: 10.3390/diagnostics13101700.
12. Peng C. C., Lee B. R. Enhancing colorectal cancer histological image classification using transfer learning and ResNet50 CNN Model, 2023 *IEEE 5th Eurasia Conference on Biomedical Engineering, Healthcare and Sustainability (ECBIOS) : proceedings*, 2023, pp. 36–40. DOI: 10.1109/ECBIOS57802.2023.10218590.
13. Srikantamurthy M. M. et al. Classification of benign and malignant subtypes of breast cancer histopathology imaging using hybrid CNN-LSTM based transfer learning, *BMC Medical Imaging*, 2023, Vol. 23, № 1. DOI: 10.1186/s12880-023-00964-0.
14. Al-Jabbar M. et al. Analyzing Histological Images Using Hybrid Techniques for Early Detection of Multi-Class Breast Cancer Based on Fusion Features of CNN and Hand-crafted, *Diagnostics*, 2023, Vol. 13, № 10, P. 1753. DOI: 10.3390/diagnostics13101753.
15. Ruiz FM. et al. CNN stability training improves robustness to scanner and IHC-based image variability for epithelium segmentation in cervical histology, *Frontiers in Medicine*, 2023, Vol. 10. DOI: 10.3389/fmed.2023.1173616.
16. Ünlülal N. et al. Histological tissue classification with a novel statistical filter based convolutional neural network, *Anatomia, Histologia, Embryologia*, 2024, Vol. 53, № 4. DOI: 10.1111/ahel.13073.
17. Dogar G. M., Shahzad M., Fraz M. M. Attention augmented distance regression and classification network for nuclei instance segmentation and type classification in histology images, *Biomedical Signal Processing and Control*, 2023, Vol. 79, P. 104199. DOI: 10.1016/j.bspc.2022.104199.
18. Xiao S. et al. A scale and region-enhanced decoding network for nuclei classification in histology image, *Biomedical Signal Processing and Control*, 2023, Vol. 83, P. 104626. DOI: 10.1016/j.bspc.2023.104626.
19. Carreras J. Celiac Disease Deep Learning Image Classification Using Convolutional Neural Networks / J. Carreras // *Journal of Imaging*. – 2024. – Vol. 10, № 8. – P. 200. – DOI: 10.3390/jimaging10080200.
20. Chandana Mani R. K., Kamalakannan J. The Comparative Study of CNN models for Breast Histopathological Image Classification, 2023 *International Conference on Computer Communication and Informatics (ICCCI), Coimbatore, India, 23–25 January 2023 : proceedings*, 2023. DOI: 10.1109/iccci56745.2023.10128352.
21. Weber A. et al. AI-Based Detection of Oral Squamous Cell Carcinoma with Raman Histology, *Cancers*, 2024, Vol. 16, № 4, P. 689. DOI: 10.3390/cancers16040689.
22. Shimizu T. et al. A trial deep learning-based model for four-class histologic classification of colonic tumor from narrow band imaging, *Scientific Reports*, 2023, Vol. 13, № 1. DOI: 10.1038/s41598-023-34750-3.
23. Karthikeyan A., Jothilakshmi S., Suthir S. Colorectal cancer detection based on convolutional neural networks (CNN) and ranking algorithm, *Measurement: Sensors*, 2024, Vol. 31, P. 100976. DOI: 10.1016/j.measen.2023.100976.
24. Amin M. S., Ahn H. FabNet: A Features Agglomeration-Based Convolutional Neural Network for Multiscale Breast Cancer Histopathology Images Classification, *Cancers*, 2023, Vol. 15, № 4, P. 1013. DOI: 10.3390/cancers15041013.
25. Sawant D., Kundale J., Varma P. Hybrid Deep Learning based Multi-Classification of Breast Cancer Approach using Histology Images, 2024 *2nd International Conference on Advancement in Computation & Computer Technologies (InCACCT), Gharuan, India, 2–3 May 2024 : proceedings*, 2024. DOI: 10.1109/incacct61598.2024.10551068.
26. Sreelekshmi V., Pavithran K., Jyothisha J. Nair SwinCNN: An Integrated Swin Transformer and CNN for Improved Breast Cancer Grade Classification, *IEEE Access*, 2024, Vol. 12, P. 1. DOI: 10.1109/access.2024.3397667.
27. Kanadath A., Jothi J. A. A., Urolagin S. CViTNet: A CNN-ViT Network with Skip Connections for Histopathology Image Classification, *IEEE Access*, 2024, Vol. 12, P. 1. DOI: 10.1109/access.2024.3448302.
28. Muniasamy A. et al. Lung cancer histopathology image classification using transfer learning with convolution neural network model, *Technology and Health Care*, 2023, Vol. 32, № 2, pp. 1–12. DOI: 10.3233/thc-231029.
29. Kousalya K., Saranya T. Improved the detection and classification of breast cancer using hyper parameter tuning, *Materials Today: Proceedings*, 2021, Vol. 81. DOI: 10.1016/j.matpr.2021.03.707.
30. Kanimozhi S., Priyadarsini S. Breast Cancer Histopathological Image Classification Using CNN and VGG-19, 2024 *Third International Conference on Intelligent Techniques in Control, Optimization and Signal Processing (INCOS), Krishnankoil, Virudhunagar distric. Tamil Nadu, India, 14–16 March 2024 : proceedings*, 2024. DOI: 10.1109/incos59338.2024.10527543.
31. Chanchal A. K. et al. Classification and grade prediction of kidney cancer histological images using deep learning [Electronic resource], *Multimedia Tools and Applications*, 2024, Vol. 83, pp. 78247–78267. DOI: 10.1007/s11042-024-18639-5.
32. Simonyan E. O., Badejo J. A., Weijin J. S. Histopathological breast cancer classification using CNN, *Materials Today: Proceedings*, 2023, Vol. 105, pp. 268–275. DOI: 10.1016/j.matpr.2023.10.154.
33. Berezsky O. et al. Synthesis of Convolutional Neural Network architectures for biomedical image classification, *Biomedical Signal Processing and Control*, 2024, Vol. 95, P. 106325. DOI: 10.1016/j.bspc.2024.106325.
34. Real E. et al. Regularized evolution for image classifier architecture search, *Proceedings of the AAAI Conference on Artificial Intelligence : proceedings*, 2019, pp. 4780–4789. DOI: 10.48550/arXiv.1802.01548.
35. Berezsky O. M., Liashchynskyi P. B. Method of generative-adversarial networks searching architectures for biomedical images synthesis, *Radio Electronics, Computer Science, Control*, 2024, № 1, P. 104. DOI: 10.15588/1607-3274-2024-1-10.
36. Radosavovic I. et al. Designing Network Design Spaces, 2020 *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) : proceedings*, 2020. DOI: 10.48550/arXiv.2003.13678.
37. Berezsky O. M., Liashchynskyi P. B. Development of the architecture of a computer aided diagnosis system in medicine, *Applied Aspects of Information Technology*, 2024, Vol. 7, № 4, P. 359–369. DOI: 10.15276/aait.07.2024.25.
38. Berezsky O. M., Pitsun O. Y. Evaluation methods of image segmentation quality, *Radio Electronics, Computer Science,*

Control, 2018, № 1, pp. 119–128. DOI: 10.15588/1607-3274-2018-1-14.

39. [Berezsky O. M., Liashchynskyi P. B. et al. Deep network-based method and software for small sample biomedical image generation and classification, *Radio Electronics*,

Computer Science, Control, 2024, № 4, P. 76. DOI: 10.15588/1607-3274-2023-4-8.

Received 13.07.2025.

Accepted 14.10.2025.

УДК 004.8

КЛАСИФІКАЦІЯ ГІСТОЛОГІЧНИХ ЗОБРАЖЕНЬ НА ОСНОВІ ЗГОРТКОВИХ НЕЙРОННИХ МЕРЕЖ

Ляшинський П. Б. – аспірант кафедри комп'ютерної інженерії, Західноукраїнський національний університет, Тернопіль, Україна.

АНОТАЦІЯ

Актуальність. Велике значення для прискорення та підвищення точності діагностики в медицині має автоматизована класифікація гістологічних зображень. Беручи до уваги складність і високу варіативність гістологічних структур, застосування глибокого навчання і згорткових нейронних мереж зокрема, є перспективним напрямком у вирішенні цієї задачі.

Об'єктом дослідження є процес класифікації гістологічних зображень з допомогою згорткових нейронних мереж для визначення та оптимізації архітектури із найвищим показником точності.

Мета роботи. Метою роботи є розробка ефективного підходу до класифікації гістологічних зображень із використанням згорткових нейронних мереж, який забезпечує високу точність за рахунок поетапної оптимізації архітектури та застосування методів розширення даних з допомогою афінних перетворень та синтезу на основі дифузійних моделей.

Метод. Дослідження включає чотири основні етапи. На першому проведено порівняльний аналіз 12 відомих архітектур згорткових нейронних мереж на базовому гістологічному датасеті. Другий та третій етапи передбачають порівняльний аналіз на розширених даних, що включають афінні перетворення та синтетичні зображення, згенеровані дифузійною моделлю відповідно. На фінальному, четвертому етапі виконується нейроеволюційний пошук оптимальної архітектурної комірки. Після її знаходження вона інтегрується у розроблену архітектуру моделі, де для кожного шару здійснюється вибір між певними блоками та знайденою коміркою. Такий підхід дозволяє автоматично сформувати оптимальну послідовність блоків у моделі, що забезпечує найвищу точність класифікації.

Результати. Запропонований підхід дозволив покращити точність класифікації гістологічних зображень порівняно з початковими архітектурами. Додавання синтетичних зображень до навчального набору забезпечило приріст продуктивності моделей. Пошук оптимальної комірки та її інтеграція в модель з подальшою оптимізацією продемонстрував додаткове покращення якості класифікації, збільшивши точність до 94,9, 96,1 та 99,8% на кожному датасеті відповідно.

Висновки. Запропонований підхід дозволяє досягнути високої точності класифікації гістологічних зображень завдяки поетапному процесу, що включає використання класичних архітектур згорткових нейронних мереж, генерацію синтетичних даних та пошук оптимальних архітектурних і гіперпараметричних конфігурацій. Розроблено програмний модуль класифікації гістологічних зображень, який може бути використаний у системі автоматичного діагностування.

КЛЮЧОВІ СЛОВА: згорткові нейронні мережі, класифікація, оптимізація, гістологія.

ЛІТЕРАТУРА

- Underwood J. C. E. More than meets the eye: the changing face of histopathology / J. C. E. Underwood // *Histopathology*. – 2016. – Vol. 70, № 1. – P. 4–9. DOI: 10.1111/his.13047.
- Use of Severity Grades to Characterize Histopathologic Changes / [K. A. Schafer et al.] // *Toxicologic Pathology*. – 2018. – Vol. 46, № 3. – P. 256–265. DOI: 10.1177/0192623318761348.
- Ho C. Immunohistochemical markers in lymphoid malignancies: Protein correlates of molecular alterations / C. Ho, S. J. Rodig // *Seminars in Diagnostic Pathology*. – 2015. – Vol. 32, № 5. – P. 381–391. DOI: 10.1053/j.semdp.2015.02.016.
- Large-scale retrieval for medical image analytics: A comprehensive review / [Z. Li et al.] // *Medical Image Analysis*. – 2018. – Vol. 43. – P. 66–84. DOI: 10.1016/j.media.2017.09.007.
- Breast cancer intelligent analysis of histopathological data: A systematic review / [F. E. Zeiser et al.] // *Applied Soft Computing*. – 2021. – Vol. 113. – P. 107886. DOI: 10.1016/j.asoc.2021.107886.
- Deep and machine learning techniques for medical imaging-based breast cancer: A comprehensive review / [E. H. Houssein et al.] // *Expert Systems with Applications*. – 2020. – Vol. 167. – P. 114161. DOI: 10.1016/j.eswa.2020.114161.
- A review: The detection of cancer cells in histopathology based on machine vision / [W. He et al.] // *Computers in Biology and Medicine*. – 2022. – Vol. 146. – P. 105636. DOI: 10.1016/j.combiomed.2022.105636.
- Komura D. Machine learning methods for histopathological image analysis: Updates in 2024 / D. Komura, M. Ochi, S. Ishikawa // *Computational and Structural Biotechnology Journal*. – 2024. – Vol. 27. – P. 383–400. DOI: 10.1016/j.csbj.2024.12.033.
- A survey on artificial intelligence in histopathology image analysis / [M. M. Abdelsamea et al.] // *WIREs Data Mining and Knowledge Discovery*. – 2022. – Vol. 12, № 6. – P. e1474. DOI: 10.1002/widm.1474.
- Convolutional Neural Network with Parallel Convolution Scale Attention Module and ResCBAM for Breast Histology Image Classification / [T. Yan et al.] // *Heliyon*. – 2024. – Vol. 10, № 10. – P. e30889. DOI: 10.1016/j.heliyon.2024.e30889.
- Detection and Classification of Histopathological Breast Images Using a Fusion of CNN Frameworks / [A. Rafiq et al.] // *Diagnostics*. – 2023. – Vol. 13, № 10. – P. 1700. DOI: 10.3390/diagnostics13101700.
- Peng C. C. Enhancing colorectal cancer histological image classification using transfer learning and ResNet50 CNN Model / C. C. Peng, B. R. Lee // *2023 IEEE 5th Eurasia Conference on Biomedical Engineering, Healthcare and Sustainability (ECBIOS) : proceedings*. – 2023. – P. 36–40. DOI: 10.1109/ECBIOS57802.2023.10218590.

13. Classification of benign and malignant subtypes of breast cancer histopathology imaging using hybrid CNN-LSTM based transfer learning / [M. M. Srikantamurthy et al.] // BMC Medical Imaging. – 2023. – Vol. 23, № 1. DOI: 10.1186/s12880-023-00964-0.
14. Analyzing Histological Images Using Hybrid Techniques for Early Detection of Multi-Class Breast Cancer Based on Fusion Features of CNN and Handcrafted / [M. Al-Jabbar et al.] // Diagnostics. – 2023. – Vol. 13, № 10. – P. 1753. DOI: 10.3390/diagnostics13101753.
15. CNN stability training improves robustness to scanner and IHC-based image variability for epithelium segmentation in cervical histology / [F.M. Ruiz et al.] // Frontiers in Medicine. – 2023. – Vol. 10. DOI: 10.3389/fmed.2023.1173616.
16. Histological tissue classification with a novel statistical filter based convolutional neural network / [N. Ünlükal et al.] // Anatomia, Histologia, Embryologia. – 2024. – Vol. 53, № 4. DOI: 10.1111/ahel.13073.
17. Dogar G. M. Attention augmented distance regression and classification network for nuclei instance segmentation and type classification in histology images / G. M. Dogar, M. Shahzad, M. M. Fraz // Biomedical Signal Processing and Control. – 2023. – Vol. 79. – P. 104199. DOI: 10.1016/j.bspc.2022.104199.
18. A scale and region-enhanced decoding network for nuclei classification in histology image / [S. Xiao et al.] // Biomedical Signal Processing and Control. – 2023. – Vol. 83. – P. 104626. DOI: 10.1016/j.bspc.2023.104626.
19. Carreras J. Celiac Disease Deep Learning Image Classification Using Convolutional Neural Networks / J. Carreras // Journal of Imaging. – 2024. – Vol. 10, № 8. – P. 200. – DOI: 10.3390/jimaging10080200.
20. Chandana Mani R. K. The Comparative Study of CNN models for Breast Histopathological Image Classification / R. K. Chandana Mani, J. Kamalakannan // 2023 International Conference on Computer Communication and Informatics (ICCCI), Coimbatore, India, 23–25 January 2023 : proceedings. – 2023. DOI: 10.1109/iccci56745.2023.10128352.
21. AI-Based Detection of Oral Squamous Cell Carcinoma with Raman Histology / [A. Weber et al.] // Cancers. – 2024. – Vol. 16, № 4. – P. 689. DOI: 10.3390/cancers16040689.
22. A trial deep learning-based model for four-class histologic classification of colonic tumor from narrow band imaging / [T. Shimizu et al.] // Scientific Reports. – 2023. – Vol. 13, № 1. DOI: 10.1038/s41598-023-34750-3.
23. Karthikeyan A. Colorectal cancer detection based on convolutional neural networks (CNN) and ranking algorithm / A. Karthikeyan, S. Jothilakshmi, S. Suthir // Measurement: Sensors. – 2024. – Vol. 31. – P. 100976. DOI: 10.1016/j.measen.2023.100976.
24. Amin M. S. FabNet: A Features Agglomeration-Based Convolutional Neural Network for Multiscale Breast Cancer Histopathology Images Classification / M. S. Amin, H. Ahn // Cancers. – 2023. – Vol. 15, № 4. – P. 1013. DOI: 10.3390/cancers15041013.
25. Sawant D. Hybrid Deep Learning based Multi-Classification of Breast Cancer Approach using Histology Images / D. Sawant, J. Kundale, P. Varma // 2024 2nd International Conference on Advancement in Computation & Computer Technologies (InCACCT), Gharuan, India, 2–3 May 2024 : proceedings. – 2024. DOI: 10.1109/incacct61598.2024.10551068.
26. Sreelekshmi V. SwinCNN: An Integrated Swin Transformer and CNN for Improved Breast Cancer Grade Classification / V. Sreelekshmi, K. Pavithran, Jyothisha J. Nair // IEEE Access. – 2024. – Vol. 12. – P. 1. DOI: 10.1109/access.2024.3397667.
27. Kanadath A. CVITS-Net: A CNN-ViT Network with Skip Connections for Histopathology Image Classification / A. Kanadath, J. A. A. Jothi, S. Urolagin // IEEE Access. – 2024. – Vol. 12. – P. 1. DOI: 10.1109/access.2024.3448302.
28. Lung cancer histopathology image classification using transfer learning with convolution neural network model / [A. Muniasamy et al.] // Technology and Health Care. – 2023. – Vol. 32, № 2. – P. 1–12. DOI: 10.3233/thc-231029.
29. Kousalya K. Improved the detection and classification of breast cancer using hyper parameter tuning / K. Kousalya, T. Saranya // Materials Today: Proceedings. – 2021. – Vol. 81. DOI: 10.1016/j.matpr.2021.03.707.
30. Kanimozhi S. Breast Cancer Histopathological Image Classification Using CNN and VGG-19 / S. Kanimozhi, S. Priyadarsini // 2024 Third International Conference on Intelligent Techniques in Control, Optimization and Signal Processing (INCOS), Krishnankoil, Virudhunagar district, Tamil Nadu, India, 14–16 March 2024 : proceedings. – 2024. DOI: 10.1109/incos59338.2024.10527543.
31. Classification and grade prediction of kidney cancer histological images using deep learning [Electronic resource] / [A. K. Chanchal et al.] // Multimedia Tools and Applications. – 2024. – Vol. 83. – P. 78247–78267. DOI: 10.1007/s11042-024-18639-5.
32. Simonyan E. O. Histopathological breast cancer classification using CNN / E. O. Simonyan, J. A. Badejo, J. S. Weijin // Materials Today: Proceedings. – 2023. – Vol. 105. – P. 268–275. DOI: 10.1016/j.matpr.2023.10.154.
33. Synthesis of Convolutional Neural Network architectures for biomedical image classification / [O. Berezsky et al.] // Biomedical Signal Processing and Control. – 2024. – Vol. 95. – P. 106325. – DOI: 10.1016/j.bspc.2024.106325.
34. Regularized evolution for image classifier architecture search / [E. Real et al.] // Proceedings of the AAAI Conference on Artificial Intelligence : proceedings. – 2019. – P. 4780–4789. DOI: 10.48550/arXiv.1802.01548.
35. Berezsky O. M. Method of generative-adversarial networks searching architectures for biomedical images synthesis / O. M. Berezsky, P. B. Liashchynskiy // Radio Electronics, Computer Science, Control. – 2024. – № 1. – P. 104. DOI: 10.15588/1607-3274-2024-1-10.
36. Designing Network Design Spaces / [I. Radosavovic et al.] // 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) : proceedings. – 2020. DOI: 10.48550/arXiv.2003.13678.
37. Berezsky O. M. Development of the architecture of a computer aided diagnosis system in medicine / O. M. Berezsky, P. B. Liashchynskiy // Applied Aspects of Information Technology. – 2024. – Vol. 7, № 4. – P. 359–369. DOI: 10.15276/aait.07.2024.25.
38. Berezsky O. M. Evaluation methods of image segmentation quality / O. M. Berezsky, O. Y. Pitsun // Radio Electronics, Computer Science, Control. – 2018. – № 1. – P. 119–128. DOI: 10.15588/1607-3274-2018-1-14.
39. Deep network-based method and software for small sample biomedical image generation and classification / [O. M. Berezsky, P. B. Liashchynskiy et al.] // Radio Electronics, Computer Science, Control. – 2024. – № 4. – P. 76. DOI: 10.15588/1607-3274-2023-4-8.

ALZHEIMER'S DISEASE PREDICTION BY USING DEEP STACKED ENSEMBLE MODEL ENHANCED WITH SQUEEZE-AND-EXCITATION ATTENTION MECHANISM

Muhammad Junaid Iqbal – Post-graduate student, Faculty of Computing, Gomal University, Dera Ismail Khan, Pakistan.

Mudasir Mahmood – Assistant Professor, Faculty of Computing, Gomal University, Dera Ismail Khan, Pakistan.

Muhammad Farhan – Lecturer, Faculty of Computing, Gomal University, Dera Ismail Khan, Pakistan.

Muhammad Sajjad Hussain – Lecturer, Faculty of Computing, Gomal University, Dera Ismail Khan, Pakistan.

ABSTRACT

Context. Alzheimer's disease (AD) is a progressive, neurological degenerative disease causing memory loss, impaired cognition, and dementia. Timely identification of AD is crucial for the provision of effective treatment and intervention. Magnetic Resonance Imaging (MRI) has also become a critical tool in understanding the structural changes in the brain that occur during Alzheimer's development. Nonetheless, the manual processing of MRI scans is time-consuming, subjective, and susceptible to human error. As such, there is increasing demand for automated and precise diagnostic technology that can support clinicians in the earlier detection and staging of Alzheimer's disease based on medical imaging data.

Objective. The present study focuses on developing and evaluating a deep learning-based stacked ensemble model for the classification and staging of Alzheimer's disease brain MRI scans. The primary objective is to improve the diagnosis accuracy and reliability through a combination of the strengths of several pre-trained convolutional neural network (CNN) architectures, combined with sophisticated attention mechanisms and meta-learning techniques.

Method. The proposed approach utilizes a deep stacked ensemble learning framework composed of three well-performing CNN architectures: MobileNetV2, ResNet50, and DenseNet121. These models are pre-trained on the ImageNet dataset, benefiting from robust feature extraction capabilities. To further improve their performance, each CNN model is enhanced with a Squeeze-and-Excitation (SE) attention module, which adaptively recalibrates channel-wise feature responses, emphasizing important features while suppressing irrelevant ones. The extracted high-level features from all three SE-augmented CNNs are then concatenated and fed into a meta-learner consisting of fully connected layers. This meta-classifier incorporates dropout and batch normalization techniques to prevent overfitting and improve generalization. The overall architecture is trained and validated on a dataset of brain MRI images categorized into different stages of Alzheimer's disease, including normal control, mild cognitive impairment, and various stages of dementia.

Results. The experimental evaluation demonstrated exceptional performance, achieving an Accuracy of 99%, a Precision of 99%, a Recall of 98%, and an F1-score of 99%. These metrics indicate the model's strong predictive capability and reliability in distinguishing between different stages of Alzheimer's disease.

Conclusions. The experimental outcomes highlight the effectiveness and robustness of the proposed deep stacked ensemble model in the automated diagnosis and staging of Alzheimer's disease using MRI scans. The integration of multiple CNNs with attention mechanisms and meta-learning significantly enhances classification performance. These findings suggest that the model can serve as a reliable decision-support system for neurologists, aiding in early diagnosis, timely intervention, and improved patient outcomes in clinical settings.

KEYWORDS: Convolution Neural Networks, Deep Learning, Alzheimer's disease, Squeeze-and-Excitation, Deep Stacked Ensemble.

ABBREVIATIONS

AD is an Alzheimer's disease;
CNN is a Convolutional Neural Network;
SE is a Squeeze-and-Excitation;
ROC is a Receiver Operating Characteristic;
MRI is a Magnetic Resonance Imaging;
AUC is an Area Under the Curve.

n is a number of samples;
 r is a reduction ratio;
 C is a channel in SE block;
 W is a width in SE block;
 s, z, e are intermediate variables in SE block;
 σ is a sigmoid activation function.

NOMENCLATURE

X is a set of input variables;
 f is a Classification function;
 Y is a set of output labels;
 H is a hypothesis space of deep learning models;
 TP is a True Positive;
 TN is a True Negative;
 FP is a False Positive;
 FN is a False Negative;
 d is a dimension of feature vectors;

INTRODUCTION

Alzheimer's disease is a neurological health problem that gradually impairs memory, thinking, and behavior, especially in older people [1–9]. AD needs to be detected early for timely treatment. According to the report, the predicted growth of U.S populations living with AD will increase from 6.5 million to 13.8 million by 2060 under the present state of affairs of medical treatment (Alzheimer Association 2023), and globally, approximately 55 million people are affected by

dementia, with over 60% of the cases in the countries described as middle and low-income countries.

In recent years, there has been some success in the automatic detection and classification of AD using neuroimaging data (such as magnetic resonance imaging scans) with the advance of deep learning models, especially Convolutional Neural Networks (CNN) [10,11]. However, there are still considerable gaps in the current models that should be filled to achieve more precise and accurate diagnoses. CNN-based models are widely used in AD detection by leveraging their ability to extract spatial features from neuroimaging data. Various studies have proposed novel CNN architectures, including 3D CNN architectures with multi-level features [12], adaptive hybrid attention networks [13], and deep feature fusion networks [14], to achieve better AD classification accuracy. However, despite technological advancements, CNN-based models are still heavily challenged in their ability to capture long-range dependencies and global context information from MRI scans. These pieces of information are key considerations in the diagnosis, particularly in the early stages of AD.

The object of study is the application of a deep stacked ensemble model with a Squeeze-and-Excitation (SE) attention mechanism for classifying and staging Alzheimer's disease (AD) using brain magnetic resonance imaging (MRI) scans.

The subject of the study focuses on applying deep learning to MRI images to enhance the reliable and easily understood multiclass classification of Alzheimer's disease, facilitating early recognition and informing medical decisions.

The purpose of the study is to develop a Deep Stacked ensemble model that integrates three strong CNN architectures (MobileNetV2, ResNet50, and DenseNet121) to augment feature extraction from MRI images for classifying Alzheimer's disease as well as introduces Squeeze-and-Excitation (SE) Attention Mechanism into every CNN to recalibrate the channel-wise feature maps, enabling the network to enhance meaningful features and reduce irrelevant features, thereby increasing diagnostic accuracy.

1 PROBLEM STATEMENT

Let $X = \{x_1, x_2, \dots, x_n\}$ be the set of input variables, where each $x_i \in R^d$ represents the feature vector extracted from an MRI scan of the brain. The problem of Alzheimer's disease classification and staging is to learn a function $f: X \rightarrow Y$ where $Y = \{y_1, y_2, \dots, y_n\}$ and each $y_i \in \{0, 1, 2, 3\}$ corresponds to one of the four Alzheimer's stages: 0 for NonDemented, 1 for VeryMildDemented, 2 for MildDemented, and 3 for ModerateDemented. The classification function f must be trained to minimize the categorical cross-entropy loss while satisfying constraints related to limited data availability, imbalanced class distribution, and computational efficiency. The model is expected to achieve high classification performance, as measured by accuracy, Precision, Recall, and F1-score.

2 REVIEW OF THE LITERATURE

Alzheimer's disease is a progressive neurodegenerative disorder that has an adverse influence on cognitive abilities, thus necessitating early and precise diagnostic criteria to support optimal management of the disease. The implementation of recent advancements in deep learning has demonstrated promising results for the automated classification and staging of Alzheimer's disease (AD) using brain MRI scans. This literature review examines recent studies that explore various deep-learning models for detecting and classifying Alzheimer's disease.

El-Latif, Chelloug, Alabdulhafith, and Hammad [12] introduce a lightweight Convolutional Neural Network (CNN) to achieve exact Alzheimer's Disease (AD) identification from MRI image data. The proposed model reached 99.22% accuracy for binary class identification and 95.93% for multi-classification, surpassing existing models in achievement. Furthermore, Sethuraman, Malaiyappan, Ramalingam, Basheer, Rashid, and Ahmad [13] apply Alexnet to identify different stages of AD from real-time ADNI datasets. The proposed model achieved an accuracy of 96.61% in binary class identification. Hazarika, Kandar, and Maji [14] developed a lightweight hybrid model that combines LeNet and AlexNet. The ensemble model achieved an accuracy of 93.58% in multiclass identification. Tuvshinjargal and Hwang [15] developed the VGG-C Transform model, achieving the best results with normalization added to VGG-C, resulting in a test accuracy of approximately 77.46%. Jraba, Elleuch, Ltifi, and Kherallah [16] utilized convolutional neural network (CNN) models to automatically identify relevant Alzheimer's disease from brain MR images. The cutting-edge performance of CNN models enables the identification of the four Alzheimer's disease stages from mild dementia through moderate dementia and their non-dementia manifestations. The study investigates three modern architectures – ResNet-152, VGG16, and Inception-V3 – to reveal complex image patterns in brain images. Hasan, Hossain, and Ullah [17] presented a weighted ensemble deep transfer learning framework that utilized two pre-trained CNN architectures, including ResNet152V2 and DenseNet201. The proposed framework's performance has been evaluated using two different MRI datasets to assess its generalization potential across diverse input data. Researchers perform an in-depth analysis of the framework's performance across various case studies. Experimental outcomes demonstrate that the weighted ensemble architecture outperforms the base models. Raza, Naseer, Tamoor, and Zafar [18] utilized a CNN-based deep learning model as the reference model. The performance of the proposed model was tested on different epochs, including 10, 25, and 50. The proposed model achieved an accuracy of 97.84%. Ramzan, Khan, Rehmat, Iqbal, Saba, Rehman, and Mehmood [19] utilize deep Residual Neural Networks (RNNs), including a transfer learning approach, for classifying the 6 AD stages. While, Alsubaie, Luo, and Shaikat [20] proposed Convolutional Neural

Networks (CNNs), Recurrent Neural Networks (RNNs), and Generative models to detect Alzheimer's Disease. Mirzaei and Adeli [21] used a Support Vector Machine, Random Forest, Convolutional Neural Network, and K-means in the detection and classification of AD with an emphasis on neuroimaging. Huang, Liu, Zhang, and Li [22] investigated the alterations in cerebral grey matter in patients with the early stage of MCI using voxel-based morphometry to diagnose the early stage of Alzheimer's disease. Jung, Luna, and Park [23] designed a new cGAN for generating high-quality 3D MR images at various stages of AD, incorporating an additional module for 3D spatial interpolation. In particular, the cGAN architecture comprises an attentional 2D generator, a 2D discriminator, and a 3D discriminator that generates ongoing 2D images along the axial view to create high-quality 3D MR volumes. Mohammed, Fakhrudeen, and Alani [24] focus on the datasets used to develop Deep Learning techniques. The study also breaks down the stages of AD diagnosis and focuses on pre-processing techniques. Reddy, Rangarajan, Shuaib, Jeribi, and Alam [25] perform a comparative analysis of VGG16, VGG19, and Alexnet models to classify Alzheimer's disease. The VGG16 and VGG19 achieved 100% accuracy, while AlexNet achieved 98.20%. Mahmud, Rahman, Fattah, Subrina, Khan, Alshamrani, and Alqahtani [26] introduce a novel viewpoint to diagnosing Alzheimer's disease by combining deep transfer learning and explainable artificial intelligence (XAI) approaches. The extensive experimentation conducted on popular pre-trained convolutional neural networks (CNNs) revealed the effectiveness of the ensembles, specifically Ensemble-1 (VGG16 and VGG19) and Ensemble-2 (DenseNet169 and DenseNet201), in terms of superior diagnostic performance, achieving up to 95% Accuracy, Precision, Recall, and F1 score. The author [27] presents a framework combining 3D Convolutional Neural Network (3D-CNN) and Recurrent Neural Network (RNN) techniques to identify Alzheimer's disease. Mehmood Khan [28] proposed a Siamese 4D-AlzNet model, achieving a remarkable 95.05% accuracy. Nehar, Bellaouar, Benyoub, and Aissa [29] employ VGG16 and VGG19 models to classify MRI images based on Alzheimer's Disease and achieved remarkable accuracy (98.14% to 99.59%). The authors [30] employ the ResNet-101 transfer learning technique to classify Alzheimer's disease. The proposed model achieves 98.21% and 97.45% accuracy on the ADNI and OASIS datasets. Khaleel and Lakizadeh [31] employ the FSBILSTM model on the OASIS dataset, achieving an accuracy of 99.6%.

In summary, the available studies indicate that deep learning models are successful in reliably identifying Alzheimer's disease from brain MRI scans. Although some models achieve higher performance through architectural innovations, others emphasize the importance of transfer learning, normalization, and data augmentation in enhancing diagnostic performance. Nevertheless, this still leaves the need for stronger, more

understandable models to generalize across different datasets and clinical situations. Incorporating attention mechanisms, such as Squeeze-and-Excitation (SE), offers excellent potential for enhancing feature extraction and model interpretability in the AD classification task.

Table 1 – Summary of related work

Serial No	Reference & Year	Model	Dataset	Classification Method (Accuracy)
1	[12] 2023	Lightweight CNN	OrDs	Binary: 99.22% & Multiclass: 95.93%
2	[13] 2023	AlexNet	ADNI	Binary: 96.61%
3	[14] 2023	Ensemble model (LetNet+ AlexNet)	ADNI	Multiclass: 93.58%
4	[15] 2022	VGG-C Transform model	Kaggle	Multiclass: 77.46%
5	[16] 2024	ResNet-152, VGG16, and Inception-V3	Images collected from different resources	Multi classifier: 98% for ResNet 152, 96% for Inception-v3, 80% for VGG 16
6	[17] 2024	Ensemble model (ResNet152V2 and DenseNet201)	(Alzheimer s-dataset-4-class-of-images) (12800 images) Kaggle Dataset (40384)	Multiclass: 89.95% for 12800 images 98% accuracy for 40384 images
7	[18] 2023	CNN	ADNI	97.84%
8	[25] 2024	VGG16 VGG19 Alex Net	Healthy Aging Data in the U.S. (250,000 Samples)	VGG16: 99.9 VGG19: 98.8 Alex Net:99.9
9	[27] 2024	Hybrid model (3D CNN and RNN)	ADNI	Hybrid Model: 99.5%
10	[30] 2024	Resnet-101	ADNI and OASIS datasets	ADNI: 98.21% OASIS: 97.45%
11	[32] 2024	DenseNet-201 EfficientNet-B0 GoogleNet Inception-v3 ResNet50 MobileNet-v2	(Alzheimer s-dataset-4-class-of-images)	DenseNet-201: 93.33 EfficientNet-B0: 90.32% GoogleNet: 92.57 % Inception-v3: 84.84% ResNet50: 88.95% MobileNet-v2: 91.02%
12	[31] 2025	FSBILSTM	OASIS	OASIS: 99.6%

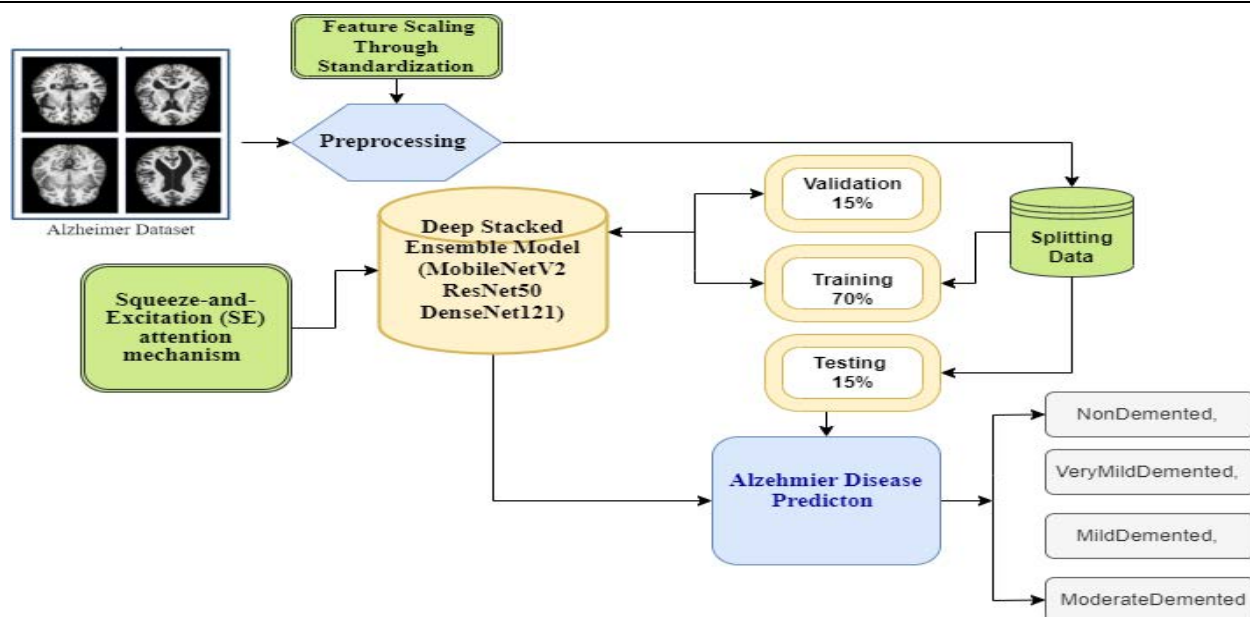


Figure 1 – Proposed Model

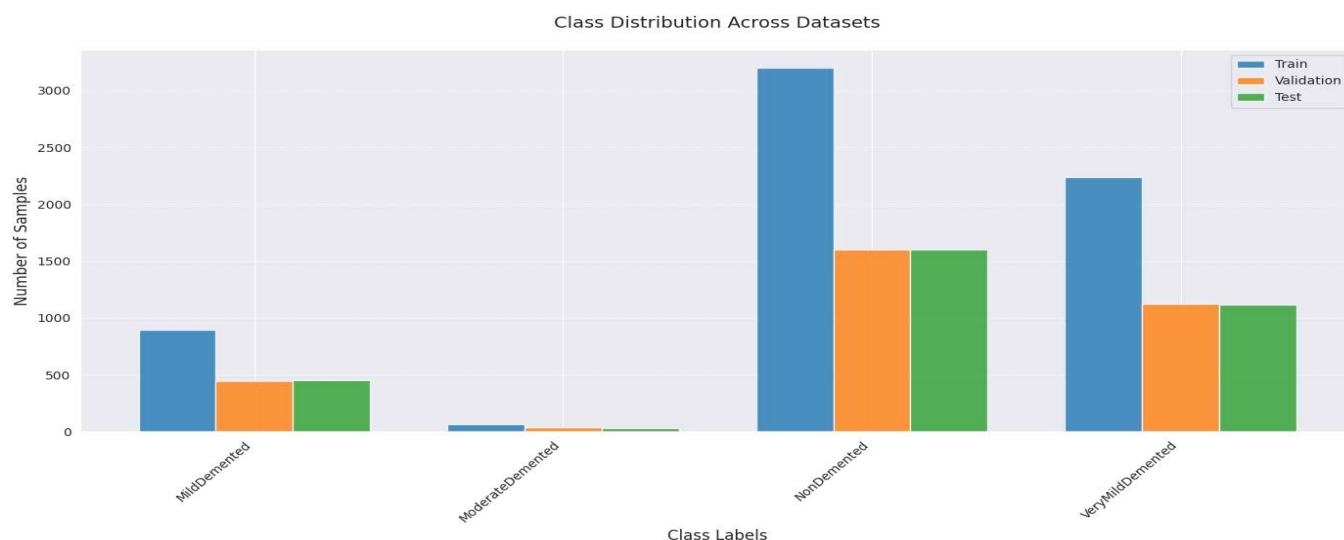


Figure 2 – Class Distribution

3 MATERIALS AND METHODS

The method developed in this research focused on designing and implementing an accurate and generalizable deep learning architecture for classifying and staging Alzheimer's disease based on brain MRI images. The overall process of the proposed model is shown in Figure 1.

The dataset used in this research is named "alzheimers-dataset-4-class-of-images". It includes four categories of brain MRI scans: Non-Demented, Very Mild Demented, Mild Demented, and Moderate Demented. The overall distribution of the dataset is shown in Figure 2. To achieve fair training and evaluation, the dataset was divided into three parts: 8.960 (70%) for training, 1.920 (15%) for testing, and 1.920 (15%) for validation.

The study proposed a deep-stacked ensemble framework that combines well-known CNN backbones,

including MobileNetV2, ResNet-50, and DenseNet-121, with feature extraction capabilities to train the classification model. Each base model was trained in advance using ImageNet to leverage transfer learning. To enhance attention to disease-relevant features in the MRI scans at the model level, each CNN is coupled with a Squeeze-and-Excitation (SE) attention block. This attention mechanism helps detect subtle variations in anatomic features that occur between different stages of AD.

After extracting the SE-enhanced features, the outputs from all three branches of the CNN are concatenated to create a single feature vector. The following steps are performed to train the data. The first step is to define the input shape. Let the shape of the input images be:

$$\text{img_shape} \leftarrow (224, 224, 3).$$

The second step is to determine the number of output classes.

The third step is to define the Squeeze-and-Excitation (SE) Block. Let $X \in R^{H \times W \times C}$ be the input tensor.

Define the SE block as follows: Compute global average pooling over spatial dimensions:

$$s \leftarrow \frac{1}{H \cdot W} \sum_{i=1}^H \sum_{j=1}^W X_{i,j}.$$

Pass through a bottleneck fully connected layer:

$$z \leftarrow \text{relu}(W_1 \cdot s + b_1), W_1 \in R^{C/r \times C}.$$

Expand back to the original channel size using sigmoid activation:

$$e \leftarrow \sigma(W_2 \cdot z + b_2), W_2 \in R^{C \times C/r}.$$

Reshape e to match X 's channel dimension and scale:

$$\text{Output} \leftarrow X \cdot \text{re.shape}(e).$$

Here, r is the reduction ratio.

The fourth step is to define the base model construction function. For any pre-trained architecture, define Global Average Pooling, Batch Normalization, a fully connected layer with 256 units, ReLU activation, and Dropout with probability 0.45.

The fifth step involves constructing three branches using different pre-trained models: MobileNetV2, ResNet-50, and DenseNet-121.

The last step is to apply shared input and concatenate Features. Finally, compile the final model with the details of the optimizer (Adamax) and the loss function (Categorical Crossentropy).

Evaluating the suggested deep stacked ensemble model for classifying Alzheimer's disease, the study employs various evaluation metrics to assess the model's performance at different stages of the disease, as determined by examining brain MRI scans. These metrics are accuracy, Precision, Recall, and F1-score.

Accuracy refers to the proportion of correct predictions in the overall model, calculated by counting the percentage of true positives and true negatives among all predictions. Mathematically, accuracy is represented as

$$\text{accuracy} = \frac{TP + TN}{TP + TN + FP + FN}. \quad (1)$$

Precision is the ratio of correctly predicted positive cases to the total predicted positives. Mathematically, Precision is represented as

$$\text{precision} = \frac{TP}{TP + FP}. \quad (2)$$

Recall (also known as sensitivity) is the percentage of true positives that the model correctly records. It is essential to reduce false negatives in medical use. Mathematically, Recall is represented as

$$\text{recall} = \frac{TP}{TP + FN}. \quad (3)$$

F1-Score is the harmonic average of Precision and Recall, yielding an aggregated number without compromising anyone. Mathematically, the F1 score is represented as

$$f1\text{-Score} = \frac{2 * (\text{precision} * \text{recall})}{(\text{precision} + \text{recall})}. \quad (4)$$

4 EXPERIMENTS

The proposed deep stacked ensemble model has been tested on the "alzheimers-dataset-4-class-of-images" dataset, which contains 12800 brain MRI scans categorized into four classes: Non-Demented, Very Mild Demented, Mild Demented, and Moderate Demented. Each category of the class is represented in Figure 3. All images were made compatible with standard deep learning models and processed uniformly by resizing them to 224×224 pixels and scaling their pixel values to the range of 0 to 1. In total, 8,960 images were used for training, while 15% (1,920 images) were included in the validation set and another 15% (1,920 images) in the testing set. Before training, specific pre-processing steps were applied to the images to enhance image quality and facilitate better model generalization. After that, the proposed deep stacked ensemble model combined the capabilities of MobileNetV2, ResNet50, and DenseNet121, which were pre-trained on the ImageNet dataset, to help accelerate the process of identifying useful features.

With every pre-trained model, the Squeeze-and-Excitation (SE) attention module was added to ensure it could adaptively control the importance of features in the layers, retaining the most important and discarding the least important ones. The Classification performance of the proposed model was qualitatively evaluated using standard evaluation criteria, including Precision, Recall, F1-score, and overall Accuracy.

5 RESULTS

The model's overall classification accuracy was impressive, amounting to 99%, and class-wise metrics are summarized in Table 2.

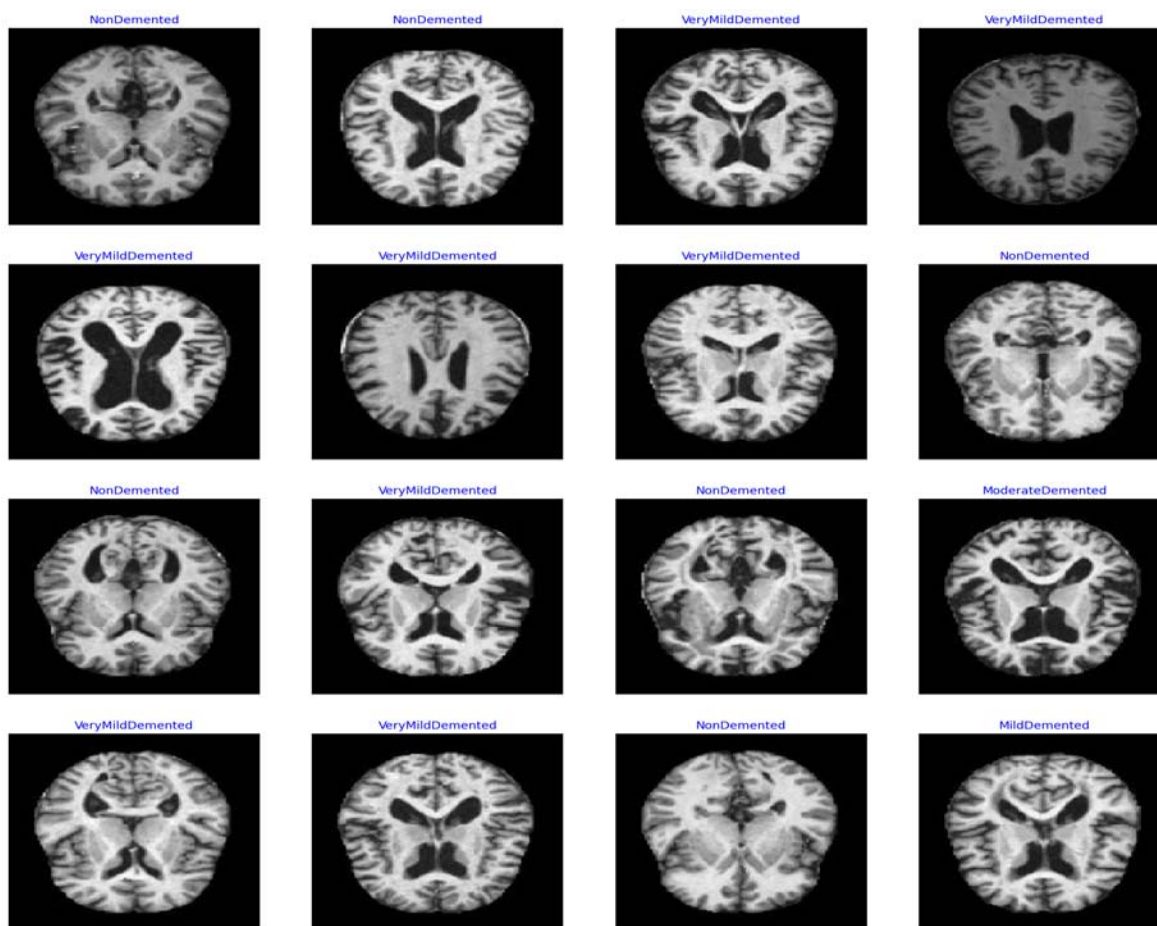


Figure 3 – Alzheimer's Disease Class Images

Table 2 – Classification Report of the Proposed Model

Class	Precision	Recall	F1-Score	Support
MildDemented	1.00	0.96	0.98	278
ModerateDemented	1.00	1.00	1.00	20
NonDemented	0.98	1.00	0.99	959
VeryMildDemented	1.00	0.98	0.99	663
Average	0.99	0.98	0.99	1920
Accuracy	99 %			

The training and validation accuracy, as well as the loss of the proposed model, are represented in Figures 4 and 5.

Figure 4 presents a visual representation of the training and validation results for the proposed deep-stacked ensemble model, showing the accuracy and loss trends over the epochs. The amounts of training and validation accuracy from the accuracy graph can improve significantly during the first few epochs, only stabilizing at 99% towards the end of the 10th epoch, demonstrating the model's learning and generalization capabilities. At the same time, the loss curve shows significant early improvements in the errors of both training and validation sets, which drop to nearly zero as the model converges. Furthermore, the confusion matrix of the proposed model is represented in Figure 6.



Figure 4 – Training and validation Loss

Figure 6 illustrates the confusion matrix for the proposed deep-stacked ensemble model in classifying Alzheimer's disease stages using brain MRI scans. A confusion matrix summarizes the performance of a classification algorithm by giving the counts of true positives, false positives, true negatives, and false negatives for each class. This matrix indicates the extent to which our model has succeeded in distinguishing the

four classes of Alzheimer's disease: Non-Demented, Very Mild Demented, Mild Demented, and Moderate Demented. The diagonal elements of the confusion matrix indicate the correlation classes for each category and demonstrate the model's efficiency, as these counts should be close to the sample sizes for each class. The ROC curve of the proposed model is represented in Figure 7.

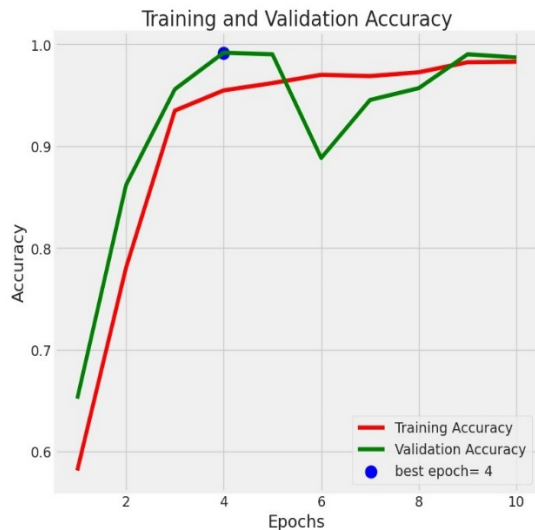


Figure 5 – Training and Validation Accuracy

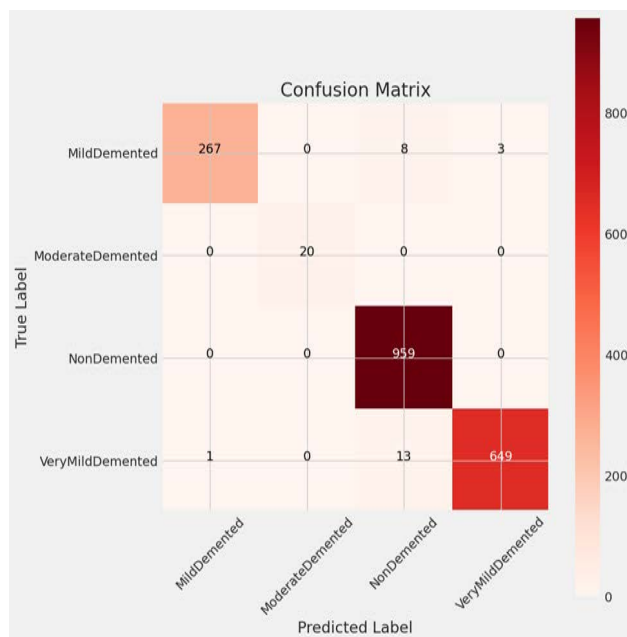


Figure 6 – Confusion Matrix of Proposed Model

Figure 7 shows the results of the ROC curve analysis for the proposed deep stacked ensemble approach, which describes the relation between the true positive rate (sensitivity) and the false positive rate (1-specificity) for different threshold settings. Figure 6 presents a plot comparing the ROC curves for the NonDemented, VeryMildDemented, MildDemented, and ModerateDemented classes, allowing for an intuitive assessment of how well the model distinguishes between stages.

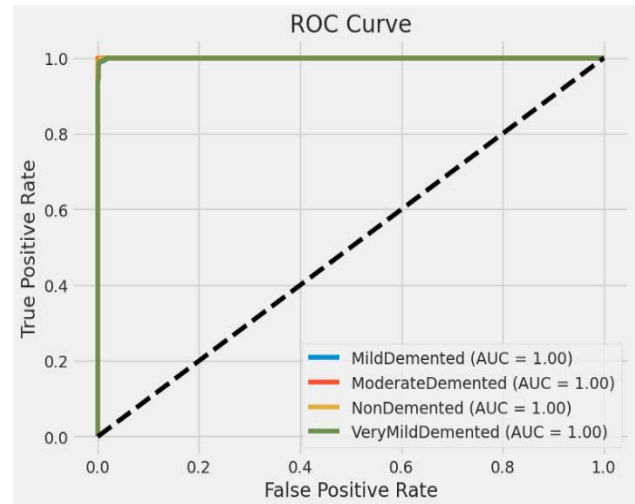


Figure 7 – ROC Curve

One key observation that results from the analysis of the ROC curve is that the AUC (1.0 for each class), which is derived from the reported accuracy, Precision, Recall, and F1-score of 99%, 99%, 98%, and 99% correspondingly, indicates that the model performance is overall excellent.

6 DISCUSSION

A comparison to the state-of-the-art models was carried out to ensure the performance of the proposed approach. The proposed deep-stacked ensemble model outperforms existing models in terms of accuracy. A comparative analysis is presented in Table 3.

Table 3 – Comparison of Proposed Model with Similar Studies

Model	Accuracy	Reference
DenseNet-201	93.33%	[32]
EfficientNet-B0:	90.32%	
GoogleNet	92.57 %	
Inception-v3	84.84%	
ResNet50	88.95%	
MobileNet-v2	91.02%	
Ensemble model (ResNet152V2 and DenseNet201)	89%	[17]
Proposed Model Deep Stacked Ensemble Model (MobileNetV2, ResNet50 , DenseNet121) With SE mechanism	99%	

The proposed model achieved 99% accuracy, surpassing previous research and demonstrating strong generalization and discriminative ability. The study also demonstrates that the SE block is crucial in recalibrating feature importance, resulting in significant gains in sensitivity levels and a reduction in false negatives –a vital step in producing accurate medical diagnoses. Therefore, the deep stacked ensemble framework establishes new ground in AD classification, achieving significant accuracy, Precision, Recall, and F1-score.

CONCLUSIONS

The present work proposes and evaluates a deep stacked ensemble model for both staging and classifying Alzheimer's disease using brain MRI images. To optimize the model, three CNN frameworks were employed: MobileNetV2, ResNet-50, and DenseNet-121. These frameworks included the Squeeze-and-Excitation (SE) attention mechanism, which enabled the model to focus on features important to the task. This design successfully addressed the issues associated with redundant features and improved classification accuracy across all four AD stages. The developed model demonstrated remarkable generalization and robustness, with an overall accuracy of 99%, impressive Precision, Recall, and F1-score, and superior performance compared to most current state-of-the-art methods. Moreover, the ROC curve also verified the model's outstanding discriminative power across all stages of Alzheimer's disease, with an AUC that approached 1.0. This highlights the practical utility of the method in clinical settings where AD detection is early and highly accurate.

Despite its promising results, some of the most critical aspects of this research remain to be addressed. One significant limitation is that the work is based on a single readily accessible dataset. However, it is diverse; it likely fails to capture the entire field of real-world clinical heterogeneity. Extending the model's robustness and generalizability is possible through an evaluation of the model against more highly representative multi-center datasets that include chronologically updated patient records. The incorporation of additional modality features, such as PET scans or clinical records, into a combined approach for deep learning has the potential to provide a more comprehensive overview of the patient's overall situation. Additionally, the use of XAI strategies may enhance clinical confidence by making the model's predictions more easily interpretable for neurologists. By optimizing the model for real-time execution at edge devices, it could be deployed readily in any resource-constrained healthcare setting, which could significantly enhance its contribution to AD in early diagnosis.

REFERENCES

- Rahbani R. G. Al, Ioannou A., Wang T. Alzheimer's disease multiclass detection through deep learning models and post-processing heuristics, *Comput. Methods Biomech. Biomed. Eng. Imaging Vis. Taylor & Francis*, 2024.
- Goyal P., Rani R., Singh K. A multilayered framework for diagnosis and classification of Alzheimer's disease using transfer learned Alexnet and LSTM, *Neural Comput. Appl. Springer Science and Business Media Deutschland GmbH*, 2024, Vol. 36, № 7, pp. 3777–3801.
- Assmi A., Elhabyb K., Benba A. et al. Alzheimer's disease classification: a comprehensive study, *Multimed. Tools Appl. Springer*, 2024, Vol. 83, № 27, pp. 70193–70216.
- Jolad B., Rao M., Kamdi S. et al. Alzheimer stages prediction using swinnet for segmentation and transfer learning based CATNet approach with SBO algorithm, *Multiscale Multidiscip. Model. Exp. Des. Springer Science and Business Media B. V.*, 2025, Vol. 8, № 2, pp. 1–16.
- Thamizhvani T. R., Hemalatha R. J. Classification of the Different Stages of Alzheimer's Disease Using 3d Transfer Learning Networks, *Research Square*, 2024, pp. 1–19.
- Ali M., Kim K., Khalid M. et al. Enhancing Alzheimer's disease diagnosis and staging: a multistage CNN framework using MRI, *Front. Psychiatry. Frontiers Media SA*, 2024, Vol. 15, P. 1395563.
- Savaş S., Damar C. Transfer-learning-based classification of pathological brain magnetic resonance images, *ETRI J. John Wiley and Sons Inc*, 2024, Vol. 46, № 2, pp. 263–276.
- Ozdemir C., Dogan Y. Advancing early diagnosis of Alzheimer's disease with next-generation deep learning methods, *Biomed. Signal Process. Control. Elsevier*, 2024, Vol. 96, P. 106614.
- Suchitra S., Krishnasamy L., Poovaraghan R. J. A deep learning-based early alzheimer's disease detection using magnetic resonance images, *Multimed. Tools Appl. Springer*, 2024, pp. 1–22.
- Fathi S., Ahmadi A., Dehnad A. et al. A Deep Learning-Based Ensemble Method for Early Diagnosis of Alzheimer's Disease using MRI Images, *Neuroinformatics. Springer*, 2024, Vol. 22, № 1, pp. 89–105.
- Bigham B., Zamanpour S. A., Zare H. Features of the superficial white matter as biomarkers for the detection of Alzheimer's disease and mild cognitive impairment: A diffusion tensor imaging study, *Heliyon. Elsevier*, 2022, Vol. 8, № 1, P. e08725.
- El-Latif A., Chelloug S., Abdulhafith M. et al. Accurate Detection of Alzheimer's Disease Using Lightweight Deep Learning Model on MRI Data, *Diagnostics*, 2023, Vol. 13, № 7, pp. 1–21.
- Sathuraman S., Malaiyappan N., Ramalling R. et al. Predicting Alzheimer's Disease Using Deep Neuro-Functional Networks with Resting-State fMRI, *Electron. 2023*, Vol. 12, P. 1031. Multidisciplinary Digital Publishing Institute, 2023, Vol. 12, № 4, P. 1031.
- Hazarika R., Maji A., Kandar D. et al. An Approach for Classification of Alzheimer's Disease Using Deep Neural Network and Brain Magnetic Resonance Imaging (MRI), *Electron*, 2023, Vol. 12, № 3.
- Tuvshinjargal B., Hwang H. VGG-C Transform Model with Batch Normalization to Predict Alzheimer's Disease through MRI Dataset, *Electron*, 2022, Vol. 11, P. 2601. Multidisciplinary Digital Publishing Institute, 2022, Vol. 11, № 16, P. 2601.
- Jraba S., Elleuch M., Ltfi H. et al. Classification of Alzheimer's Disease with Transfer Learning Using Deep Learning Models. *Lect. Notes Networks Syst. Springer, Cham*, 2024, Vol. 1046 LNNS, P. 363–374.
- Hasan M. R., Hossain A. B. M. A., Ullah S. M. A. Ensemble ResDenseNet: Alzheimer's disease staging from brain MRI using deep weighted ensemble transfer learning, *Int. J. Comput. Appl. Taylor and Francis Ltd.*, 2024, Vol. 46, № 7, pp. 539–554.
- Raza N., Naseer A., Tamoor M. et al. Alzheimer Disease Classification through Transfer Learning Approach, *Diagnostics. Multidisciplinary Digital Publishing Institute (MDPI)*, 2023, Vol. 13, № 4, P. 801.
- Ramzan F., Khan M., Rehmat A. et al. A Deep Learning Approach for Automated Diagnosis and Multi-Class Classification of Alzheimer's Disease Stages Using Resting-State fMRI and Residual Neural Networks, *J. Med. Syst. Springer*, 2020, Vol. 44, № 2, pp. 1–16.
- Alsubaie M. G., Luo S. Alzheimer's Disease Detection Using Deep Learning on Neuroimaging: A Systematic Review, *Mach. Learn. Knowl. Extr.*, 2024, Vol. 6, P. 464–505. Multidisciplinary Digital Publishing Institute, 2024, Vol. 6, № 1, pp. 464–505.

21. Mirzaei G., Adeli H. Machine learning techniques for diagnosis of alzheimer disease, mild cognitive disorder, and other types of dementia, *Biomed. Signal Process. Control. Elsevier*, 2022, Vol. 72, P. 103293.
22. Huang H., Zheng S., Yang Z. et al. Voxel-based morphometry and a deep learning model for the diagnosis of early Alzheimer's disease based on cerebral gray matter changes, *Cereb. Cortex. Oxford Academic*, 2023, Vol. 33, № 3, pp. 754–763.
23. Jung E., Luna M., Park S. H. Conditional GAN with 3D discriminator for MRI generation of Alzheimer's disease progression, *Pattern Recognit. Pergamon*, 2023, Vol. 133, P. 109061.
24. Mohammed E. M., Fakhrudeen A. M., Alani O. Y. Detection of Alzheimer's disease using deep learning models: A systematic literature review, *Informatics Med. Unlocked. Elsevier*, 2024, Vol. 50, P. 101551.
25. Reddy C. K., Rangarajan A., Rangarajan D. et al. A Transfer Learning Approach: Early Prediction of Alzheimer's Disease on US Healthy Aging Dataset, *Mathematics*, 2024, Vol. 12, № 14, pp. 1–32.
26. Mahmud T., Barua K., Habiba S. et al. An Explainable AI Paradigm for Alzheimer's Diagnosis Using Deep Transfer Learning, *Diagnostics*, 2024, Vol. 14, P. 345. Multidisciplinary Digital Publishing Institute, 2024, Vol. 14, № 3, P. 345.
27. Ramani R., Ganesh S., Rao S. et al. Integrated Multi-Modal 3D-CNN and RNN Approach with Transfer Learning for Early Detection of Alzheimer's Disease, *Iran. J. Sci. Technol.- Trans. Electr. Eng. Springer Nature*, 2024, Vol. 49, № 1, pp. 383–407.
28. Mahmood A., Shahid F., Khan R. et al. Utilizing Siamese 4D-AlzNet and Transfer Learning to Identify Stages of Alzheimer's Disease, *Neuroscience. Pergamon*, 2024, Vol. 545, pp. 69–85.
29. Nehar A., Bellaouar S., Benyoub N. et al. Alzheimer's Disease Detection using Transfer Learning, *Proc. – 2024 Int. Conf. Telecommun. Intell. Syst. ICTIS 2024. Institute of Electrical and Electronics Engineers Inc.*, 2024.
30. Pandey P., Pruthi J., Alzahrani S. et al. Enhancing healthcare recommendation: transfer learning in deep convolutional neural networks for Alzheimer disease detection, *Front. Med. Frontiers Media SA*, 2024, Vol. 11, P. 1445325.
31. Khaleel Z. S., Lakizadeh A. Early Diagnosis of Alzheimer's Disease with Transfer Learning Techniques Via ResNet50 and FSBI-LSTM, *Inform.*, 2025, Vol. 49, № 11, pp. 149–164.
32. Ali M., Hussain S., Khalid M. et al. MRI-Driven Alzheimer's Disease Diagnosis Using Deep Network Fusion and Optimal Selection of Feature, *Bioeng.*, 2024, Vol. 11, P. 1076. Multidisciplinary Digital Publishing Institute, 2024, Vol. 11, № 11, P. 1076.

Received 13.07.2025.

Accepted 20.10.2025.

УДК 004.93

ПРОГНОЗУВАННЯ ХВОРОБИ АЛЬЦГЕЙМЕРА ЗА ДОПОМОГОЮ ГЛИБОКО СКЛАДОВОЇ АНСАМБЛЕВОЇ МОДЕЛІ, ПОСИЛЕНОЇ МЕХАНІЗМОМ УВАГИ СТИСНЕННЯ ТА ЗБУДЖЕННЯ

Мухаммад Джунаїд Ікбал – аспірант, факультет обчислювальної техніки, Університет Гомалі, Дера Ісмаїл Хан, Пакистан.

Мудасір Махмуд – доцент, факультет обчислювальної техніки, Університет Гомалі, Дера Ісмаїл Хан, Пакистан.

Мухаммад Фархан – викладач, факультет обчислювальної техніки, Університет Гомалі, Дера Ісмаїл Хан, Пакистан.

Мухаммад Саджад Хуссейн – викладач, факультет обчислювальної техніки, Університет Гомалі, Дера Ісмаїл Хан, Пакистан.

АНОТАЦІЯ

Актуальність. Хвороба Альцгеймера (ХА) – це прогресуюче неврологічне дегенеративне захворювання, що спричиняє втрату пам'яті, порушення когнітивних функцій та деменцію. Своєчасне виявлення ХА має вирішальне значення для забезпечення ефективного лікування та втручання. Магнітно-резонансна томографія (МРТ) також стала критично важливим інструментом для розуміння структурних змін у мозку, що відбуваються під час розвитку хвороби Альцгеймера. Тим не менш, ручна обробка МРТ-сканів є трудомісткою, суб'єктивною та схильною до людських помилок. Як наслідок, зростає попит на автоматизовані та точні діагностичні технології, які можуть допомогти клініцистам у ранньому виявленні та стадіювання хвороби Альцгеймера на основі даних медичної візуалізації.

Мета. Це дослідження зосереджено на розробці та оцінці багатослової ансамблевої моделі на основі глибокого навчання для класифікації та стадіювання МРТ-сканувань головного мозку при хворобі Альцгеймера. Основною метою є підвищення точності та надійності діагностики шляхом поєднання сильних сторін кількох попередньо навчених архітектур згорток нейронних мереж (CNN) у поєднанні зі складними механізмами уваги та методами метанавчання.

Метод. Запропонований підхід використовує структуру глибокого багатослового ансамблю навчання, що складається з трьох високопродуктивних архітектур CNN: MobileNetV2, ResNet50 та DenseNet121. Ці моделі попередньо навчені на наборі даних ImageNet, використовуючи потужні можливості витягу ознак. Для подальшого покращення їхньої продуктивності кожна модель CNN покращена модулем уваги стиснення та збудження (SE), який адаптивно перекалібрує реакції на ознаки по каналах, підкреслюючи важливі ознаки, водночас пригнічуючи нерелевантні. Виділені високорівневі ознаки з усіх трьох SE-доповнених CNN потім об'єднуються та подаються в мета-навчальну систему, що складається з повністю зв'язаних шарів. Цей мета-класифікатор включає методи відсіву та пакетної нормалізації, щоб запобігти перенавчанню та покращити узагальнення. Загальна архітектура навчається та перевіряється на наборі даних зображень МРТ мозку, класифікованих за різними стадіями хвороби Альцгеймера, включаючи нормальний контроль, легкі когнітивні порушення та різні стадії деменції.

Результати. Експериментальна оцінка продемонструвала виняткову продуктивність, досягнувши точності 99%, прецизійності 99%, повноти 98% та F1-оцінки 99%. Ці показники вказують на високу прогностичну здатність та надійність моделі у розрізненні різних стадій хвороби Альцгеймера.

Висновки. Експериментальні результати підкреслюють ефективність та надійність запропонованої моделі глибоко укладеного ансамблю в автоматизованій діагностиці та стадіювання хвороби Альцгеймера за допомогою МРТ-сканувань. Інтеграція кількох згорткових нейронних мереж (CNN) з механізмами уваги та метанавчання значно покращує ефективність класифікації. Ці результати свідчать про те, що модель може служити надійною системою підтримки рішень для неврологів, допомагаючи у ранній діагностиці, своєчасному втручанні та покращенні результатів лікування пацієнтів у клінічних умовах.

КЛЮЧОВІ СЛОВА: згорткові нейронні мережі, глибоке навчання, хвороба Альцгеймера, стиснення та збудження, глибоко укладений ансамбль.

ЛІТЕРАТУРА

1. Rahbani R. G. Al., Alzheimer's disease multiclass detection through deep learning models and post-processing heuristics /

R. G. AI Rahbani, A. Ioannou, T. Wang // Comput. Methods Biomech. Biomed. Eng. Imaging Vis. Taylor & Francis. – 2024.

2. Goyal P. A multilayered framework for diagnosis and classification of Alzheimer's disease using transfer learned Alexnet and LSTM / P. Goyal, R. Rani, K. Singh // *Neural Comput. Appl.* Springer Science and Business Media Deutschland GmbH. – 2024. – Vol. 36, № 7. – P. 3777–3801.
3. Alzheimer's disease classification: a comprehensive study/[A. Assmi, K. Elhabyb, A. Benba et al.] // *Multimed. Tools Appl.* Springer. – 2024. – Vol. 83, № 27. – P. 70193–70216.
4. Alzheimer stages prediction using swinnet for segmentation and transfer learning based CATNet approach with SBO algorithm/[B. Jolad, M. Rao, S. Kamdi et al.] // *Multiscale Multidiscip. Model. Exp. Des.* Springer Science and Business Media B.V. – 2025. – Vol. 8, № 2. – P. 1–16.
5. Thamizhvani T.R. Classification of the Different Stages of Alzheimer's Disease Using 3d Transfer Learning Networks /T. R. Thamizhvani, R. J. Hemalatha // *Research Square*. – 2024. – P. 1–19.
6. Enhancing Alzheimer's disease diagnosis and staging: a multitask CNN framework using MRI / [M. Ali, K. Kim, M. Khalid et al.] // *Front. Psychiatry.* Frontiers Media SA. – 2024. – Vol. 15. – P. 1395563.
7. Savaş S. Transfer-learning-based classification of pathological brain magnetic resonance images/S. Savas, C. Damar // *ETRI J.* John Wiley and Sons Inc. – 2024. – Vol. 46, № 2. – P. 263–276.
8. Ozdemir C. Advancing early diagnosis of Alzheimer's disease with next-generation deep learning methods / C. Ozdemir, Y. Dogan // *Biomed. Signal Process. Control.* Elsevier. – 2024. – Vol. 96. – P. 106614.
9. Suchitra S. A deep learning-based early alzheimer's disease detection using magnetic resonance images / S. Suchitra, L. Krishnasamy, R. J. Poovaraghan // *Multimed. Tools Appl.* Springer. – 2024. – P. 1–22.
10. A Deep Learning-Based Ensemble Method for Early Diagnosis of Alzheimer's Disease using MRI Images/[S. Fathi, A. Ahmadi, A. Dehnad et al.] // *Neuroinformatics.* Springer, 2024. Vol. 22, № 1. P. 89–105.
11. Bigham B. Features of the superficial white matter as biomarkers for the detection of Alzheimer's disease and mild cognitive impairment: A diffusion tensor imaging study / B. Bigham, S. A. Zamanpour, H. Zare // *Heliyon.* Elsevier. – 2022. – Vol. 8, № 1. – P. e08725.
12. Accurate Detection of Alzheimer's Disease Using Lightweight Deep Learning Model on MRI Data / [A. El-Latif, S. Chelloug, M. Abdulhafith et al.] // *Diagnostics*. – 2023. – Vol. 13, № 7. – P. 1–21.
13. Predicting Alzheimer's Disease Using Deep Neuro-Functional Networks with Resting-State fMRI / [S. Sathuraman, N. Malaiyappan, R. Ramaling et al.] // *Electron*. – 2023. – Vol. 12. – P. 1031. Multidisciplinary Digital Publishing Institute, 2023. –Vol. 12, № 4. – P. 1031.
14. An Approach for Classification of Alzheimer's Disease Using Deep Neural Network and Brain Magnetic Resonance Imaging (MRI)/ [R. Hazarika, A. Maji, D. Kandar et al.] // *Electron*. – 2023. – Vol. 12, № 3.
15. Tuvshinjargal B. VGG-C Transform Model with Batch Normalization to Predict Alzheimer's Disease through MRI Dataset / B. Tuvshinjargal, H. Hwang // *Electron*. – 2022. – Vol. 11. – P. 2601. Multidisciplinary Digital Publishing Institute. – 2022. – Vol. 11, № 16. – P. 2601.
16. Classification of Alzheimer's Disease with Transfer Learning Using Deep Learning Models / [S. Jraba, M. Elleuch, H. Lufi et al.] // *Lect. Notes Networks Syst.* Springer, Cham. – 2024. – Vol. 1046 LNNS. – P. 363–374.
17. Hasan M. R. Ensemble ResDenseNet: Alzheimer's disease staging from brain MRI using deep weighted ensemble transfer learning / M. R. Hasan, A. B. M. A. Hossain, S. M. A. Ullah // *Int. J. Comput. Appl.* Taylor and Francis Ltd. – 2024. – Vol. 46, № 7. – P. 539–554.
18. Alzheimer Disease Classification through Transfer Learning Approach/ [N. Raza, A. Naseer, M. Tamoor et al.] // *Diagnostics*. Multidisciplinary Digital Publishing Institute (MDPI). – 2023. – Vol. 13, № 4. – P. 801.
19. A Deep Learning Approach for Automated Diagnosis and Multi-Class Classification of Alzheimer's Disease Stages Using Resting-State fMRI and Residual Neural Networks/ [F. Ramzan, M. Khan, A. Rehmat et al.] // *J. Med. Syst.* Springer. – 2020. – Vol. 44, № 2. – P. 1–16.
20. Alsubaie M.G. Alzheimer's Disease Detection Using Deep Learning on Neuroimaging: A Systematic Review / M. G. Alsubaie, S. Luo // *Mach. Learn. Knowl. Extr.* – 2024. – Vol. 6. – P. 464–505. Multidisciplinary Digital Publishing Institute. – 2024. –Vol. 6, № 1. – P. 464–505.
21. Mirzaei G. Machine learning techniques for diagnosis of alzheimer disease, mild cognitive disorder, and other types of dementia / G. Mirzaei, H. Adeli // *Biomed. Signal Process. Control.* Elsevier. – 2022. – Vol. 72. – P. 103293.
22. Voxel-based morphometry and a deep learning model for the diagnosis of early Alzheimer's disease based on cerebral gray matter changes/ [H. Huang, S. Zheng, Z. Yang et al.] // *Cereb. Cortex.* Oxford Academic. – 2023. – Vol. 33, № 3. – P. 754–763.
23. Jung E. Conditional GAN with 3D discriminator for MRI generation of Alzheimer's disease progression / E. Jung, M. Luna, S. H. Park // *Pattern Recognit.* Pergamon. – 2023. – Vol. 133. – P. 109061.
24. Mohammed E. M. Detection of Alzheimer's disease using deep learning models: A systematic literature review / E. M. Mohammed, A. M. Fakhrudeen, O. Y. Alani // *Informatics Med. Unlocked.* Elsevier. – 2024. –Vol. 50. – P. 101551.
25. A Transfer Learning Approach: Early Prediction of Alzheimer's Disease on US Healthy Aging Dataset/ [K. Reddy C, A. Rangarajan, D. Rangarajan et al.] // *Mathematics*. – 2024. – Vol. 12, № 14. – P. 1–32.
26. An Explainable AI Paradigm for Alzheimer's Diagnosis Using Deep Transfer Learning/ [T. Mahmud, K. Barua, S. Habiba et al.] // *Diagnostics*. – 2024. – Vol. 14, P. 345. Multidisciplinary Digital Publishing Institute. – 2024. – Vol. 14, № 3. – P. 345.
27. Integrated Multi-Modal 3D-CNN and RNN Approach with Transfer Learning for Early Detection of Alzheimer's Disease/ [R. Ramani, S. Ganesh, S. Rao et al.] // *Iran. J. Sci. Technol. - Trans. Electr. Eng.* Springer Nature. – 2024. –Vol. 49, № 1. – P. 383–407.
28. Utilizing Siamese 4D-AlzNet and Transfer Learning to Identify Stages of Alzheimer's Disease/ [A. Mahmood, F. Shahid, R. Khan et al.] // *Neuroscience.* Pergamon. – 2024. – Vol. 545. – P. 69–85.
29. Alzheimer's Disease Detection using Transfer Learning/ [A. Nehar, S. Bellaouar, N. Benyoub et al.] // *Proc. - 2024 Int. Conf. Telecommun. Intell. Syst. ICTIS 2024.* Institute of Electrical and Electronics Engineers Inc., 2024.
30. Enhancing healthcare recommendation: transfer learning in deep convolutional neural networks for Alzheimer disease detection/ [P. Pandey, J. Pruthi, S. Alzahrani et al.] // *Front. Med.* Frontiers Media SA. – 2024. – Vol. 11. – P. 1445325.
31. Khaleel Z. S. Early Diagnosis of Alzheimer's Disease with Transfer Learning Techniques Via ResNet50 and FSBi-LSTM/ Z. S. Khaleel, A. Lakizadeh // *Inform*. – 2025. – Vol. 49, № 11. – P. 149–164.
32. MRI-Driven Alzheimer's Disease Diagnosis Using Deep Network Fusion and Optimal Selection of Feature / [M. Ali, S. Hussain, M. Khalid et al.] // *Bioeng*. – 2024. – Vol. 11. – P. 1076. Multidisciplinary Digital Publishing Institute. – 2024. – Vol. 11, № 11. – P. 1076.

THE “PRISM AND RAYS METHOD” FOR RESEARCHING SUBJECTIVIZATION OF PERCEPTION OF HUMAN-MACHINE INTERACTION OBJECTS

Pukach A. I. – PhD, Assistant of Automated Control Systems Department, Institute of Computer Sciences and Informational Technologies, Lviv Polytechnic National University, Lviv, Ukraine.

Teslyuk V. M. – Dr. Sc., Professor, Head of Automated Control Systems Department, Institute of Computer Sciences and Informational Technologies, Lviv Polytechnic National University, Lviv, Ukraine.

ABSTRACT

Context. The task of representation formalizing as well as possibility of further research of the scientific and applied problem of perception subjectivization of the human-machine interaction objects by relevant interaction subjects, in the context of global problems of automation and intellectualization of the component processes and environments of human-machine interaction, is considered, investigated, and resolved in scope of current research. **The object of research** are processes of perception subjectivization of the human-machine interaction objects by relevant subjects of same interaction. **The subject of research** are methods and means of mathematical, computer, and simulation modeling. **Objective** – development of a method for researching the perception subjectivization of the human-machine interaction objects by relevant subjects of the same interaction.

Method. The development of the “prism and rays method” (author’s name of the method) is proposed and performed, which provides the possibility to resolve the scientific and applied problem of representation formalizing as well as possibility of further research into the processes of perception subjectivization of the human-machine interaction objects by relevant interaction subjects (of the same interaction).

Results. The results of the developed method – are corresponding models that represent and allow to investigate the researched processes of perception subjectivization of the human-machine interaction objects by relevant interaction subjects (of the same interaction). The developed method provides the possibility of both formalization as well as further interpretation and investigation of the researched processes of perception subjectivization of human-machine interaction objects. As a practical approbation, the developed method has been applied for synthesis of the basic model of perception subjectivization of the object of complex (comprehensive) support (one of the most highlighted examples of HMI/HCI) of software product(s), using the example case of an experimental task of estimating the approximate processing time of client request by the department of customer-and-technical support of the given software product.

Conclusions. The developed “prisms and rays method” solves the declared task of representation formalizing as well as possibility of further research of the scientific and applied problem of perception subjectivization of the human-machine interaction objects by relevant interaction subjects, in the context of global problems of automation and intellectualization of the component processes and environments of human-machine interaction. At the same time, the obtained results of experimental practical approbation of the developed method confirm its effectiveness and efficiency in the context of solving relevant practical applied tasks of the scientific and applied problem of human-machine interaction objects’ perception subjectivization.

KEYWORDS: human-machine interaction, object of interaction, perception subjectivization, impact factors, automation, intellectualization.

ABBREVIATIONS

AAR is an After Action Review;
AI is an Artificial Intelligence;
ApHT is an Approximate Handle Time;
ANN is an Artificial Neural Networks;
EI is an Emotional Intelligence;
HCI is a Human-Computer Interaction;
HMI is a Human-Machine Interaction;
ML is a Machine Learning;
MP is a Multilayer Perceptron;
PRM is a “Prisms and Rays Method”.

NOMENCLATURE

$Chr_{[i]}$ is an i -th (original) characteristic of the researched object;

$Chr_{[i]}^{[j-1]}$ is an i -th characteristic of the object at the entrance $(j-1)$ -th prism of perception;

$Chr_{[i]}^{[j]}$ is an i -th characteristic of the object at the entrance j -th prism of perception;

$Chr_{[io]}^{[1]}$ is an io -th original characteristic of the researched object at entrance of the 1-st perception prism;

$Chr_{[noc]}^{[1]}$ is a noc -th original characteristic of the object at the entrance of the 1-st perception prism;

$Fpr_{[1]}()$ is a transformation function of the 1-st perception prism of the existing prism system;

$Fpr_{[j]}()$ is a transformation function of the j -th prism of perception;

$Fpr_{[npr]}()$ is a transformation function of the npr -th (the last one, in case there are npr prisms in current prism system configuration) perception prism of the existing prism system;

$Fpr_{[j]}.RAc()$ is a subfunction of the j -th prism's transformation function, which returns the refraction angle value of current prism depending on the value of the argument (current characteristic of the researched object);

$Fpr_{[j]}.BWc()$ is a subfunction of the j -th prism's transformation function, which returns the bandwidth value of current prism depending on the value of the argument (current characteristic of the researched object);

$McPfcSDp()$ is a complex subjectivization function (of the human-machine interaction object);

$Obj[Chr_{[1]...[noc]}]$ is a set of objective (true, original) characteristics of the researched human-machine interaction object, where noc – count of these characteristics;

$SbP_{[i]}$ is an i -th (subjective) characteristic of the perception (subjective) of the researched object;

$SbP_{[is]}$ is an is -th subjective characteristic of the perception of researched object (which corresponds to the io -th original characteristic of the same researched object);

$Subj[SbP_{[1]...[nss]}]$ is a set of subjective (relative, modified) characteristics of the researched human-machine interaction object, where nss – count of these characteristics.

INTRODUCTION

One of the relevant areas of research nowadays is HMI/HCI, which is a complex concept that combines an extremely large variety of derivative problems, areas, and tasks. With the active introduction of AI and ML technologies, some of the problems of HMI/HCI have gained new opportunities in terms of conceptual approaches, methods, and means of research and resolution. Despite that, there are still a fairly significant number of relevant scientific and applied issues in this area which require both developing new and revision of existing methods and means for solving them. One of such actual issues is the task of representation formalizing (with the possibility of further research) of the scientific and applied problem of perception subjectivization of the HMI/HCI objects by relevant interaction subjects, in the context of global problems of automation and intellectualization of the component processes and environments of HMI/HCI. Since perception subjectivization problematic is extremely important, as it allows to understand the reasons of (both existing and potential) “incompatibilities” of the subjects – participants of the HMI/HCI, as well as, in general, subjects of any other intersubjective interactions, including the “human-human” class as well.

The object of research are processes of perception subjectivization of the human-machine interaction objects by relevant subjects of same interaction. **The subject of research** are methods and means of mathematical, computer, and simulation modeling.

The objective of the research consists in the development of a method for studying the processes of human-

machine interaction objects' perception subjectivization by the relevant interaction subjects, which will provide the possibility of solving the declared task of representation formalizing (as well as possibility of further research) of the scientific and applied problem of perception subjectivization of the human-machine interaction objects by relevant interaction subjects, in the context of global problems of automation and intellectualization of the component processes and environments of HMI/HCI.

1 PROBLEM STATEMENT

Let's consider the formalization of the problem of human-machine interaction object(s)' perception subjectivization in the form of a problem of multi-criteria multifunctional cascade dependence. In such case, the input variables of the problem are set of objective (true, original) characteristics of the researched human-machine interaction object: $Obj[Chr_{[1]}, Chr_{[2]}, ..., Chr_{[noc]}]$.

While the output (i.e., resulting) variables of the problem are set of subjective (relative, modified) characteristics of the same researched object, but already within the framework of its perception by the subject(s), in the context of their existing human-machine (and/or computer, as well as, in general – any other intersubjective) interaction: $Subj[SbP_{[1]}, SbP_{[2]}, ..., SbP_{[nss]}]$.

Let's consider a set of multi-criteria multifunctional cascade dependencies (given by some complex subjectivization function, named as $McPfcSDp()$, representing the processes of perception subjectivization of the researched object (of interaction) by the relevant subjects (of same interaction):

$$Subj[SbP_{[1]...[nss]}] = McPfcSDp(Obj[Chr_{[1]...[noc]}]). \quad (1)$$

The main necessary criterion of the problem is the finiteness of corresponding sets of both objective and subjective characteristics (of the researched object), as well as the previously agreed and defined dimension of the multi-criteria multifunctional cascade dependence existing between these sets.

Problem limitations:

1. The complex subjectivization function $McPfcSDp()$ should provide the “many-to-many” relation type/scheme between the sets of both objective and subjective characteristics (of the researched object), because only such relation type/scheme corresponds to the proposed methodology for HMI/HCI objects' perception subjectivization within the framework of current research (while the “one-to-one” or “one-to-many” relation types/schemas doesn't correspond the declared principles of the proposed methodology).

Expression (1) provides the possibility of representation formalizing of perception subjectivization of the researched object(s) (of HMI/HCI, as well as any intersubjective interaction in general).

However, such format provides an exclusively basic generalized representation. Thus, there arise a need for

further, more deeper disclosure of the subjectivization function itself, which would provide the possibility of not only pure formalization, but also further research into the complex processes of the perception subjectivization of any object(s) in the context of both HMI/HCI, as well as any intersubjective interaction in general as the most universalized approach.

The purpose of the research is to highlight the developed “prisms and rays method” (which is author’s name of the method) for studying the processes of HMI/HCI objects’ perception subjectivization, intended to solve the declared task of representation formalizing as well as possibility of further research of the scientific and applied problem of perception subjectivization of the human-machine interaction objects by relevant interaction subjects, in the context of global problems of automation and intellectualization of the component processes and environments of HMI/HCI.

2 REVIEW OF THE LITERATURE

HMI/HCI contains an extremely wide variety of different existing aspects and problems. For example, a recent study [1] presented a sociolinguistic model of HMI which provides opportunity for studying interaction of technological capabilities, user cognitive awareness, and language strategies, and is based on these three continua: technological capabilities, user cognitive awareness, and language strategies. At the same time, the author of another recent study [2] analyzed the evolution of HCI in the era of AI, examining the path of AI in the context of this interaction from competitor to partner, emphasizing the importance of preserving a human will in this partnership between men and machines.

While, research [3] considers some aspects of modern achievements in HMI, AI, and robotics. Meanwhile, authors of another research [4] conducted a comprehensive study in the context of the issues of building relationships in HMI, namely, in the direction of confirming the concept of the influence of self-disclosure on mutual understanding in coaching chatbots using psychophysiological measurements. In another research [5], the author conducted an ethnographic study which confirms the obvious shortcomings in emotional interaction and social feedback in the context of HMI, and therefore: the ability of machines to understand emotions and cope/manage with complex scenarios still requires further development, improvement, and refinement.

The author of the work [6] studied the physiological and psychological characteristics and limitations of humans in human-machine systems, as well as the rules of human-human and human-machine interaction, applying the results of fundamental research to the development of human-machine systems in order to obtain safe, convenient, and effective human-machine systems.

The team of authors of the work [7] investigates the issue of probabilistic human-machine cooperation in the context of product personalization, proposing a user interaction paradigm for recommendation and configuration systems based on appropriate Pass Bayesian Reasoning

and Suitability Probability Tables approaches. In work [8], the author explores the interaction between humans and AI using AI-artifacts that can understand the relevant properties of their users (e.g., states, skills, needs) and accordingly personalize the interaction in a way that preserves transparency, user control, and trust.

The authors of research [9] conducted a comprehensive review of the current state of the issue of EI, which includes a review of main methods for measuring EI, as well as the problems, which EI faces with.

At the same time, authors of research [10] substantiated the need to improve adaptation mechanisms in human-machine systems, including: by using a functional networks to solve ergonomic design problems, as well as by combining functional and neural networks, which, in aggregate, provides the properties of adaptive human-machine interaction.

So, as it can be seen from the conducted review of existing researches, the issue of HMI/HCI is extremely relevant and actual nowadays, and covers a whole range of various aspects, tasks, issues and problems, most of which have already been sufficiently well-disclosed and researched, in particular, for example, the issue of HMI / HCI personalization.

However, in the context of HMI/HCI object(s)’ perception subjectivization, there is still a significant gap among the conducted researches. In turn, this leads to the emergence of a relevant task of representation formalizing as well as possibility of further research of the scientific and applied problem of perception subjectivization of the human-machine interaction objects by relevant interaction subjects, in the context of global problems of automation and intellectualization of the component processes and environments of HMI/HCI.

3 MATERIALS AND METHODS

The main concept of the proposed “prisms and rays method” (for researching the HMI/HCI objects’ perception subjectivization) is to represent the complex subjectivization function as a set of prisms (which, in turn, are representing the existing defined impact factors, influencing the perception subjectivization of the object – by relevant subjects of interaction with this object), which appropriately modify the rays (which emanate from an imaginary “light source” and pass through the object, representing its characteristics), refracting those rays (due to prisms’ refraction angle) and changing their intensity (due to prisms’ bandwidth), and forming at the output of the prism system spectral characteristics that corresponds to the subjectivized perception of the researched object – by each relevant subject (who acts as an “observer” of the researched object).

Thus, the complex spectral characteristic of the object of observation, passing through the perception prisms’ system, is changed (modified, or “subjectivized”) accordingly by this system, receiving at the output new values (of this complex spectral characteristic) that reflect the subjectivized perception of the object of observation – by the appropriate relevant subject (observer).

At the same time, each spectral component of the incoming “beam” of rays (at the exit of observation object) necessarily and mandatory (without exception) passes through each of the perception prisms.

As for the perception prisms themselves, they have two basic characteristics/properties, such as:

1. The bandwidth of the perception prism, which is responsible for the loss/gain coefficient of the intensity of each individual (depending on its current spectrum) spectral component (entering the prism), as a result of which, at the output of this prism this spectral component receives a new intensity value;

2. The angle of refraction, which is responsible for changing the value of the spectrum of the input spectral characteristic itself, as a result of which, at the output of this prism – this spectral characteristic receives a new spectral value, which, in turn, directly affects its further processing by the next prism of the existing cascade of perception prisms.

Thus, each individual spectral component of the incoming “beam” of rays (representing the original “objective” characteristics of the object) passes through each of the prisms of perception, enters the input of each separate prism with the input values of its current spectrum and intensity, and exits the same prism with completely new values of its updated spectrum and intensity, with which it, in turn, enters the input of the next prism of perception, and so on, until it exits the last prism (of the entire existing system of perception prisms), representing the corresponding subjective perception (subjective interpretation) of these individual specific original characteristics of the object, turning them from “objective” to “subjective”.

Accordingly, the resulting subjective perception of the object is represented by a set of subjective characteristics, the dominant one of which (i.e. the one whose output ray’s intensity is maximal among all other rays of the remaining resulting subjective characteristics’ “beam”) actually represents this resulting subjective perception of the object. So, the more rays from the perception prism system fall into a specific output subjective characteristic, and the greater is the intensity of these rays at the output of the prism system, – the greater will be, accordingly, the share of this specific characteristic in the resulting subjective perception of the researched object. In turn, this feature provides the possibility of using the proposed “prisms and rays method” for solving the classification problem(s) as well.

Figure 1 below represents a graphical interpretation of corresponding simulation model of the proposed “prism and rays method”.

It is important to note that the number of input characteristics of the researched object does not necessarily have to coincide with the number of output characteristics of the subjective perception of the same object, because the whole essence of subjective perception is that any subjective perception of an object is not, directly, this object itself. At the same time, the existing spectral components of the researched object(s)’ characteristics can either merge with each other or be divided into subcomponents due to the dispersion property (if such possibility is provided by the existing prism system, specific prisms of this system, as well as clear conditions for such a scenario).

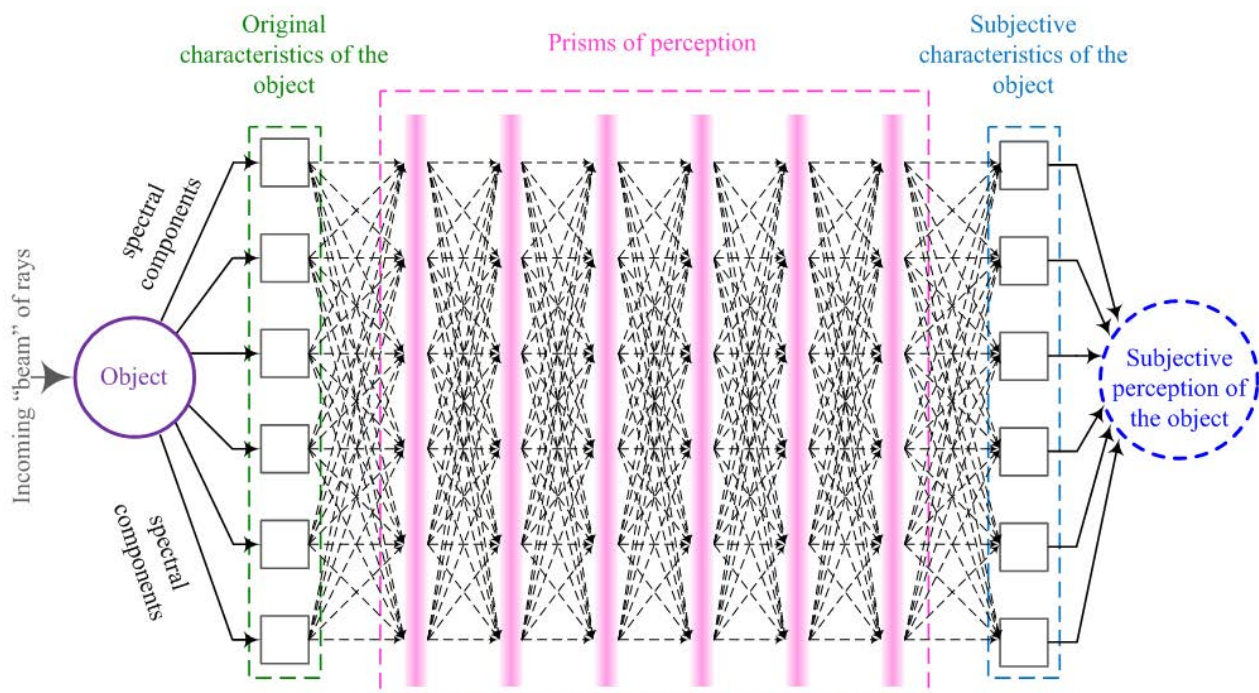


Figure 1 – Graphical interpretation of the simulation model of proposed “prism and rays method”

However, it should also be noted that the practical implementation of such a scenario requires additional refinement (improvement) precisely in the context of the most detailed coordination of all possible scenarios and conditions of both the merging of characteristics as well as their dispersion. That is why, within the framework of current research, the “classical” option is considered, according to which the number of output characteristics of the object’s perception is equal to the number of input characteristics of the same object. So, in this “classical” version/variant, each output characteristic of the object’s perception could represent, in essence, a subjective perception of a certain input original characteristic of the same object. While the whole subjective perception of the object itself is, actually, a set of subjective perceptions of each of its original (objective) characteristics.

Thus, according to the developed simulation model of the proposed “prism and rays method”, the system of perception prisms, in fact, plays the role of the complex perception function.

The main feature of the developed simulation model, as well as PRM in general, is the possibility of combining separate prisms in a prism system (in accordance with various scenarios, situations, or conditions), as well as the possibility of forming and preparing certain “templates” of sets (or systems) of prisms that correspond to some “classic” (known or frequently repeated) scenarios, situations, or conditions.

The general strategy for positioning: both the input original characteristics of the research object and the prisms of perception in the prism system, as well as the impact factors within the corresponding prisms, consists in the sequential placement of each of these elements according to their decreasing priority within their groups, that is: from the highest priority at the initial position of the corresponding group – to the lowest priority at the final position of the same group. While the priorities themselves, in turn, may depend on the specific objectives of researching the investigated processes of subjectivization. At the same time, activation of each individual impact factor from the pre-defined set of declared impact factors within their prisms of perception is carried out in accordance with a specific scenario(s) or so-called situational case(s).

Therefore, the next stage of implementation of the proposed PRM is the development of an appropriate basic mathematical model of perception subjectivization of the HMI/HCI object(s).

Expression (2) represents the moment of passage of a separate component characteristic’s ray (of the researched object) through one separate specific prism of perception:

$$Chr_{[i]}^{[j]} = Fpr_{[j]}(Chr_{[i]}^{[j-1]}). \quad (2)$$

At the same time, each individual transformation function works on the basis of the principle laid down by the above-mentioned structural and functional feature of the

corresponding prism of perception itself, relevant to this transformation function.

This feature is common for absolutely all prisms of perception, and lies in the basic properties of the prisms: the bandwidth and the refraction angle.

Thus, the operation principle of absolutely all transformation functions, which detail the function(s) itself, can be described by expression (3) below:

$$Fpr_{[j]}(Chr_{[i]}^{[j-1]}) = \begin{cases} Fpr_{[j]}.BWd(Chr_{[i]}^{[j-1]}) \times Chr_{[i]}^{[j-1]}; \\ Fpr_{[j]}.RAc(Chr_{[i]}^{[j-1]}) \times Chr_{[i]}^{[j-1]}. \end{cases} \quad (3)$$

In the context of expression (3), it is important to note that the basic properties of prisms (namely: the bandwidth and the refraction angle) act precisely as functions, the arguments of which are the input characteristic(s) of the researched perception object. Thus, both coefficients (the bandwidths and the prism refraction angle) are not static constants, but the dynamic ones, which provide greater flexibility and adaptability of the proposed solution and PRM-method in general.

Accordingly, any (arbitrary) resulting subjective characteristic at the output of the model, which represents a component characteristic of the corresponding subjective perception of the researched object, is, in essence, the result of performing a cascade of transformation functions applied sequentially (by each of the perception prisms) in relation to a certain input (separately taken) original characteristic of this object.

Therefore, expression (4) below reflects the described regularity pattern:

$$SbP_{[is]} = Fpr_{[npr]} \left(Fpr_{[npr-1]} \left(\dots \left(Fpr_{[1]} \left(Chr_{[io]}^{[1]} \right) \right) \right) \right), \quad (4)$$

where: npr – number of perception prisms of the existing prism system.

Thus, the possibility of declaring a complex subjectivization function of the researched HMI/HCI object(s), represented by expression (5) below, is provided:

$$McPfCsDp = \begin{cases} Fpr_{[npr]} \left(\dots \left(Fpr_{[1]} \left(Chr_{[1]}^{[1]} \right) \right) \right); \\ Fpr_{[npr]} \left(\dots \left(Fpr_{[1]} \left(Chr_{[2]}^{[1]} \right) \right) \right); \\ \vdots \\ Fpr_{[npr]} \left(\dots \left(Fpr_{[1]} \left(Chr_{[noc]}^{[1]} \right) \right) \right), \end{cases} \quad (5)$$

where noc – the number of objective (true, original) characteristics of the researched HMI/HCI object.

A key feature of the developed basic mathematical model is its complete correspondence to the relevant form of declared multi-criteria multifunctional cascade dependence.

The final stage of implementation of the declared “prism and rays method” is the development of an appropriate algorithm of perception of HMI/HCI objects’ perception subjectivization, which fully corresponds to the previously developed simulation model, as well as the relevant basic mathematical model.

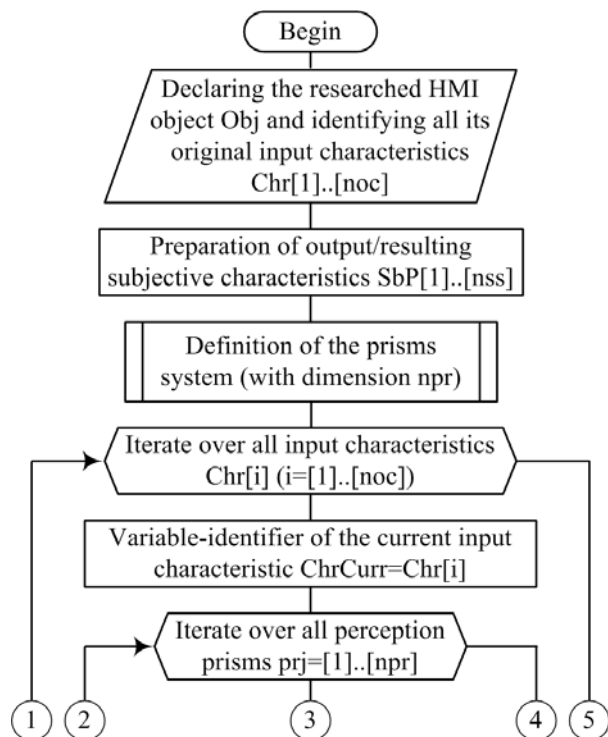


Figure 2 below represents a flowchart of the developed algorithm of HMI/HCI objects’ perception subjectivization in the context of declared PRM.

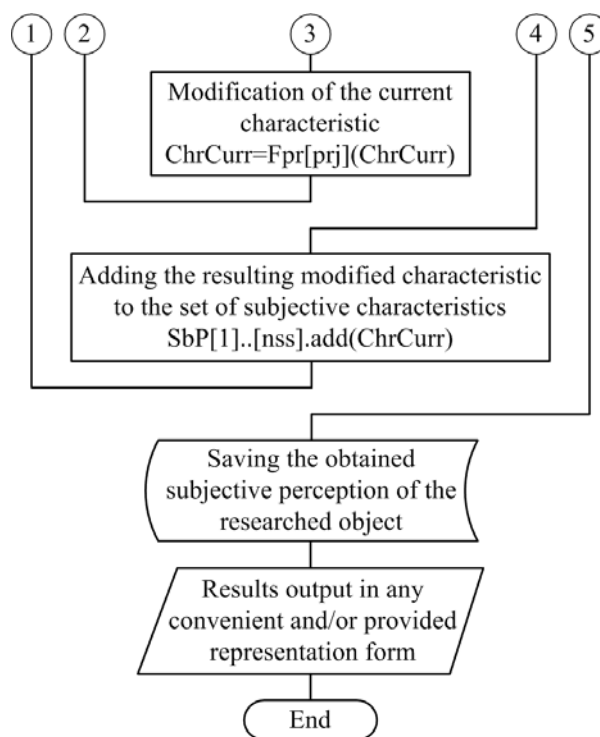


Figure 2 – Flowchart of the developed algorithm of HMI/HCI objects’ perception subjectivization in the context of declared PRM

Thus, the implementation process of the declared PRM (for researching the perception subjectivization of the objects of HMI) has been completely finalized.

In addition, it should be emphasized that the prisms of perception are constructively synthesized from the relevant impact factors, which, in an appropriate way(s), affect the perception subjectivization of any investigated object(s) – by the existing relevant subject(s) of interaction with this object(s).

In turn, such a constructive feature of the fundamental-conceptual structure of the prisms of perception provides an additional, extremely relevant, feature of the developed PRM, which consists in its capabilities for fundamental and comprehensive research of both: directly the prisms of perception themselves, as well as the relevant impact factors which, actually, form/construct these prisms of perception. Because, it is the impact factors that are the basic elements that forms the individual personalized characteristics (of objects’ perception) of each separate subject (interacting with this object).

Therefore, the key to understanding the individuality of perception of any subject lies precisely in the study / research of the impact factors which form individual (and in many cases, even unique and unrepeatable) personal-

ized systems of perception prisms of each separate individual subject.

4 EXPERIMENTS

The experiment consists in the step-by-step execution of all stages of the specialized developed algorithm of the declared PRM, as well as in the usage of the developed models (both simulation and basic mathematical ones) for investigating the perception subjectivization of the experimental researched object of the investigated HMI / HCI.

At the same time, it is extremely important to determine (as accurately as possible) the original characteristics of the researched object, as well as the system of perception prisms in the context of the specific vector of perception that will be investigated in any separate/specific individual case of research.

In addition, it is equally important to clearly identify the subjects of perception that are closely related to the prism system, in order to (if necessary) further adjust the prism system in the context of the selected subject(s) of performed investigation/research.

5 RESULTS

The main results obtained by the developed PRM – are the corresponding simulation models that provide the possibility of initial formalization of any problem of HMI/HCI object(s)' perception subjectivization, with the subsequent possibility of a more detailed investigation of the component processes of subjectivization of each individual characteristic(s) of the researched object(s).

In particular, as a practical approbation of the developed PRM, a relevant scientific and applied problem of synthesizing a basic model of perception subjectivization of an object of software product's comprehensive support has been resolved, using the example of an experimental case of estimating the approximate processing time for a client's request – by the customer-and-technical support department of pre-defined software product.

Comprehensive support of software products is one of the most highlighted examples of HMI/HCI, and includes such mandatory components as:

- an object of interaction, or an object of support (which is the supported software product itself, or the processes related to its comprehensive support);
- interaction/support subjects (which interact with the support object, and directly provide and implement the components of its comprehensive support);
- as well as the impact factors that affect/influence the perception subjectivization of the support object – by relevant existing subjects, which, in turn, directly affects executing by subjects their functional duties regarding the comprehensive support of the supported object, as well as the efficiency and the effectiveness of the relevant components of such comprehensive support.

As a specific example of an experimental case, the task of evaluating the ApHT indicator of a client's request, received by the customer-and-technical support service of a specific pre-defined supported software product, by the employees of this service has been considered in scope of current research.

The approximate time of client's requests processing by the customer-and-technical support service department of software development companies – is an extremely relevant and actual practical applied problem, as it provides the possibility(-ies) to rationalize resources, as well as prevent the occurrence(s) of any problematic situations related to the potential delays in processing client's requests. Unfortunately, there is still no any "one-size-fits-all"-solution/mechanism for assessing this ApHT indicator, since it involves a lot of human factors (some of which will be presented, in particular, in the relevant list(s) of provided prism factors described below in the text), so most existing solutions are based primarily on the empirical (i.e., heuristic) experience of relevant professionals and specialists (i.e., experts) in this subject area.

Thus, in this particular specific considered investigated case: the researched "object" (of a defined software product's comprehensive support) – will be, in fact, the client's request (namely: its true ApHT indicator); the subjects – will be employees, members of the customer-

and-technical support team (of the defined supported software product); the input characteristics of the object – will be the data obtained from the request; and the output characteristics of the resulting perception of the researched object – will be, respectively, subjective ApHT indicator(s).

The resolution of the given experimental exemplary practical applied problem depends on several consecutive stages, namely:

- determining the input characteristics of the object (based on the data/information from obtained client's request);
- identification of the perception prisms and the impact factors;
- declaration of resulting characteristics (i.e. ApHT indicators).

So, the first stage in resolving the declared practical applied problem – is to determine the input characteristics of the object of perception, which is carried out on the basis of data/information contained directly in the obtained client's request, since this is the only existing source of objective information at this stage/phase of investigation in the context of given task.

Thus, among the main input characteristics, following ones could be distinguished, in particular, such as:

- request's original priority set directly by the client himself;
- problematic area of the supported product (system, subsystem, area, environment, node, interface, etc.);
- type of request (not working, delay in operation/functioning, unstable operating/functioning, incorrect operating/functioning, how exactly should it work, can it work differently, need it to work differently, etc.);
- scale of the request (classified based on a general/total number of affected users);
- receipt date and time of the request (to identify whether it is "business hours" or not);
- availability of all necessary (according to previously agreed/defined request-filling templates) information/data from the client at this stage of entering the request into the customer-and-technical support service department/team.

In turn, among the main prisms of perception and impact factors, in the context of solving the given experimental practical applied problem, the following ones could be highlighted:

1. Professional prism, where the relevant impact factors could be:

- 1.1. employee's current position (within current place of work);
- 1.2. employee's experience in scope of the current project;
- 1.3. employee's experience in scope of the current company;
- 1.4. employee's experience at the relative global market;
- 1.5. employee's experience working with the given problematic area of the product;
- 1.6. employee's experience working with the given specific type of request;

1.7. presence of any specific features of working with this separate given client;

1.8. experience working with this separate given client and/or its specifics/features;

1.9. having experience working with similar requests;

1.10. having the necessary/needed level of general education (eruditeness, intelligence, specialized education, certification, etc.).

2. Organizational prism, where the relevant impact factors could be:

2.1. the reliability of the information/data contained in the request;

2.2. availability and serviceability of access to the client's environment (if required);

2.3. day of the week and time of the day (to identify whether it is "business hours" or not);

2.4. availability and serviceability of all necessary hardware;

2.5. availability and serviceability of all necessary software;

2.6. availability and serviceability of all the material and technical equipment necessary for the organization of the workplace;

2.7. skills for effective use of all needed and available hardware;

2.8. skills for effective use of all needed and available software;

2.9. skills for effective use of all needed and available material and/or technical equipment;

2.10. availability and completeness of all necessary technical documentation (functional requirements, knowledge base, requests archive, frequently asked questions, etc.);

2.11. skills for effective use of all needed and available technical documentation;

2.12. availability and accessibility of all necessary teams (of the development company) and their relevant and necessary representatives;

2.13. availability and completeness of all necessary protocol (procedural, as well as situational) documentation;

2.14. skills for effective use of all needed and available protocol (procedural, as well as situational) documentation;

2.15. the presence of clear criteria for assessing the quality of work, and the complete/full familiarity of the customer-and-technical support department's employee with these criteria.

3. Motivational prism, where the relevant impact factors could be:

3.1. personal intangible level of motivation of the given employee of the customer-and-technical support department/team (to what extent the resolution of this separate/specific client's request/problem is a "personal challenge" for this employee, without the presence of any other additional material and/or non-material incentive);

3.2. the employee's sense of its personal image and reputation;

3.3. the employee's personal well-being (in all aspects);

3.4. the employee's favorability level of any interaction conflictedness (since the presence of any conflicts between subjects negatively affects the level of their motivation, as well as the level of their interaction quality, efficiency and effectiveness);

3.5. personal, intangible level of motivation of the employees of other departments/teams involved in processing the current given client's request;

3.6. availability of material incentives (bonuses, allowances, etc.);

3.7. availability of corporate/organizational incentives (rating system, opportunity for promotion and career growth, etc.);

3.8. the presence of a positive reaction, feedback, approval from the client;

3.9. presence of a "team spirit", ability and desire to work in a team;

3.10. the employee's sense of corporate image and reputation;

3.11. presence of a punitive mechanism for poor performance of functional duties (financial penalties, demotion, dismissal, etc.).

4. Communication prism, where the relevant impact factors could be:

4.1. completeness and clarity (i.e., quality) of the information contained directly in the obtained client's request itself;

4.2. level of proficiency in the language of the client/request (since very often nowadays the companies of customers/users, the software development companies, and the customer-and-technical support departments/teams may be located in completely different countries, even overseas);

4.3. openness to communication, understanding of its importance, and absence of individual / personal communication complexes and/or problems;

4.4. client/user availability for communication;

4.5. availability of necessary members of other involved teams to enable communication with (and between) them;

4.6. availability and accessibility of all necessary channels and means of communication;

4.7. skills in using existing channels and means of communication (both for "external" and "internal" communication);

4.8. personal communication characteristics of the given employee of the customer-and-technical support department/team;

4.9. personal communication characteristics of the client/user;

4.10. personal communication characteristics of the rest members of other involved teams of the development company.

It is also important to note that each of the impact factors can bring both negative and positive connotation to the perception subjectivization of the researched object, depending on the context and prerequisites.

That is why, in order to simplify the understanding of the exact connotation of each separate impact factor, in practice, when designing/synthesizing basic simulation models (of HMI/HCI objects' perception subjectivization), it is recommended to use not only single impact factor(s), but instead – a pairs of the “protagonist-antagonist” type for each separate impact factor, which will ensure the most unambiguous understanding of the connotation of this impact factor during simulation modeling of almost any arbitrary investigated situational case(s).

The final stage of solving the declared experimental practical applied problem – is declaration of the output / resulting characteristics (i.e. ApHT indicators).

To do this, the ApHT indicator(s) are presented in the format of a compliance coefficient(s) of the resolution time of obtained client's request (with given type / class / priority), regulated by the relevant customer-and-technical-support contract/agreement concluded with each separate client.

In this case, the following possible values of the output/resulting characteristics of the subjective perception of the ApHT assessment process (by a relevant employee(s) of the customer-and-technical support department/team of the supported software product development company) based on the received client's request have been proposed/highlighted in scope of resolving the declared practical applied problem:

- <0.5 , which means that the obtained client's request will be processed more than twice faster then the expected processing time;
- $[0.5-1]$ – from expected time to twice faster;
- $[1-1.5]$ – from the expected time to one and a half times longer;
- $[1.5-2]$ – from one and a half times longer than expected – to twice longer then expected;
- >2 – more than twice longer then expected.

Thus, the identification of all necessary components, carried out at the previous stages of solving the given practical applied problem, makes it possible to obtain the necessary basic model of perception subjectivization of the object of software product's comprehensive support, using the example of an experimental case of estimating the approximate processing time of client's request by the customer-and-technical support department/team employee(s) of a defined software product.

In particular, the corresponding simulation model is presented below, at Figure 3.

The components of each of the perception prisms (marked at Figure 3 by numbered elements 1.1–1.10, 2.1–2.15, 3.1–3.11, 4.1–4.10) correspond to the same positions in the above-provided list of impact factors that are part of the corresponding four perception prisms (namely: professional, organizational, motivational, and communication), given above in the text, just as the input and the output characteristics correspond to the same items in the appropriate above-provided list(s) of elements of the cor-

responding stages of solving the given practical applied problem, described above as well.

In fact, in this case, the developed simulation model operates in a classification mode, which is also available for the developed PRM, where the resulting/output perception of the object (i.e., the subjective ApHT indicator) is presented in the format of the probabilities of each of the categories (the resulting output characteristics). Accordingly, the category (the resulting output subjective characteristic) that will gain the maximal value among all others – will represent the result of the subjective perception of the researched object (in this case, the ApHT indicator).

In particular, at Figure 4, provided below, there are also examples of few specific experimental cases of simulation modeling performed on the basis of a synthesized simulation model (of this experimental research on the assessment of ApHT indicator), which represent specific situational cases (scenarios) based on the analysis of some previously processed archived client requests (of this specific given researched supported software product, provided for investigation and resolution of this separate specific experimental practical applied problem).

The main objective of this experiment is to investigate the possibility of using the developed “prisms and rays method” and, in particular, the basic simulation models of perception subjectivization of the researched object(s) (in this particular case – the ApHT indicator, as such researched object) synthesized by means of the developed PRM, – as an environment for modeling procedures of the AAR-methodology class, since this methodology is extremely effective for a better understanding of the researched processes and their cause-and-effect relationships, which, in turn, is quite relevant in the context of investigating the relevant issues of perception subjectivization of the objects of HMI/HCI.

As examples of modeling, four specific situational cases of processing client's requests and subjective forecasting of the ApHT indicator (which was later compared with the real one, after the request's resolution has been completely finalized), have been presented.

Each of these 4 experimental cases, provided below at Figure 4, reflects precisely those key impact factors (within the relevant perception prisms), established through the AAR-interview with the subject, which, in fact, played a major determining and/or decisive role in the subject's decision-making regarding the expected (by this subject) value of the ApHT indicator when processing this specific case.

The obtained results of modeling provided experimental cases of the investigated practical applied problem (of client's request(s) ApHT estimation) confirm the effectiveness of the relevant simulation models synthesized by means of the developed PRM in providing opportunities and possibilities for modeling the processes of perception subjectivization of the objects (as well as the processes) of HMI/HCI – by the relevant existing subjects of this interaction.

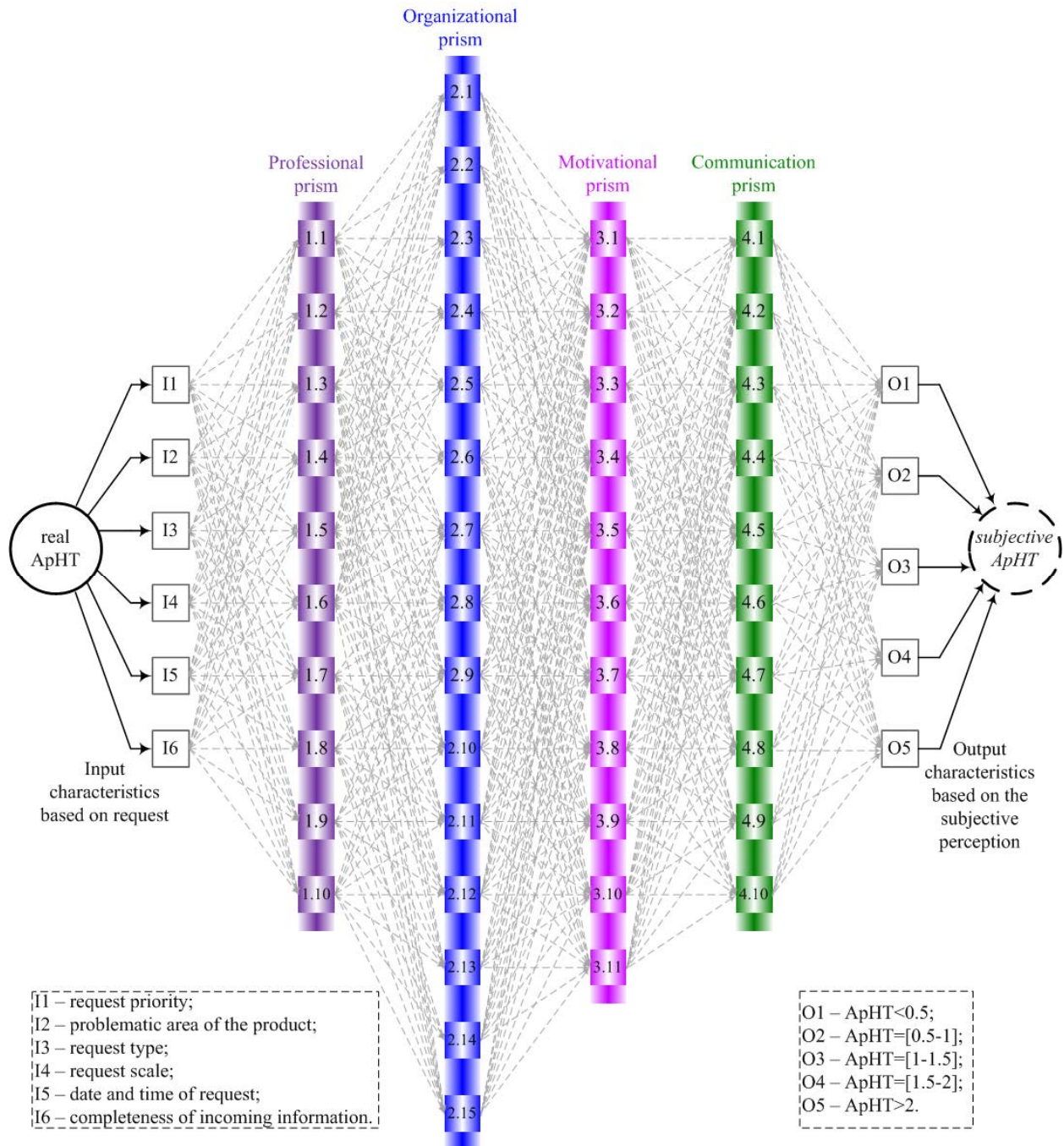


Figure 3 – An example of a synthesized basic simulation model of perception/estimation subjectivization of the approximate processing time for a client's request by the customer-and-technical support department/team employee of the supported software product

The developed simulation model (of the object's perception subjectivization) is a fundamental component, or a kind of "coordinate system" for a specific investigated problem, which provides the possibility of further complex research into the processes of perception subjectivization of HMI/HCI objects in scope of this specific "co-ordinate system".

In particular, it provides possibilities for modeling the dynamics and tendency of perception subjectivization of researched HMI/HCI object – by the same investigated subject, changing only the parametric indicators (i.e.,

bandwidth and refraction angle) of the perception prisms for specific impact factors.

In this way, by selecting different experimental values of the parametric indicators (i.e., bandwidth and refraction angle) of the component perception prisms, it is possible to achieve certain optimal results in which the resulting subjective perception of the researched object (for example, investigated subjective ApHT indicator) will be as close as possible to the real object (i.e., real ApHT obtained already from the results of final processing of appropriate client's request).

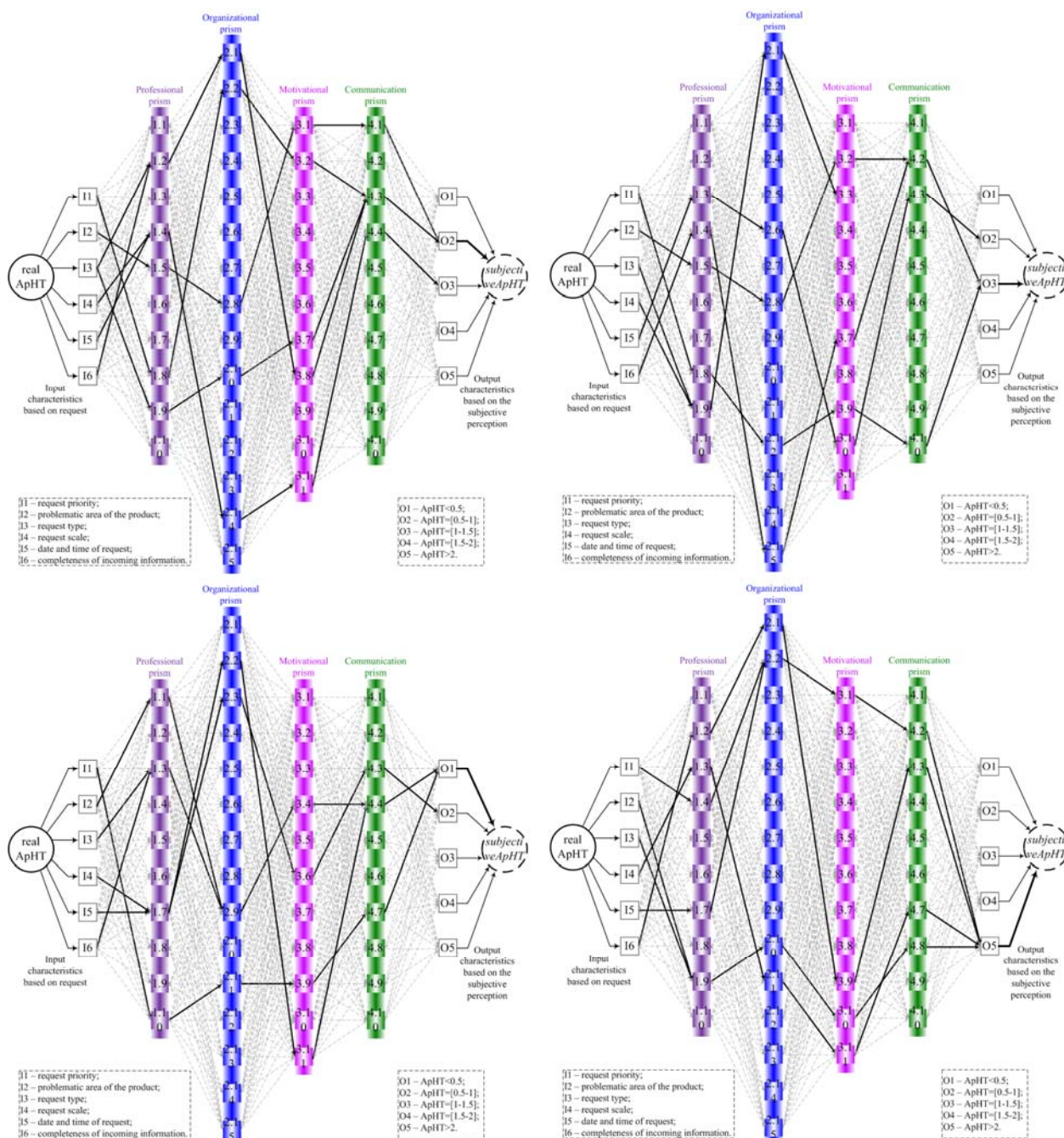


Figure 4 – Examples of experimental cases of simulation modeling performed on the basis of a generalized subjective perception model synthesized by the developed PRM

Thus, the developed “prisms and rays method” (for researching the HMI/HCI objects’ perception subjectivization) has been practically approbated on the example of solving an experimental case of relevant practical applied problem of estimating the approximate processing time for a client’s request by the customer-and-technical support department/team employee(s) of a relevant supported software product. In turn, the obtained approbation results confirm the effectiveness and efficiency of the developed “prisms and rays method” for researching both the scientific and applied problem of HMI/HCI object(s)’ perception subjectivization, as well as the relevant derived prac-

tical applied problems, both in the field of human-machine interaction as well as in the context of any inter-subjective interaction in general.

Besides that, the developed method provides additional possibilities for a comprehensive research of the impact factors that directly affect the subjectivization of the perception of HMI/HCI objects – by relevant subjects of the same interaction.

6 DISCUSSION

In research [11] authors explore the role of digital twin technology in many aspects of designing, modeling, and

optimizing complex systems, as well as the possibilities of integrating this technology into simulation, data analytics, machine learning, the Internet of Things, networks, and human-machine interaction. In particular, as an example of such integration, the human-machine interaction model of an intelligent digital twin presented in [12] is given, which improves user interaction by visualizing big data in a way that makes it easy to understand and process, including through the usage of new interaction methodologies, thanks to which the user has full control over the data flow, using gaze, gestures and a voice recognition interface that will provide contextual information.

The research presented in work [13] is devoted to the study of the problems of human perception based on text information, where the authors carried out a comparative analysis of methods like Decision Tree, K nearest neighbors, Naive Bayes, and generalized linear models in the context of human perception analysis based on text mining using a precisely selected data set(s) of customer support discussions at one of the social networks. As a result of the comparative analysis, it was found that the decision tree algorithm is the best, followed by Naive Bayes.

At the same time, the work [14] investigates the issue of personalized individual perception of AI chatbots by users through the mediating role of the categories of usefulness and ease of use of the latter, based on cross-temporal measurement of the relevant indicators using a specialized questionnaire, using, among other things, partial least squares structural equation modeling, as a result of which a significant positive direct relationship between the quality of an AI chatbot and its perception by a human user during their human-machine/computer interaction was confirmed, and the results of the indirect relationship show that the expected usefulness and ease of use partially mediated the relationship between the quality of AI chatbots and their perception by the users (human).

While, in the work [15], modern aspects of human-machine interaction in the field of systems engineering are investigated, due to the active intervention into this area of ML, AI, autonomous systems, and the issue of how these technologies change whole this activity field, positioning people not only as end users, but also as critical components in extensive software ecosystems, exploring the complex interaction between technical and human factors that affect the mutual perception of subjects (human and machines) on both sides of their joint human-machine interaction.

Thus, existing researches confirm the relevance of the researched issues and the significant both pure scientific as well as scientific-practical interest of scientists, researchers and professionals in the field of HMI/HCI and a large number of its constituent components and derived scientific-applied problems and tasks, one of which is, in fact, the issue of perception subjectivization of the HMI/HCI objects by the relevant subjects of this interaction.

At the same time, the research of this specific component of HMI/HCI requires additional actualization and updating, since there is a certain lack of attention of re-

searchers (scientists, and/or professionals) to this component, which, nevertheless, is one of the basic, fundamental and conceptual components of HMI/HCI, since it provides the possibility of formalizing the fundamental basis of the manifestations of any further HMI/HCI problems and issues, the root cause of which, in a global sense, is, among other things, different perception subjectivization of the same HMI/HCI object by various subjects (of this interaction) as all these subjects could perceive this same object in completely different ways, which is due to the individual properties of each subject to personalize and subjectivize their own perception of the objects (both HMI/HCI objects, as well as any other objects in general case).

In turn, the developed "prisms and rays method" provides needed possibility(-ies), carrying out both preliminary formalization of the researched processes of perception subjectivization of HMI/HCI objects, as well as the possibility of further processing, researching, and modeling of these processes, thereby filling the existing gap in a scope of fundamental understanding of the root causes of an extremely large number of derivative HMI/HCI problems, caused, precisely, by differences in the individual personalized perception(s) of the same HMI/HCI objects by different subjects.

As a prospect for further research and practical application of the developed MPR, authors see the potential of its use in solving both scientific-applied and practical-applied problems of automation and intellectualization of software products' comprehensive support, where the perception subjectivization of the support objects plays one of the key roles both in the context of the quality of customer support services (and/or any other parts of comprehensive support) and in the context of the effectiveness of intersubjective interaction of those subjects who directly implement and provide such comprehensive support of software products.

CONCLUSIONS

The "prism and rays method" (author's name of the method) has been developed to study the perception subjectivization of the objects of HMI/HCI. The main scientific and applied problem solved by the developed method is the problem of representation formalizing as well as possibility of further research of the scientific and applied problem of perception subjectivization of the HMI/HCI objects by relevant interaction subjects, in the context of global problems of automation and intellectualization of the component processes and environments of HMI/HCI. The input data of the developed method are the original (i.e. objective) characteristics of the researched object of perception. While, the output results of the developed method is the subjective perception of the researched object, represented by the subjectivized values of its objective characteristics.

The main feature of the developed method is its ability to conduct fundamental and comprehensive research not only on the processes of interaction objects' perception subjectivization, but also on the perception prisms them-

selves, and therefore on relevant impact factors which form the subjective personalized perception and vision of the same objects by different subjects of interaction.

A simulation model (of HMI/HCI objects' perception subjectivization) has been developed, key feature of which is the possibility of combining individual prisms in a relevant prism system in accordance to various scenarios, situations, or conditions, as well as the possibility of forming and preparing certain "templates" of prisms sets (or systems) which correspond to some "classic" (known or frequently repeated) scenarios, situations, or conditions.

Also, a basic mathematical model (of HMI/HCI objects' perception subjectivization) has been developed, the main feature of which is its complete correspondence to the declared form of the relevant multi-criteria multifunctional cascade dependence. In addition, a specialized algorithm of HMI/HCI objects' perception subjectivization has been developed, which provides possibility of further automation (including computer modeling) of the researched processes.

The scientific novelty consists in development of the "prism and rays method" (author's name of the method) for researching the perception subjectivization of HMI/HCI objects, which provides possibility of solving the declared scientific and applied task of representation formalizing as well as possibility of further research of the scientific and applied problem of perception subjectivization of the HMI/HCI objects by relevant interaction subjects, in the context of global problems of automation and intellectualization of the component processes and environments of HMI/HCI.

The practical significance consists in development of: a simulation model, a basic mathematical model, and the algorithm of HMI/HCI objects' perception subjectivization processes, which provide the possibility of comprehensive research and modeling of these processes, as well as lay the foundations for ensuring the possibilities of further computer/software implementation and modeling of these processes by developing (on their basis) appropriate dedicated specialized software.

Prospects for further research consist of two key areas of potential improvement and expansion of application area of the developed "prism and rays method". In particular, the first direction is the encapsulation of ANN, including, possibly, a MP, especially at the initial stage, taking into account the constructive similarity of the developed simulation model (of objects' perception subjectivization) with the structure of the MP itself, but not only limited to it, but also considering and investigating the prospects for encapsulating other neural networks – into the simulation models, synthesized by the developed PRM, which will provide opportunities and possibilities for increasing the level of automation and intellectualization of the scientific and applied potential of the developed method. While the second direction will provide the possibility of research and further disclosure of the entire inherent scientific and applied potential of the developed method, not being limited exclusively to a narrow direc-

tion of research in the context of only the HMI/HCI objects' perception subjectivization, but instead – investigating the potential of the developed PRM in the context of the perception subjectivization of any objects in any environments of any intersubjective interaction, such as, for example: socionics, conflictology, psychology, and an extremely wide variety of other relevant areas of intersubjective interaction, both in the context of relevant scientific and applied as well as derivative practical applied problems.

ACKNOWLEDGEMENTS

This research is proactive. It was carried out as a part of the scientific activity of the authors outside of the working hours at their main positions.

REFERENCES

1. Lotze N. Human-Machine Interaction as a Complex Socio-Linguistic Practice, *Voice Assistants in Private Homes*, 2024, pp. 105–144. DOI: 10.1515/9783839472002-005
2. Jain N. The evolution of human-computer interaction in the AI era, *International Journal of Research in Computer Applications and Information Technology*, 2025, Vol. 8, Iss. 1, pp. 144–151. DOI: 10.34218/ijrcait_08_01_015
3. Solanes J. E., Gracia L., Valls Miro J. Advances in Human-Machine Interaction, *Artificial Intelligence, and Robotics. Electronics*, 2024, Vol. 13, Iss. 19, pp. 1–5. DOI: 10.3390/electronics13193856
4. Mai V., Savanevska N., Richert A. et al. Relationship building in human-machine interaction: A proof of concept on the influence of self-disclosure on rapport in chatbot coaching using psychophysiological measurements, *Chatbot coaching, Hauptartikel*, 2024, Vol. 16, pp. 1–27. DOI: 10.25598/JKM/2024-16.2
5. Guo S. H. Ethnographic Research On Human-Computer Interaction Behavior In The Intelligent Era, *World Journal of Management Science*, 2024, Vol. 2, Iss. 3, pp. 57–61. – DOI: 10.61784/wms3035
6. Guozhen Z. Human Ergonomics, *The ECPH Encyclopedia of Psychology*, 2024, pp. 1–4. DOI: 10.1007/978-981-99-6000-2_774-1
7. Dangelmaier M., Hölzle K., Krieg S. et al. Probabilistic Human-Machine Cooperation In Product Personalization [Electronic resource], *11th International Conference on Customization and Personalization MCP 2024 The Power of Customization and Personalization in the Digital Age*, 2024. pp. 97–101. Mode of access: https://mcp-ce.org/wp-content/uploads/2024/11/15_Dangelmaier.pdf (date of access: 01.07.2025). Title from screen.
8. Conati C. AI-Driven Personalization to Support Human-AI Collaboration, *In Companion Proceedings of the 16th ACM SIGCHI Symposium on Engineering Interactive Computing Systems (EICS '24 Companion)*, 2024, pp. 5–6. – DOI: 10.1145/3660515.3661324
9. Verma A., Yadav C. S., Sharma N. et al. Emotional Intelligence in Machine Interaction, *Advances in Computational Intelligence and Robotics*, 2024, pp. 173–189. DOI: 10.4018/979-8-3693-6806-0.ch009

10. Lavrov E., Siryk O., Kirichenko I. et al. The Methodology of Managed Functional Networks for Organizing Effective and Adaptive Human-Machine Dialogue in Automated Systems [Electronic resource], *International Conference on Information and Communication Technologies in Education, Research, and Industrial Applications, CEUR Workshop Proceedings*, 2021, Vol. 3013, pp. 428–437. Mode of access: <https://ceur-ws.org/Vol-3013/20210428.pdf> (date of access: 01.07.2025). – Title from screen.
11. Adekola F., Folasade K., Olujimi A. et al. Digital Twin Technology: A Tool in Computer Engineering Task – A Review, *American Journal of Computer Engineering*, 2024, Vol. 12, Iss. 10, pp. 1–7. Mode of access: https://www.researchgate.net/publication/388955822_Digital_Twin_Technology_A_Tool_in_Computer_Engineering_Task_-_A_Review (date of access: 01.07.2025). Title from screen.
12. Dingli A., Haddod F. Interacting with Intelligent Digital Twins, *Lecture Notes in Computer Science*, 2019, Vol. 11584, pp. 3–15. DOI: 10.1007/978-3-030-23541-3_1
13. Hoque A., Munmun F. Y., Ifte S. I. et al. Human Perception Based on Textual Analysis, *International Journal of Intelligent Systems and Applications*, 2024, Vol. 16, Iss. 6, pp. 84–93. – DOI: 10.5815/ijisa.2024.06.05
14. Abdallah W., Harraf A., Mosusa O. et al. Investigating Factors Impacting Customer Acceptance of Artificial Intelligence Chatbot: Banking Sector of Kuwait, *International Journal of Applied Research in Management and Economics*, 2022, Vol. 5, Iss. 4, pp. 45–58. DOI: 10.33422/ijarme.v5i4.961
15. Abrahão S., Grundy J., Pezzè M. et al. Software Engineering by and for Humans in an AI Era, *ACM Transactions on Software Engineering and Methodology*, 2025, pp. 1–44. DOI: 10.1145/3715111

Received 31.07.2025.

Accepted 15.10.2025.

УДК 004.5

МЕТОД «ПРИЗМ ТА ПРОМЕНІВ» ДЛЯ ДОСЛІДЖЕННЯ СУБ'ЄКТИВІЗАЦІЇ СПРИЙНЯТТЯ ОБ'ЄКТІВ ЛЮДИНО-МАШИННОЇ ВЗАЄМОДІЇ

Пукач А. І. – канд. техн. наук, асистент кафедри Автоматизованих Систем Управління Інституту Комп'ютерних Наук та Інформаційних Технологій Національного Університету «Львівська Політехніка», Львів, Україна.

Теслюк В. М. – д-р техн. наук, професор, завідувач кафедри Автоматизованих Систем Управління Інституту Комп'ютерних Наук та Інформаційних Технологій Національного Університету «Львівська Політехніка», Львів, Україна.

АНОТАЦІЯ

Актуальність. Розглянуто задачу формалізації представлення з можливістю подальшого дослідження науково-прикладної проблеми суб'єктивізації сприйняття об'єктів людино-машинної взаємодії – суб'єктами цієї взаємодії, в контексті глобальної проблематики автоматизації й інтелектуалізації складових процесів та середовищ людино-машинної взаємодії. **Об'єктом дослідження** є процеси суб'єктивізації сприйняття об'єктів людино-машинної взаємодії – суб'єктами цієї взаємодії. **Предметом дослідження** є методи та засоби математичного, комп'ютерного, а також імітаційного моделювання.

Метою роботи – розроблення методу для дослідження суб'єктивізації сприйняття об'єктів відповідними суб'єктами людино-машинної взаємодії.

Метод. Запропоновано розроблення методу «призм та променів» (авторська назва методу), що забезпечує можливість вирішення науково-прикладної задачі формалізації представлення з можливістю подальшого дослідження процесів суб'єктивізації сприйняття об'єктів людино-машинної взаємодії відповідними суб'єктами цієї взаємодії.

Результати. Результатами роботи розробленого методу є відповідні моделі, що репрезентують суб'єктивізацію сприйняття досліджуваного об'єкта людино-машинної взаємодії відповідними суб'єктами цієї взаємодії. Розроблений метод забезпечує можливість як формалізації, так і подальшої інтерпретації та дослідження процесів суб'єктивізації сприйняття об'єктів людино-машинної взаємодії. В якості апробації, розроблений метод застосовано для синтезу базової моделі суб'єктивізації сприйняття об'єкта комплексної підтримки (як одного з найбільш поширених прикладів людино-машинної взаємодії) програмних продуктів, на прикладі експериментального кейсу оцінки приблизного часу опрацювання клієнтського запиту відділом клієнтської та технічної підтримки програмного продукту.

Висновки. Розроблений метод вирішує поставлену задачу формалізації представлення з можливістю подальшого дослідження науково-прикладної проблеми суб'єктивізації сприйняття об'єктів людино-машинної взаємодії – суб'єктами цієї взаємодії, в контексті глобальної проблематики автоматизації й інтелектуалізації складових процесів та середовищ людино-машинної взаємодії. Водночас, отримані результати експериментальної апробації розробленого методу підтверджують його дієвість в контексті вирішення прикладних практичних задач науково-прикладної проблематики суб'єктивізації сприйняття об'єктів людино-машинної взаємодії.

КЛЮЧОВІ СЛОВА: людино-машинна взаємодія, об'єкт взаємодії, суб'єктивізація сприйняття, фактори впливу, автоматизація, інтелектуалізація.

ЛІТЕРАТУРА

1. Lotze N. Human-Machine Interaction as a Complex Socio-Linguistic Practice / N. Lotze // *Voice Assistants in Private Homes*. – 2024. – P. 105–144. DOI: 10.1515/9783839472002-005
2. Jain N. The evolution of human-computer interaction in the AI era / N. Jain // *International Journal of Research in Computer Applications and Information Technology*. – 2025. – Vol. 8, Iss. 1. – P. 144–151. DOI: 10.34218/ijrcait_08_01_015
3. Solanes J. E. Advances in Human-Machine Interaction / J. E. Solanes, L. Gracia, J. Valls Miro // *Artificial Intelligence, and Robotics. Electronics*. – 2024. – Vol. 13, Iss. 19. – P. 1–5. DOI: 10.3390/electronics13193856
4. Relationship building in human-machine interaction: A proof of concept on the influence of self-disclosure on rapport in chatbot coaching using psychophysiological measurements / [V. Mai, N. Savanevska, A. Richert et al.] // *Chatbot coaching, Hauptartikel*. – 2024. – Vol. 16. – P. 1–27. DOI: 10.25598/JKM/2024-16.2
5. Guo S.H. Ethnographic Research On Human-Computer Interaction Behavior In The Intelligent Era / S. H. Guo // *World Journal of Management Science*. – 2024. – Vol. 2, Iss. 3. – P. 57–61. DOI: 10.61784/wms3035
6. Guozhen Z. Human Ergonomics / Z. Guozhen // *The ECPH Encyclopedia of Psychology*. – 2024. – P. 1–4. DOI: 10.1007/978-981-99-6000-2_774-1
7. Probabilistic Human-Machine Cooperation In Product Personalization [Electronic resource] / [M. Dangelmaier, K. Hölzle, S. Krieg et al.] // *11th International Conference on Customization and Personalization MCP 2024 The Power of Customization and Personalization in the Digital Age*. – 2024. – P. 97–101. Mode of access: https://mcp-ce.org/wp-content/uploads/2024/11/15_Dangelmaier.pdf (date of access: 01.07.2025). – Title from screen.
8. Conati C. AI-Driven Personalization to Support Human-AI Collaboration / C. Conati // *In Companion Proceedings of the 16th ACM SIGCHI Symposium on Engineering Interactive Computing Systems (EICS '24 Companion)*. – 2024. – P. 5–6. DOI: 10.1145/3660515.3661324
9. Emotional Intelligence in Machine Interaction / A. Verma, C. S. Yadav, N. Sharma et al. // *Advances in Computational Intelligence and Robotics*. – 2024. – P. 173–189. DOI: 10.4018/979-8-3693-6806-0.ch009
10. The Methodology of Managed Functional Networks for Organizing Effective and Adaptive Human-Machine Dialogue in Automated Systems [Electronic resource] / [E. Lavrov, O. Siryk, I. Kirichenko et al.] // *International Conference on Information and Communication Technologies in Education, Research, and Industrial Applications, CEUR Workshop Proceedings*. – 2021. – Vol. 3013. – P. 428–437. – Mode of access: <https://ceur-ws.org/Vol-3013/20210428.pdf> (date of access: 01.07.2025). – Title from screen.
11. Digital Twin Technology: A Tool in Computer Engineering Task – A Review / [F. Adekola, K. Folasade, A. Olujimi et al.] // *American Journal of Computer Engineering*. – 2024. – Vol. 12, Iss. 10. – P. 1–7. – Mode of access: https://www.researchgate.net/publication/388955822_Digital_Twin_Technology_A_Tool_in_Computer_Engineering_Task_A_Review (date of access: 01.07.2025). – Title from screen.
12. Dingli A. Interacting with Intelligent Digital Twins / A. Dingli, F. Haddod // *Lecture Notes in Computer Science*. – 2019. – Vol. 11584. – P. 3–15. – DOI: 10.1007/978-3-030-23541-3_1
13. Human Perception Based on Textual Analysis / [A. Hoque, F. Y. Munmun, S. I. Ifte et al.] // *International Journal of Intelligent Systems and Applications*. – 2024. – Vol. 16, Iss. 6. – P. 84–93. DOI: 10.5815/ijisa.2024.06.05
14. Investigating Factors Impacting Customer Acceptance of Artificial Intelligence Chatbot: Banking Sector of Kuwait / [W. Abdallah, A. Harraf, O. Mosusa et al.] // *International Journal of Applied Research in Management and Economics*. – 2022. – Vol. 5, Iss. 4. – P. 45–58. DOI: 10.33422/ijarme.v5i4.961
15. Software Engineering by and for Humans in an AI Era / [S. Abrahão, J. Grundy, M. Pezzè et al.] // *ACM Transactions on Software Engineering and Methodology*. – 2025. – P. 1–44. DOI: 10.1145/3715111

OPTIMIZED MODEL FOR PREDICTING THE AVAILABILITY OF OBJECTS BASED ON DEEP LEARNING AND GEOSPATIAL FEATURES

Tryhuba A. M. – Dr. Sc., Professor, Head of the Department of Information Technologies, Lviv National Environmental University, Lviv, Ukraine.

Ratushny R. T. – Dr. Sc., Professor, Professor of the Department of Law and Management in the Field of Civil Protection, Lviv State University of Life Safety, Lviv, Ukraine.

Koval L. S. – Associate Professor, Department of Information Technology and Electronic Communications Systems, Lviv State University of Life Safety, Lviv, Ukraine.

Ratushnyi A. R. – Associate Professor, Department of Information Technologies and Electronic Communications Systems, Lviv State University of Life Safety, Lviv, Ukraine.

ABSTRACT

Context. Today, predicting the availability of objects in spatially distributed systems remains one of the areas of computer science that constantly attracts the attention of researchers. There are many reasons for this. There is an increase in the amount of spatial information. New types of infrastructure networks are emerging, as well as the need for rapid decision-making in changing conditions. At the same time, traditional analysis methods do not always cope with the tasks of processing multidimensional data. This is especially true when it comes to complex or unstable environments. This opens up opportunities for applying deep learning methods that demonstrate high efficiency where classical approaches fail.

Objective. The study aims to optimize the model for predicting the availability of objects in spatially distributed systems by defining an efficient deep learning architecture that uses spatial and other infrastructure features to improve prediction accuracy and generalizability.

Method. To achieve this goal, deep learning architectures were used, including feed-forward models (FNN), convolutional neural networks (CNN), and recurrent neural networks (RNN, GRU, LSTM). During the modeling, methods of data normalization, training regularization, and a comprehensive system for evaluating the accuracy of forecasts using the mean square error, mean absolute error, and coefficient of determination were used.

Results. An optimized architecture of a recurrent neural network was built for the study, which includes a combination of two recurrent layers, Dropout regularization layers, and a fully connected layer. The analysis has shown that the proposed model provides high accuracy in predicting the availability of objects, demonstrating stability over a wide range of spatial data. Comparison of actual and predicted values confirmed the effectiveness of the proposed solution.

Conclusions. The proposed approach to building an optimized deep learning model for predicting the availability of objects provides a high level of generalization and accuracy, which creates prerequisites for its use in systems of intelligent decision support in spatially distributed environments.

KEYWORDS: deep learning, forecasting, object availability, geospatial data, RNN optimization, intelligent systems.

ABBREVIATIONS

CNN is a convolutional neural network;

FNN is a feed-forward neural network;

GRU is a gate mechanism in recurrent neural networks;

LSTM is a long short-term memory as one of the artificial architectures of recurrent neural networks;

NetworkX is a Python library for graph and network research;

OpenStreetMap is an open project aimed at collecting, preserving, and distributing publicly available geospatial data;

OSMNx is a Python library that allows you to capture spatial geometries and model, design, visualize, and analyze real street networks from OpenStreetMap APIs;

RNN is a recurrent neural network.

NOMENCLATURE

Δ_0 is a gradient of a loss function calculated on the current batch;

η is a learning rate;

θ is a set of network parameters (weights and offsets of all layers);

σ is a sigmoidal function in the output layer used to obtain a probabilistic forecast of availability;

A_i is an area of the community in square kilometers;

b is a displacement vector;

b_{ij} is a presence of the j -th obstacle on the territory of the i -th community;

$\vec{b}^{(l)}$ is a displacement vector;

E is a set of edges representing road segments with specified attributes (length, type of pavement, number of lanes);

$F()$ is a model architecture;

$f(\cdot)$ is an activation function;

f_i is an activation function of hidden layers (for example, ReLU);

q_{kj} is a value of the k -th quality metric for the model, which includes indicators of accuracy, stability, and computational efficiency;

$\vec{h}^{(l-1)}$ is an output of the previous layer;

h_t is a hidden state at a given time t ;

$\vec{h}_T$ is the last state of the LSTM after processing the entire sequence;

$\vec{k}$ is a convolution kernel;

L_i is a total length of all roads within the community;

N is a number of examples in the batch;

n is a number of examples in the test sample;

P_i is a population of the i -th community, which allows to estimate the potential load on rescue forces in case of emergencies;

V is a set of vertices corresponding to the geographical coordinates of nodes (intersections, turns);

w is a vector of weight parameters;

$W^{(l)}$ is a weight matrix of the l -th layer;

w_j is a weighting factor of the corresponding obstacle on accessibility;

w_i is a weighting factor of the importance of the i -th metric, determined in accordance with the research priorities;

W_i are respectively weight matrices;

x is an initial value of the feature;

x_s is a feature vector;

x_t is an input vector;

$x_{i+m,j+n}$ is a local feature matrix;

x_{ji} is a separate j -th characteristic, which is a numerical or categorical description of the features of the i -th settlement;

$x_{\min}$ is a minimum value of the feature;

$x_{\max}$ is a maximum value of the feature;

$x_{\min}, x_{\max}$ are respectively minimum and maximum values of the feature;

$\langle x, y \rangle$ is an initial sample in the form of a set of cases;

$\langle x', y' \rangle$ is a subsample from the original set of cases;

y_i is an actual value of the availability of voluntary rescue groups on the territory of communities for the i -th settlement;

$\bar{y}$ is an average of the actual values of y_i ;

$\hat{y}_i$ is a predicted value of the availability of voluntary rescue groups on the territory of communities for the i -th settlement based on the model;

y_s is a target value.

INTRODUCTION

One of the important areas of modern informatics is the development of models for predicting the availability of objects in spatially distributed systems [1–3]. With the increasing complexity of infrastructure networks and the growing volume of geospatial data, there is a need to improve methods capable of providing accurate analysis of spatial dependencies and support for decision-making processes.

Traditional analytical approaches based on static maps or normative estimates are often not effective enough when dealing with high-dimensional data and variable spatial characteristics [4–6]. The complexity of the object environment, the heterogeneity of the distribution of features, and the dependence on numerous infrastructure and social factors pose new challenges to researchers regarding the accuracy and adaptability of forecasting models [7–10].

In this context, deep learning methods are of particular importance. The use of multi-layer neural networks, such as feed-forward networks (FNNs), convolutional neural networks (CNNs), and recurrent neural networks (RNNs, GRUs, LSTMs), opens up opportunities for building models capable of processing large amounts of spatial data and identifying complex hidden dependencies. The development of optimized architectures of such networks for predicting the availability of objects is an important step in the development of intelligent decision support systems.

Object of the study: the process of predicting the level of availability of objects in spatially distributed systems based on the analysis of geospatial and other infrastructure characteristics.

Subject of the study: models and architectures of deep neural networks, their optimization, and application for predicting the availability of objects within intelligent decision support systems.

The purpose of this work is to develop an approach to the formation of an input spatial and infrastructural data set for predicting the availability of objects in spatially distributed systems and to substantiate an optimized model for predicting the level of availability of objects based on deep learning using spatial and infrastructural features.

Tasks of the research:

- To propose an approach to optimizing the model for predicting the availability of objects in spatially distributed systems by determining an effective deep learning architecture that uses spatial and infrastructural features to improve the prediction accuracy and generalization ability;

- To generate an input dataset based on geospatial information for training deep learning models focused on predicting the level of availability of objects in space-distributed systems;

- To justify the choice of deep neural network architecture, to optimize it according to the criteria of accuracy of predicting the availability of objects, and to evaluate the possibility of using the created model in decision support systems in spatial planning.

Deep learning methods used to analyze complex and high-dimensional data allow not only to identify hidden spatial patterns, but also to create effective predictive models for solving problems related to assessing the accessibility of objects. The combination of geoinformation resources with the capabilities of deep neural networks provides an increase in forecasting

accuracy by taking into account the current state of the road network, spatial obstacles, route length, and other critical factors.

The study is aimed at substantiating the model for predicting the availability of objects based on geospatial features and deep learning, which involves the stages of building an input spatial set of characteristics, training a neural network on representative data, and further verifying the accuracy of the created model for practical use in strategic spatial analysis.

1 PROBLEM STATEMENT

Let the initial sample of geospatial and infrastructure data be given in the form of a set of precedents $\langle x, y \rangle$, where $x = \{x_s\}$, $y = \{y_s\}$, $x_s = \{x_{sj}\}$, $s = 1, 2, \dots, S$, $j = 1, 2, \dots, N$. In this case, each element of the data sample is a feature vector x_s of dimensionality N , which corresponds to the target value y_s .

The task of synthesizing a neural network model is to find a pair of $\langle F(), w \rangle$, where $F()$ – is the model architecture, which is usually set by the user or system designer, and w – is a vector of weight parameters that is configured based on the training set. Thus, the predicted value of the level of availability of objects is denoted as:

$$y_s^* = F(w, x_s),$$

and the criterion of training quality is the optimization of the error function:

$$f(F(), w, \langle x, y \rangle) \rightarrow \min.$$

An important aspect of building an effective model for predicting the level of availability of objects is also the task of forming a subsample from the original sample $\langle x, y \rangle$. This is formally reduced to finding a pair $\langle x', y' \rangle$ such that:

$$x' \subset \{x_s\}, \quad y' = \{y_s \mid x_s \in x'\}, \quad S' < S, \quad N' = N,$$

for which the value of the loss function when training on the subsample provides the best agreement with the full sample:

$$f(\langle x', y' \rangle, \langle x, y \rangle) \rightarrow \min.$$

Thus, a two-component problem arises: 1) building the architecture and parameters of the model; 2) selecting the most informative subsample for its training, which is especially important in cases with large or unevenly represented data.

2 LITERATURE REVIEW

Ensuring the availability of objects in spatially distributed systems remains one of the key issues of modern informatics and applied geographic information analytics. The uneven location of infrastructure facilities in many regions leads to a violation of the expected normative access time, which significantly affects the level of efficiency of territorial systems [11–14]. According to the researchers, the decrease in spatial accessibility is one of the factors of increasing vulnerability of territories, especially in conditions of limited resources and underdeveloped transport infrastructure.

Traditional approaches to assessing accessibility are usually based on normatively defined radii of action or administrative boundaries [15–18]. However, such methods do not take into account the actual topography of the place, the state of the road network, building density, and spatial barriers, which limit their applicability in complex and changing environments.

In recent years, the use of open geoinformation resources for building spatial analysis models has attracted considerable interest. In particular, OpenStreetMap (OSM) data are actively used to estimate the density of the road network, model routes, and identify infrastructure objects [19–21]. Thanks to the availability of automated data processing tools, such as the OSMNX and NetworkX libraries, it becomes possible to form structured feature sets that accurately reflect the real conditions of movement in space.

In parallel, there is an active development of methods for predicting the state of objects using deep learning algorithms [22–25]. Multilayer neural networks, convolutional architectures, and models based on graph structures demonstrate high efficiency in simple classification and regression tasks [26–28]. Such approaches allow for taking into account complex nonlinear relationships between spatial, infrastructural, and social characteristics.

At the same time, the integration of deep learning methods with open geospatial data to predict the level of accessibility of objects remains insufficiently developed. In particular, there is a scientific task of creating models that can comprehensively take into account both the topographic features of the territory and the parameters of transport accessibility, building density, the presence of physical obstacles, and other key features that determine the real accessibility of objects in space.

Thus, there is a scientific and applied task of creating a model capable of accurately predicting the level of accessibility of objects based on open geodata. Such a model will take into account not only the geographical distance but also the features of the road network, the presence of obstacles, building density, and other indicators necessary for making management decisions.

3 MATERIALS AND METHODS

Building a model for predicting the level of availability of objects in spatially distributed systems

requires an integrated approach that combines the stages of geospatial data processing, formation of a set of features, selection and training of deep neural networks, as well as evaluation of the accuracy of the built models and analysis of the results. The proposed research methodology is based on the use of open geoinformation resources, spatial analysis tools, and deep learning algorithms to identify patterns in multidimensional spatial data (Fig. 1).

Step 1. Collecting and forming a database of spatial and infrastructural characteristics of the territories.

The first stage involves the creation of a holistic structured data set that includes key spatial and infrastructural features of the object environment [29–31]. The purpose of this stage is to obtain the input information necessary for the construction of features used in the process of training neural prediction networks. The main source of spatial information is open data containing information about the road network, building boundaries, infrastructure facilities, and natural features of the territory.

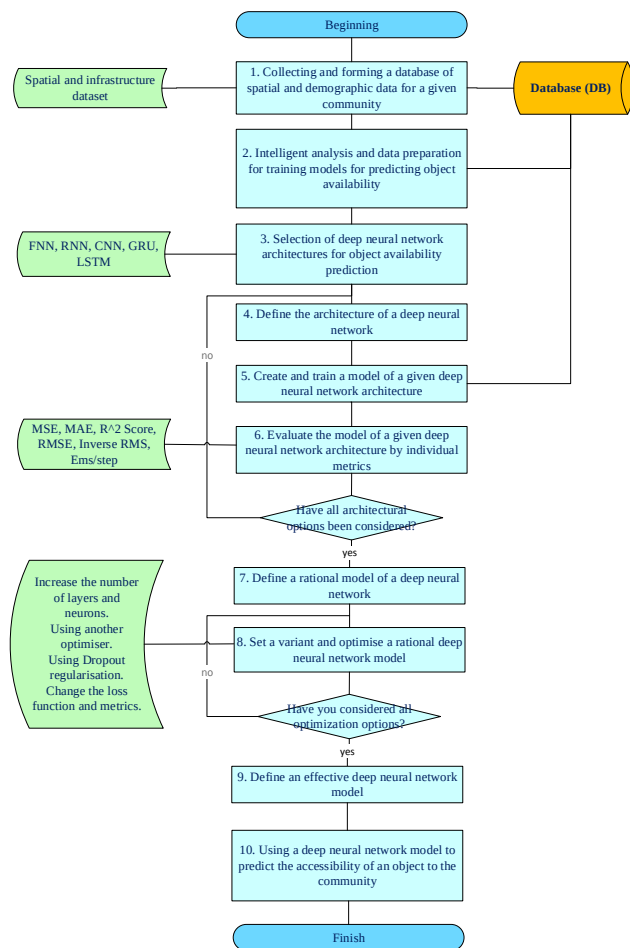


Figure 1 – An algorithm for researching to substantiate and optimize the architecture of a deep neural network model for predicting the availability of objects in spatially distributed systems

Building a model for predicting the level of accessibility of objects in spatially distributed systems involves an integrated approach that combines the stages of geospatial data processing, formation of an input set of features, building a deep learning model, and evaluating its accuracy.

At the first stage, we construct a directed graph of the road network that models the topological structure of the territory:

$$G = (V, E). \quad (1)$$

This graph is used to calculate the reachability metrics between points, which are used as features for modeling spatial accessibility.

At the same time, a set of spatial and infrastructural characteristics of the territory is formed by analyzing polygonal boundaries, road networks, the location of infrastructure facilities, and natural barriers. Spatial filtering allows you to select objects located within the analyzed area.

After that, important spatial indicators are calculated [32–34]. For example, the density of the road network for the i -th community is defined as:

$$\rho_i = \frac{L_i}{A_i}. \quad (2)$$

Indicator (2) allows us to assess the potential accessibility of settlements within the community.

To take into account natural obstacles, a barrier index is formed, which reflects the number of route intersections with natural obstacles (water bodies, forests) and is used as a factor in reducing operational accessibility:

$$B_i = \sum_{j=1}^n w_j \cdot b_{ij}. \quad (3)$$

The weighting factor w_j for the impact of a particular obstacle on accessibility has different values (e.g., 0.7 for a river, 0.3 for a forest).

The population density D_i is calculated based on official statistical data according to the formula:

$$D_i = \frac{P_i}{A_i}. \quad (4)$$

Step 2. Intelligent analysis and preparation of input data. At this stage, a preliminary analysis of spatial and infrastructural data is carried out to form a structured set of features for training deep learning models [35–38]. The main task is to identify the key characteristics that determine the level of spatial accessibility of objects within the analyzed area.

Within the constructed dataset, the target variable is determined – the accessibility index, which represents the

presence or absence of conditions for ensuring access within a given spatial area. For each unit of analysis, a description is generated that contains a number of quantitative predictors that take into account the spatial, infrastructural, social, and transportation characteristics of the environment.

During the data cleaning, a number of procedures were carried out to improve the quality of input features for modeling. In particular, duplicates with the same geographical coordinates were eliminated to avoid the influence of duplicate records on the model training results. After the initial processing, the features were selected: attributes that did not carry an information load due to the lack of variability of values were excluded from the set. Such attributes included characteristics whose values were equal to zero in all records, which justified their removal from further analysis.

As a result of the preliminary analysis and data cleaning, eight informative features were formed, which are used to build models for predicting the spatial accessibility of objects. The selected features include characteristics of the road network, infrastructure provision and transport accessibility of the territory.

The formation of a compact set of features reduces the risk of model overtraining, avoids the influence of redundant or trivial parameters, and improves the quality and stability of the training process.

To minimize the impact of different scales of numerical characteristics, data are normalized [39]. In particular, indicators such as building density, number of infrastructure facilities, length of the road network, and area of territorial zones are brought to a unified scale [0;1] using minmax normalization:

$$x' = \frac{x - x_{\min}}{x_{\max} - x_{\min}}. \quad (5)$$

This approach ensures the same scale of input data for the model, which has a positive effect on the stability of weighting optimization and the speed of convergence during training.

The analysis of the scaling results showed that some spatial characteristics have a narrow range of variation, which indicates a significant uneven distribution of infrastructure elements within different territories. Normalization made it possible to adapt these features to a single scale without losing their impact on the modeling results.

As a result, a vector of input features is formed for each unit of analysis:

$$\bar{x}_i = [x_{1i}, x_{2i}, \dots, x_{ni}]. \quad (6)$$

Preparing input data involves not only aggregating numerical indicators but also conceptually structuring spatial information in a form suitable for submission to deep learning models [40]. This approach provides comprehensive coverage of the factors that affect the accessibility of objects.

Step 3. Selecting the types of deep neural network architectures. After forming a structured set of spatial and infrastructural features for each territorial unit, the task is to choose the deep neural network architecture that best suits the data features and requirements for predicting the level of accessibility of objects.

The choice of architecture is determined by the nature of the input data, the type of forecasting task, and the requirements for accuracy and stability of the results. In this case, the input data is presented in the form of a tabular numeric array, where each row is a vector of features of the corresponding territorial unit formed according to (6).

The main goal is to build a model capable of predicting the value of the target variable in a binary classification task.

The basic option for working with tabular data is a Fully Connected Neural Network (FNN), which is a sequence of tightly connected layers, each of which performs a linear transformation of the input vector followed by a nonlinear activation:

$$\bar{h}^{(l)} = f(W^{(l)}\bar{h}^{(l-1)} + \bar{b}^{(l)}). \quad (7)$$

The FNN architecture is a universal function approximator and is well suited for processing structured numerical data. Depending on the complexity of the problem and the number of input features, the number of layers and the number of neurons in each layer are determined, which allows the model to be flexibly customized to specific forecasting conditions.

In cases where the input features have a pronounced spatial dependence, it is advisable to use convolutional neural networks (CNN). Such networks allow for effective detection of local spatial structures and patterns in the data by performing a convolutional operation:

$$s_{ij} = (\bar{x} * \bar{k})_{ij} = \sum_m \sum_n x_{i+m, j+n} \cdot k_{m,n}. \quad (8)$$

In this context, CNNs can be adapted to process spatial representations of data, for example, in the form of density distribution maps or environmental characteristic matrices.

If the study involves the processing of data sequences with time dynamics, it is advisable to use recurrent neural networks (RNNs), which allow modeling dependencies in time by preserving the internal state:

$$\bar{h}_t = f(W_x \bar{x}_t + W_h \bar{h}_{t-1} + \bar{b}). \quad (9)$$

Since classical RNN architectures are prone to the problem of gradient disappearance or explosion when working with long sequences, their modifications are often used in practice – Long Short-Term Memory (LSTM) and Gated Recurrent Units (GRU). These architectures store information better over long time

intervals and demonstrate high efficiency in time-dependent forecasting tasks.

In cases where the data have both spatial and temporal structure, it is advisable to use combined architectures, in particular, Recurrent Convolutional Neural Networks (RCNN). They allow for to simultaneously take into account of local spatial features and temporal dynamics of features, which makes them promising for modeling the evolution of the state of the spatial environment over time.

Step 4. Formalizing the basic architectures of deep neural networks. After determining the types of architectures suitable for solving the problems of predicting the availability of objects in spatially distributed systems, their structure is specified by formally setting the basic parameters. At this stage, the key characteristics of the network are determined, including the number of layers, the number of neurons in each layer, types of activation functions, loss function, optimization algorithm, and training parameters, such as the size of the data packet and the number of epochs.

For a Fully-connected Neural Network (FNN) model, the basic architecture usually includes one input layer, several hidden layers, and one output layer. The network structure is adapted to the dimensionality of the input feature vector. For example, when working with a feature vector of nnn features, the basic architecture may include an input layer with nnn neurons, two or three hidden layers with 128 and 64 neurons, respectively, and an output layer with one neuron to implement binary classification:

$$\begin{aligned}\bar{h}^{(1)} &= f_1(W^{(1)}\bar{x} + \bar{b}^{(1)}), \\ \bar{h}^{(2)} &= f_2(W^{(2)}\bar{h}^{(1)} + \bar{b}^{(2)}), \\ y &= \sigma(W^{(3)}\bar{h}^{(2)} + b^{(3)}).\end{aligned}\quad (10)$$

The choice of the number of layers and the number of neurons is made taking into account the dimensionality of the input data, the complexity of the problem, and the requirements for the generalization ability of the model.

In the case of Convolutional Neural Networks (CNNs), the basic architecture usually includes two or three convolutional layers with kernels of size 3×3 or 5×5 , each of which is accompanied by subsampling operations to reduce the dimensionality of the features. After the convolutional layers, a block of fully connected layers is formed, which ensures the integration of local features into global representations. This structure allows the model to efficiently detect spatial patterns in data represented as density matrices or images.

For recurrent architectures, such as LSTM (Long Short-Term Memory) or GRU (Gated Recurrent Unit), the input data is represented in the form of sequences. The basic configuration of such models involves one or two memory layers with the number of neurons in the range of 64–128, after which the result is processed through the output dense layer:

$$\bar{h}_t = \text{LSTM}(\bar{x}_t, \bar{h}_{t-1}), \quad y = \sigma(W\bar{h}_T + b). \quad (11)$$

Regardless of the chosen architecture, binary cross-entropy is used to solve the problem of binary classification of the target variable as a function of losses:

$$L = -\frac{1}{N} \sum_{i=1}^N [y_i \cdot \log(\hat{y}_i) + (1 - y_i) \cdot \log(1 - \hat{y}_i)]. \quad (12)$$

A loss function of this type contributes to the correct training of the model in tasks with two possible classes of outcomes.

In the process of training deep neural networks for predicting the availability of objects, adaptive optimization algorithms are used, among which Adam and RMSProp are the basic ones. These methods provide efficient updating of weights based on the moments of gradients, which contributes to fast convergence even in multidimensional parameter spaces.

Regardless of the chosen architecture, binary cross-entropy is used to solve the problem of binary classification of the target variable as a function of loss:

Typical training parameter settings include:

- batch size – 32 or 64 examples per iteration;
- number of epochs – from 50 to 150, depending on the dynamics of the loss function during training;
- initial learning rate – usually set to 0.001 when using the Adam optimizer.

The initial setting of the basic architecture and optimization parameters forms the initial basis for the subsequent stage of fine-tuning the model to achieve the best forecasting results.

Step 5. Creating and training deep learning models. After defining the basic architecture of each model, its implementation is performed by initializing the network in accordance with the specified layer structure, activation functions, optimization algorithm, and loss function.

Let the input dataset be represented by a matrix $X \in \mathbb{R}^{n \times m}$, where n – is the number of examples, and m – is the number of features for each example. The target vector $\bar{y} \in \{0,1\}^n$ contains the corresponding values of the target variable.

The data set is divided into three parts – training, validation, and test samples in the ratio of 70:15:15. This division makes it possible to simultaneously train the model and evaluate its ability to generalize to unknown data.

The training process is aimed at minimizing the loss function $L(\theta)$, which for the binary classification task has the form of cross-entropy:

$$L(\theta) = -\frac{1}{n} \sum_{i=1}^n [y_i \cdot \log(\hat{y}_i) + (1 - y_i) \cdot \log(1 - \hat{y}_i)]. \quad (13)$$

Training is performed iteratively by epochs. During each epoch, the full set of training examples is divided

into packages (battles) of fixed BBB size (e.g., 32 or 64 elements). The weights are updated after each battle is processed according to the stochastic gradient descent method or its adaptive variations:

$$\theta_{t+1} = \theta_t - \eta \cdot \Delta \theta L_B. \quad (14)$$

At each step, the loss function and model accuracy are evaluated on the validation sample. If signs of overtraining are detected, when the validation error increases while the training error decreases, an early stopping mechanism is applied – fixing the best model state in accordance with the minimum value of the loss function on the validation data.

Thus, the process of creating and training models ensures the transition from a conceptual description of the architecture to the construction of a ready-made predictive model capable of effectively solving classification problems based on spatial and structural features.

Step 6. Evaluation of model performance by accuracy metrics. After completing the training stage, the effectiveness of the built models is evaluated on a test sample that was not used in the training and validation process. This approach allows us to determine the ability of the models to generalize and the accuracy of forecasting on new, previously unknown data.

To quantify the quality of forecasts, a set of metrics is used to comprehensively characterize the accuracy and stability of the models.

One of the basic metrics is the Mean Squared Error (MSE), which determines the average square of the difference between the predicted $\hat{y}_i$ and actual y_i values:

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2. \quad (15)$$

MSE is sensitive to large deviations, so it reflects the overall level of error.

To supplement the estimate, the Mean Absolute Error (MAE) is used:

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i|. \quad (16)$$

MAE is less sensitive to single large deviations and better demonstrates the average level of forecast error.

Another important metric is Root Mean Squared Error (RMSE):

$$RMSE = \sqrt{MSE}. \quad (17)$$

The RMSE has the same dimension as the predicted variable, which makes it easier to interpret the results.

The inverse of RMSE is used to estimate the inverse of the error:

$$\text{Inverse RMSE} = \frac{1}{RMSE}. \quad (18)$$

The value of (18) increases as the model accuracy improves.

The coefficient of determination (R^2) is another key metric that shows how much of the variance in the dependent variable is explained by the model:

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2}. \quad (19)$$

A value R^2 close to 1 indicates that the model is a good fit to the actual data.

In addition to the classical metrics, the average error per step (EMS/step) is used to assess the stability of the models, which determines the average difference between the forecasts and the actual values at each step in the conditional scenario space. This metric allows us to assess the uniformity of forecasting quality across the entire dataset.

The analysis of the results allows us to identify the architecture that provides the best forecasting quality, the smallest errors, and the highest level of generalization. The selected model is further used for optimization and practical application in the tasks of predictive analysis of spatial characteristics.

Step 7. Selection of a rational deep neural network model. After building several architectural options and training the corresponding models, a comparative analysis of the results is carried out to select the most rational model. In this context, rationality is understood as the optimal ratio between forecasting accuracy and computational costs required for practical application.

The process of model selection is based on a multi-criteria approach, in which each model is evaluated by a vector of quality metrics:

$$Q(M_j) = (q_{1j}, q_{2j}, \dots, q_{kj}). \quad (20)$$

The computational complexity of the model may include such characteristics as the number of parameter parameters, training time, and the amount of RAM required for operation.

To make an objective choice, a trade-off is made between forecasting accuracy and computational efficiency. One approach is to build an aggregate quality function:

$$F(M_j) = \sum_{i=1}^k w_i \cdot q_{ij}. \quad (21)$$

The weights can be determined by experts depending on whether forecasting accuracy is preferred (higher weights for MSE, RMSE) or computational efficiency.

Alternatively, the Pareto-dominance method can be applied, when a model is selected that is not inferior to any other model in all key metrics simultaneously.

In addition to quantitative characteristics, the stability of the learning process is also taken into account: smoothness of the loss function decline, absence of overfitting, and speed of gradient convergence. Models that demonstrate an excessive number of epochs before reaching the minimum or significant deviations of predictions on the test set are not considered rational, even with high accuracy rates on the training data.

Thus, the choice of a rational model is based on a comprehensive analysis of its quality, resource efficiency, and stability of the learning process.

Step 8. Optimization of the rational model architecture. After determining the rational architecture of the neural network, the stage of its targeted optimization is performed. The purpose of optimization is to reduce the forecasting error, increase the generalization ability of the model, and ensure computational efficiency without significantly complicating the structure.

Optimization begins with setting the quantitative characteristics of the architecture, in particular, determining the number of hidden layers L and the number of neurons n_l on each layer l . The next step is to choose the activation functions for each layer.

The standard activation function is the ReLU (Rectified Linear Unit), which is defined as:

$$f(x) = \max(0, x). \quad (22)$$

To improve the handling of negative input values, modifications such as leaky ReLU or ELU can be applied, e.g:

$$f(x) = \begin{cases} x, & \text{if } x \geq 0, \\ \alpha x, & \text{if } x < 0, \end{cases} \quad (23)$$

To prevent overtraining, regularization mechanisms are integrated into the architecture, in particular the Dropout method, which involves randomly “disconnecting” a certain part of neurons during each training iteration. If the probability of maintaining neuronal activation is denoted as p , then the output of the neuron during training looks like:

$$\tilde{h}^{(l)} = \text{Dropout}(f(W^{(l)}\tilde{h}^{(l-1)} + \tilde{b}^{(l)}), p), \quad (24)$$

where $z \sim \text{Bernoulli}(p)$.

Additionally, the Batch Normalization method is used to stabilize the learning process by normalizing the output values of each layer to zero mean and unit variance.

This stage results in the final version of the model that combines high forecasting accuracy, training stability, consistency with input data characteristics, and efficiency in terms of computational costs.

Step 9. Determination of an effective model. At the initial stage of optimization, an effective deep learning model is formally defined as one that has achieved the best results in terms of a set of criteria for prediction accuracy, stability of the learning process, and computational performance.

An effective model is defined as an architecture that minimizes the loss function L , demonstrates a stable value of the coefficient of determination R^2 , and provides the lowest root mean square error (RMSE):

$$\text{RMSE} = \sqrt{\text{MSE}} \rightarrow \min. \quad (25)$$

The final selected model is stored in the form of weighting parameters and a structural layout of layers. It is used as a basic solution for further forecasting tasks that require analysis of spatial or structural data.

Step 10. Use the model to predict the accessibility of objects. Once an effective model is determined, it is applied in the format of a predictive module that estimates the probability of a new object belonging to a certain accessibility class based on its spatial and infrastructural features.

Let the input vector of features of a new object be denoted as x^* . The model generates a prediction of the probability of belonging to a positive class according to the formula:

$$\hat{y} = f(x^*, \theta). \quad (26)$$

The classification is based on the threshold principle: if $\hat{y} \geq \tau$, the object belongs to the availability class, if $\hat{y} < \tau$, the object is classified as inaccessible. The threshold value τ is usually set to 0.5, but can be adjusted depending on the requirements for sensitivity or specificity of the forecast.

The use of such a model allows for a quick assessment of the characteristics of new objects based on open spatial data, which expands the capabilities of analytical and forecasting systems in the field of spatial planning.

4 EXPERIMENTS

To perform experimental studies, the procedure for forming an input data set was implemented to solve the problem of creating a model for predicting the level of availability of objects in spatially distributed systems. Particular attention was paid to the use of open sources of geospatial information that provide the necessary level of detail of spatial, infrastructure, and logistics features. This approach allows to form a structured input dataset suitable for training deep learning models.

To demonstrate the practical application of the proposed approach, an example of forming an input feature database based on OpenStreetMap data was implemented. The choice of the source is due to the high availability, relevance, and structuredness of the data, which cover the key characteristics of road infrastructure,

location of critical infrastructure facilities, types of buildings, street network density, and other relevant indicators.

The territory of Zolochiv territorial community in Lviv region (Ukraine) was chosen as an example of a spatial zone for building the input set. This region has a pronounced rural structure with significant internal differentiation by infrastructure features, which allows testing the model on data with a high level of variability. In the process of preparation, 16 features were identified for each of the community's settlements, including the density of the road network, the number of educational,

healthcare, and critical infrastructure facilities, geographic accessibility indices, and building indicators.

All the obtained values were normalized using the Min-Max method to the interval , after which the training, validation, and test samples were formed. The created dataset was used as a basis for training deep learning models, the architecture of which was described in detail in the previous section. To determine a rational model, five basic architectures of deep neural networks were chosen (Fig. 2).

Layer (type)	Output Shape	Param #
dense (Dense)	(None, 64)	576
dense_1 (Dense)	(None, 32)	2,080
dense_2 (Dense)	(None, 1)	33

a

Layer (type)	Output Shape	Param #
conv1d (Conv1D)	(None, 7, 32)	96
flatten (Flatten)	(None, 224)	0
dense_3 (Dense)	(None, 1)	225

b

Layer (type)	Output Shape	Param #
simple_rnn (SimpleRNN)	(None, 32)	1,088
dense_4 (Dense)	(None, 1)	33

c

Layer (type)	Output Shape	Param #
gru (GRU)	(None, 32)	3,360
dense_5 (Dense)	(None, 1)	33

d

Layer (type)	Output Shape	Param #
lstm (LSTM)	(None, 32)	4,352
dense_6 (Dense)	(None, 1)	33

e

Figure 2 – Variants of deep neural network architectures for predicting the availability of volunteer rescue groups in rural communities: a – Fully-connected Neural Network (FNN); b – Convolutional Neural Network (CNN); c – Recurrent Neural Network (RNN); d – Gated Recurrent Unit (GRU); e – Long Short-Term Memory (LSTM)

They are applied to the task of binary classification of the availability of voluntary rescue services based on the given characteristics of rural communities. The first model is constructed as a classical feed-forward neural network (FNN) containing two hidden layers with 64 and 32 neurons, respectively, and uses the ReLU activation function, and the sigmoidal function is used as an output to obtain the probability of availability.

The second model is a convolutional neural network (CNN) that uses a Conv1D layer with 32 filters and a kernel size of 2. The input data is transformed into the form (8, 1), i.e. each of the eight normalized indicators is treated as a time step. After convolution, the data is flattened and processed in a dense layer with 16 neurons.

The third model implements the architecture of a simple recurrent network (RNN), which consists of a SimpleRNN layer with 32 units and has an output layer for classification. Similarly, a fourth model based on GRU elements was constructed, which replace the simple RNN with a more complex and resistant to gradient reduction structure. The number of GRU units in the layer is 32.

The fifth model is built using an LSTM layer that also uses 32 memory elements. This option is the most complex of the presented models due to the internal structure of the LSTM, which allows for more efficient modeling of dependencies in the sequence of input features. All models have an output layer with one neural

unit and a sigmoid function to predict the probability of class 0 or 1. Visualizations of the architectures are created using the plot_model() function of the Keras library, which displays the shape of each layer and its parameters.

At the compilation stage, each of the models was configured with the same parameters: 1) the optimizer was Adam with a learning rate of 0.001; 2) the loss function was binary_crossentropy; 3) the evaluation metric was accuracy. This unified setting allows for a correct comparison between architectures, since they are all trained under the same conditions.

Training was performed for 50 epochs with a batch size of 8. For each model, after training was completed, an evaluation was performed on the test set using a special function evaluate_model, which returned the selected metrics (15–19). The obtained values were stored in the metrics_results dictionary, which allowed us to summarize the information for the subsequent visualization of the results and determination of the most efficient architecture.

5 RESULTS

As a result of the experimental study, the territory of the Zolochiv territorial community of the Lviv region (Ukraine) was chosen to build the input dataset. The information obtained is presented in the form of a DataFrame (df) (Table 1).

Table 1 – A fragment of the database for predicting the availability of volunteer rescue teams in the Zolochiv community of Lviv Oblast (Ukraine)

	community	village	lat	lon	area_km2	road_length_km	road_density	n_fire_stations	accessibility_index	n_schools	n_hospitals
0	Золочівська громада	Підлисса	49.933105	24.822783	0.0	128.522	0	0	0	3	0
1	Золочівська громада	Гавареччина	49.919228	24.884085	0.0	57.904	0	1	1	2	2
2	Золочівська громада	Гутище	49.918442	24.947210	0.0	88.026	0	1	1	3	1
3	Золочівська громада	Розваж	49.911228	24.797859	0.0	33.890	0	0	0	3	0
4	Золочівська громада	Бужок	49.895750	24.820013	0.0	129.492	0	0	0	3	0
5	Золочівська громада	Білий Камінь	49.896170	24.835809	0.0	113.608	0	0	0	3	0
6	Золочівська громада	Ушня	49.890568	24.919470	0.0	84.781	0	1	1	2	0
7	Золочівська громада	Побіч	49.890293	24.964476	0.0	86.296	0	1	1	1	0
8	Золочівська громада	Черемошня	49.893113	24.858039	0.0	95.729	0	0	0	3	0
9	Золочівська громада	Бір	49.880768	24.953941	0.0	85.989	0	1	1	2	0

In the resulting df, each record corresponds to one community settlement and contains numerical, spatial and categorical attributes that reflect the real conditions of the territory. This completes the formation of the input dataset, which will be the basis for further analysis, construction of neural network models and forecasting the level of availability of volunteer rescue groups.

The description of the df attributes, which contains data for modeling the availability of volunteer rescue groups in rural communities, is presented in Table 2.

In the process of building a model for predicting the level of accessibility of objects in spatially distributed systems, the main target variable is the accessibility_index indicator, which reflects the presence or absence of spatial coverage of a given object, taking into account infrastructure and transport characteristics. To train the model, quantitative spatial and structural features describing individual territorial units are used (Table 2).

In particular, the set of input features includes the length of the road network, building density, the number of social and functional facilities (educational institutions, medical facilities, religious buildings, industrial zones), as well as the presence or absence of access to major transport arteries. Additionally, geographical coordinates (latitude and longitude) can be used, which are either entered into the model as predictors or used to visualise the results in the spatial environment.

The use of such features allows us to comprehensively describe the spatial and infrastructural profile of an object and improve the accuracy of the forecast of its availability. The scaled values of the input parameters used to train the models are shown in Table 3.

As can be seen from Table 3, the scaling of the features made it possible to bring all indicators to a single range [0;1], which is necessary for the stable functioning of deep learning models. Some features (e.g.,

n_fire_stations, n_hospitals) contain fragmented information that takes into account the partial presence of functional objects, reflected in the form of fractional values.

Table 2 – Characterization of the attributes of df, which contains data for modeling the availability of volunteer rescue groups in rural communities

Attribute name	Type	Description
community	object (str)	Name of the territorial community
village	object (str)	Name of the village or settlement within the community
lat	float64	Geographical latitude of the village centroid
lon	float64	Geographical longitude of the village centroid
area_km2	float64	Area of the buffer zone around the settlement, km ²
road_length_km	float64	Total length of roads in the village buffer zone, km
road_density	float64	Density of the road network (road_length_km/area_km2)
n_fire_stations	int64	Number of fire stations (including voluntary formations) within the village buffer
accessibility_index	int64	Binary sign (0 or 1) – whether the village has access to a voluntary rescue group (1 – has, 0 – does not)
n_schools	int64	Number of schools within the village (indicator of critical social infrastructure)
n_hospitals	int64	Number of medical facilities (hospitals, outpatient clinics, rural health posts)
n_churches	int64	Number of places of worship (churches, chapels, etc.) – an indicator of places of mass gathering
n_industrial	int64	Number of industrial facilities (industrial zones, factories, workshops) – potential sources of risk
n_buildings	int64	Total number of buildings within the buffer
building_density	float64	Building density: Number of buildings per unit area (buildings/km ²)
major_road_access	int64	Binary feature (0 or 1) – whether there is a main road nearby (primary or secondary according to OSM)

Table 3 – Fragment with the results of feature scaling

	road_length_km	n_fire_stations	n_schools	n_hospitals	n_churches	n_industrial	n_buildings	major_road_access
0	0.362206	0.0	0.272727	0.000000	0.285714	0.764706	0.271600	0.0
1	0.126874	0.5	0.181818	0.333333	0.523810	0.352941	0.323860	0.0
2	0.227254	0.5	0.272727	0.166667	0.619048	0.352941	0.170513	0.0
3	0.046848	0.0	0.272727	0.000000	0.238095	0.176471	0.100515	0.0
4	0.365438	0.0	0.272727	0.000000	0.333333	0.058824	0.014686	0.0

Particularly important is the `major_road_access` feature, which, despite its binary nature, allows the model to take into account the strategic accessibility of objects in terms of transport links, which has a significant impact on the formation of the forecast. The obtained set of features forms a vector of spatial representation of the object, which is the basis for the classification forecast of the level of accessibility in the deep learning model.

At the stage of analysing the relationships between spatial and infrastructural features and the target variable `accessibility_index`, a correlation matrix based on Pearson's coefficients was built. Such a matrix allows us to assess the strength and direction of the linear relationship between the variables included in the forecasting model (Fig. 3).

The analysis of the results showed that the highest level of positive correlation with the target variable is for `n_fire_stations` ($r = 0.97$), which indicates an almost linear relationship between the availability of facilities with the appropriate functionality and an increased level of accessibility.

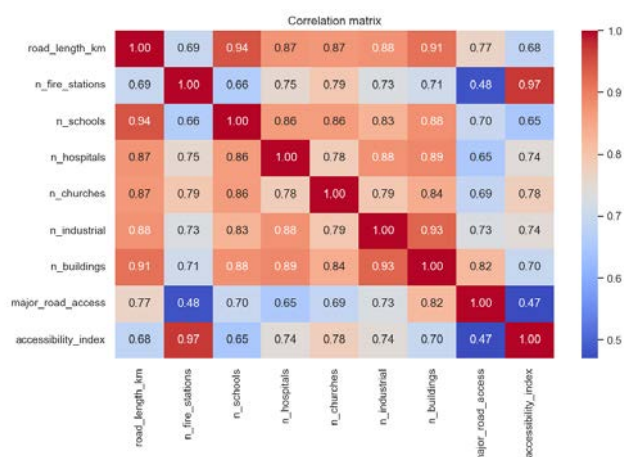


Figure 3 – Correlation matrix of relationships between input features and the target feature “accessibility_index”

Weaker, but statistically significant relationships are also observed for the variables `n_hospitals` ($r = 0.47$), `n_industrial` ($r = 0.44$), and `n_buildings` ($r = 0.36$). Their positive correlation may indicate the accompanying presence of infrastructure elements in areas with a higher level of general accessibility.

Moderate positive correlation is also demonstrated by `road_length_km` ($r = 0.68$) and `n_churches` ($r = 0.30$), while the relationship with `n_schools` is neutral-negative

($r = -0.06$), which may be due to indirect effects or the peculiarities of the distribution of educational institutions in the sample.

`Major_road_access` has the lowest correlation with the target indicator ($r = 0.47$), which indicates the presence of independent factors that do not provide a direct effect between transport accessibility and the predicted variable.

In general, the analysis allowed us to identify the most informative features that will have a decisive impact in the classification model. In particular, access to a voluntary rescue unit.

The study compared the effectiveness of five deep neural network architectures that implement different approaches to processing pro- tectively structured data. The corresponding results are presented in Fig. 4, which shows the numerical values of metrics characterizing the quality of the built models.

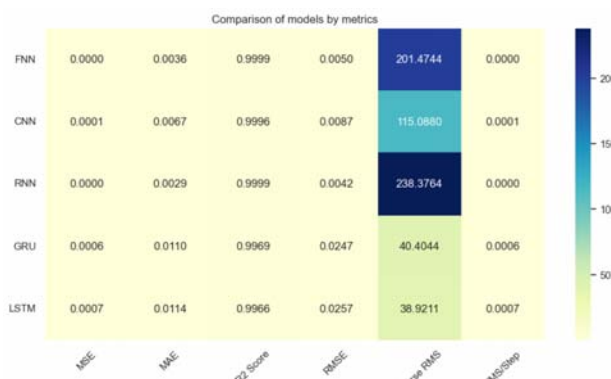


Figure 4 – Results of comparing the performance of deep neural network architectures

Among the evaluation indicators used are the mean square error (MSE), which allows us to determine the average square deviation of forecasts from actual values, and the mean absolute error (MAE), which reflects the average level of absolute deviation between the forecast and the target value. The coefficient of determination (R^2) is used to estimate the explained variance of the target variable, while the root mean square error (RMSE) provides a convenient interpretation of accuracy in the same units as the original variable. In addition, the Inverse RMSE allows for a unified presentation of the inverse effect of precision, providing an alternative interpretation. The generalized error per step (EMS/Step) allows you to evaluate the dynamics of error changes in the forecasting process over the entire data set.

The application of these metrics made it possible not only to quantitatively compare the accuracy of different architectures, but also to identify models with the highest level of stability, adaptability, and generalizability for further practical use.

Based on the results shown in Fig. 4, it can be concluded that the recurrent neural network (RNN) model provided the best forecasting quality among all the considered architectures. In particular, it demonstrated the lowest mean square error ($MSE=0.000018$) and mean absolute error ($MAE=0.002879$), which indicates a high accuracy of reproducing the dependencies between the features. The coefficient of determination for this model was $R^2=0.999910$, which indicates an almost complete ability to explain the variation of the target variable on the test sample. The RMSE was the smallest among all the models and amounted to only 0.004195, and the Inverse RMSE value reached 238.38, which further confirms the excellent consistency between the predicted and actual values.

Among the competitors, the closest results were shown by the FNN model, for which $MSE=0.000025$, $MAE=0.003578$, and $R^2=0.999874$. It is also characterized by a low RMSE value (0.004963), although it is inferior to RNN in all major accuracy criteria. The CNN model turned out to be slightly less accurate, with $MSE=0.000075$ and $RMSE=0.008689$, but at the same time retained a high value of the coefficient of determination ($R^2=0.999614$), which makes it competitive for predicting spatial dependencies.

Instead, the GRU and LSTM architectures showed significantly worse results. The RMSE value for GRU was 0.024750, and for LSTM – 0.025693, which is several times higher than the corresponding values of RNN and FNN models. In addition, the MAE in these cases exceeded 0.011, and the Inverse RMSE values were significantly lower – 40.40 and 38.92, respectively. This indicates less stability and worse generalization ability.

Thus, among the considered architectures, it is the RNN model that demonstrates the best balance between accuracy, stability, and the ability to reproduce complex spatial patterns. This model was chosen for further optimization and structural improvement, the results of which are presented in Fig. 5.

The optimized recurrent neural network model includes one layer of SimpleRNN with 32 neurons, which processes the input sequence of features and generates the corresponding representation of temporal dependencies. This layer provides a parameterized representation of spatially structured input data in the context of temporal dynamics. After the recurrent layer, a dense layer (Dense) with one output neuron and a sigmoidal activation function is applied, which implements the binary classification mechanism.

The total number of model parameters is 1,121, of which all are trained. Such a compact architecture allows to maintain high performance during training while providing sufficient flexibility to model dependencies between features. The model was compiled using the

Adam optimizer and the binary_crossentropy loss function, which is the standard for two-class classification problems.

Layer (type)	Output Shape	Param #
simple_rnn_4 (SimpleRNN)	(None, 32)	1,088
dense_17 (Dense)	(None, 1)	33

Total params: 1,121 (4.38 KB)
Trainable params: 1,121 (4.38 KB)
Non-trainable params: 0 (0.00 B)

Figure 5 – Architecture of an optimized recurrent neural network for modeling the level of availability of objects in spatially distributed systems based on deep learning

The model was trained for a specified number of epochs with controlled validation monitoring. Upon completion of the process, graphs were plotted showing the change in the loss function and accuracy in the training and validation samples. A visualization of the training dynamics is shown in Figure 6.

The analysis of the training dynamics of the optimized recurrent neural network presented in the graphs (Fig. 6) demonstrates an effective reduction in the loss function and a steady increase in model accuracy throughout the training process. The graph on the left shows that the values of the loss function for the training and validation sets decrease rapidly during the first 10 epochs, after which they stabilize at a low level. This behavior of the loss curves indicates the absence of overfitting and indicates the model's ability to generalize the input data.

The model accuracy graph (right) confirms the conclusions about the quality of training: the classification accuracy increases to over 98% at the initial stages and remains stable for both the training and validation sets. This indicates that the model is well matched to the training data and does not lose performance on new examples, which indicates the effectiveness of its architectural solution.

The achieved level of stability and accuracy is the result of the use of a deeper multilevel structure, Dropout type regularization, and the use of an optimized weight update algorithm (Adam). The visualized results confirm the feasibility of the chosen architecture for the classification task in spatially distributed systems.

Based on the above graph (Fig. 7), we can conclude that the optimized recurrent neural network model is highly accurate in reproducing the target values. The line of actual values, constructed on the basis of the real values of the accessibility variable, almost completely coincides with the graph of predicted values, which indicates the absence of significant errors in the classification of test examples.

The model correctly reproduces the binary structure of the target variable, clearly distinguishing objects with a high level of accessibility (designation 1) from objects with insufficient or no coverage (value 0). Particularly noteworthy is the complete coincidence of predicted and actual values in areas of sharp class change, which indicates the high sensitivity of the model to changes in the simple infrastructure characteristics of the input set.

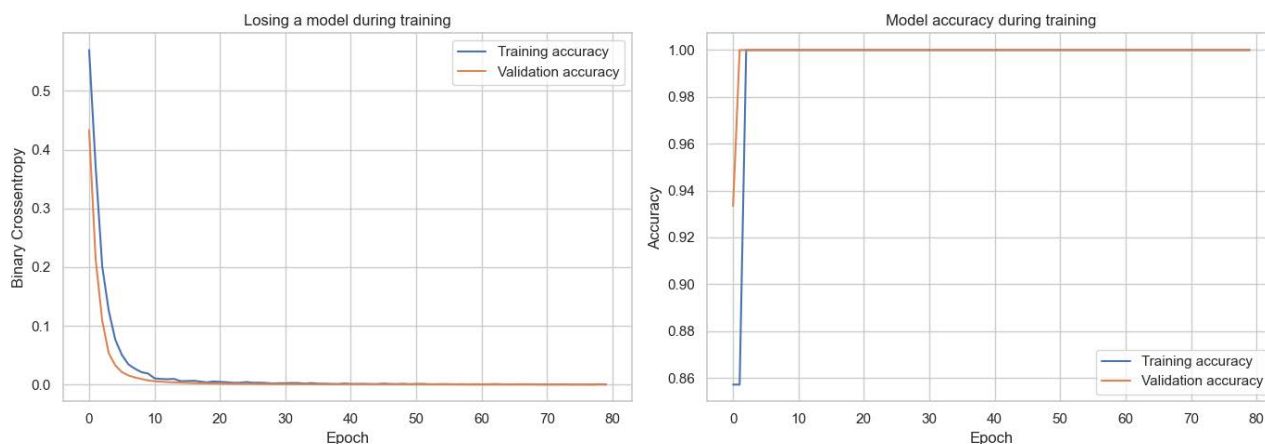


Figure 6 – Graphs of the loss function and accuracy of an optimized recurrent neural network during training for a model for predicting the availability of objects in spatially distributed systems

As a result of our research, a graph comparing the actual and predicted values of the target variable in the test sample using the optimized RNN model was constructed (Fig. 7).

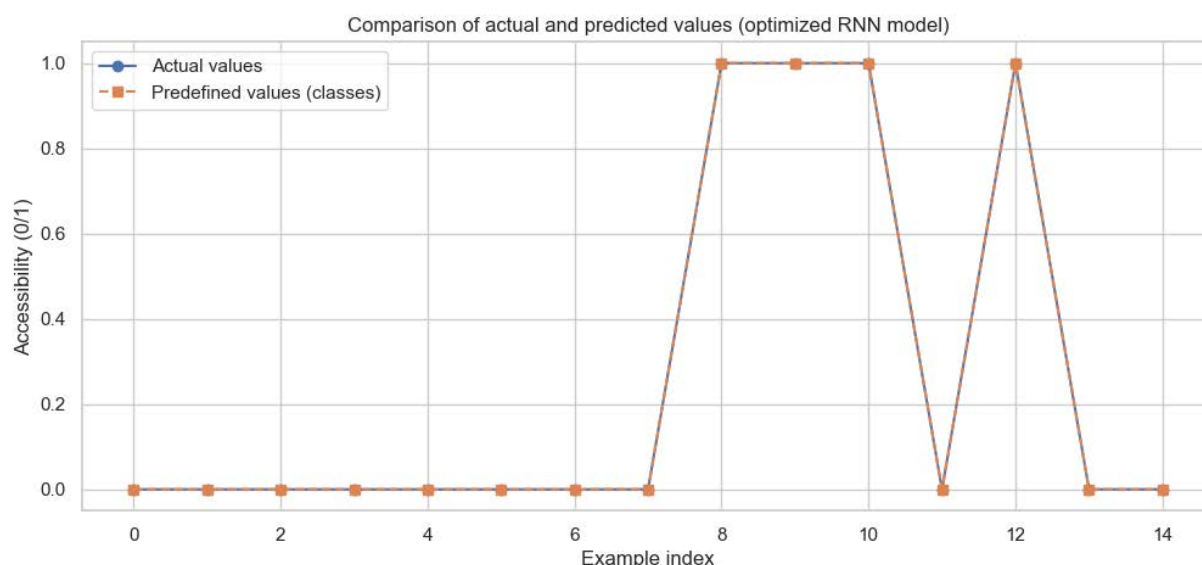


Figure 7 – Graph comparing the actual and predicted values of the target variable in the test sample using the optimized RNN model

The presence of only minor fluctuations or their complete absence confirms that the optimized model demonstrates not only the ability to highly accurate classification, but also the stability of prediction even under conditions of variability in the structure of test examples. This confirms the suitability of the constructed architecture for further use in predictive modeling tasks in spatial analysis systems.

The comparison between the baseline and optimized RNN model showed a significant improvement in the quality of forecasting after the implementation of the optimized architecture (Fig. 8).

Based on the results visualized in Figure 8, we can conclude that the quality of forecasting has significantly improved after optimizing the architecture of the recurrent neural network. All six presented metrics demonstrate a

significant advantage of the optimized model compared to the baseline. In particular, the mean square error (MSE) decreased from 1.76×10^{-5} to 7.66×10^{-8} , which indicates a significant decrease in the mean square deviation between predicted and actual values. The mean absolute error (MAE) demonstrates a similar trend, decreasing from 0.002879 to 0.000209, indicating an increase in accuracy at the level of individual examples.

The coefficient of determination (R^2) reached the ideal value of 1.000 in the case of the optimized model, as opposed to 0.99991 in the baseline, which confirms that the variation in the target variable on the test sample is fully explained. The root mean square error (RMSE) also decreased significantly, from 0.004195 to 0.000277, which strengthens the arguments in favor of high accuracy of predictions.

Comparison of Basic and Optimized RNN Models

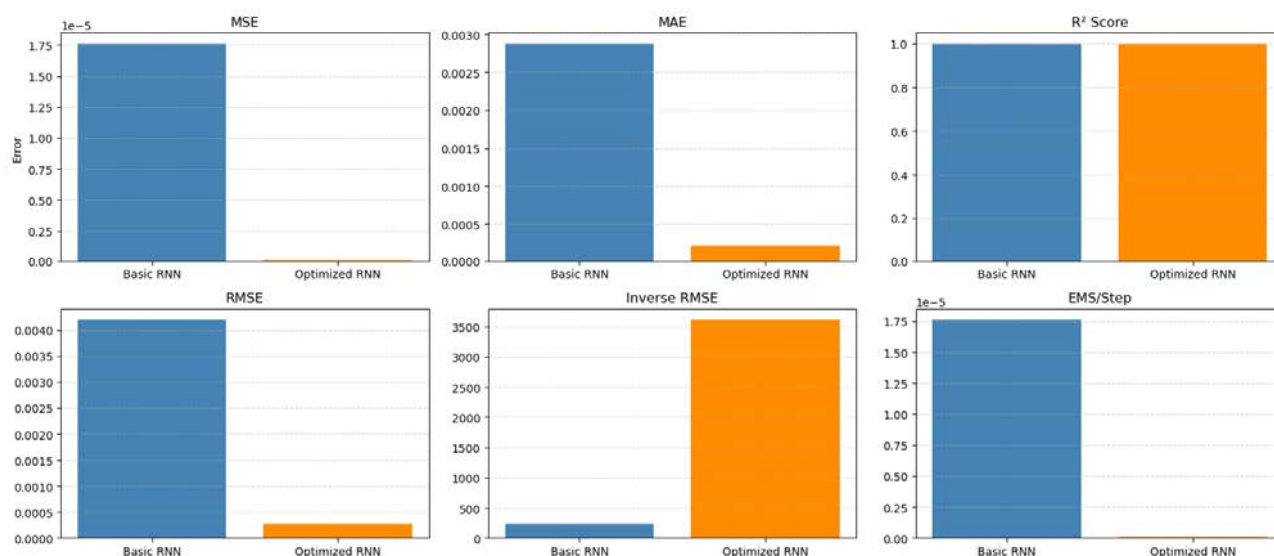


Figure 8 – Comparison of accuracy indicators between the baseline and optimized recurrent neural network model by the main metrics (MSE, MAE, R², RMSE, Inverse RMSE, EMS/Step)

The largest difference was recorded for the Inverse RMSE, which increased from 238.38 to 3612.75. This confirms that the optimized model provides significantly higher accuracy at the inverse transform scale. Similarly, EMS/Step, which characterizes the generalized error per observation unit, decreased by an order of magnitude: from 1.76×10^{-5} to 7.66×10^{-8} , which indicates stable model behavior in a variable spatial environment.

Thus, the comparative analysis confirms the effectiveness of the optimization. The model based on the improved architecture demonstrates not only an increase in forecasting accuracy, but also increased stability, generalizability, and reliability when applied to new data.

6 DISCUSSION

The study was aimed at developing a model for predicting the level of accessibility of objects in spatially distributed systems using deep neural networks and geospatial data. A comprehensive analysis of the subject area was carried out, the peculiarities of forming an input dataset based on open sources of geoinformation, in particular OpenStreetMap, were outlined, and key areas for further improvement of the modeling process were identified.

The study identified several critical challenges related to forecasting the availability of facilities, including high heterogeneity of spatial characteristics and the need for models to be adaptive to changes in infrastructure conditions. To overcome these challenges, we propose an approach that combines the formation of a high-quality input set of features using open geodata and the optimization of the architecture of a deep neural network.

Based on the results of experimental training of the basic architectures, it was found that the recurrent neural network (RNN) provides the highest quality of forecasting, demonstrating the values of the main accuracy metrics, in particular the coefficient of

determination and low levels of root mean square error (RMSE), which indicates a high correspondence of the model to real spatial dependencies.

Optimization of the RNN model by increasing the number of recurrent layers, applying Dropout regularization with a parameter of 0.2, and adjusting the number of neurons in the hidden layers significantly improved the results. After optimization, the model achieved even better accuracy, training stability, and generalization ability.

The results obtained showed that the proposed optimized architecture is able to effectively predict the level of accessibility of objects in complex spatial conditions. Prospects for further research are associated with scaling the approach to larger areas, as well as with the integration of the built model into geographic information systems for the operational analysis of spatial characteristics. An additional direction of development is the use of deep neural networks for automated processing of satellite images, which will optimize the formation of input features and increase the efficiency of planning spatial processes in various scenarios.

CONCLUSIONS

As a result of the study, an approach was proposed that involves 10 steps to optimize the model for predicting the availability of objects in spatially distributed systems by building an effective deep learning architecture focused on the use of spatial and infrastructure features to improve the accuracy of forecasting and generalization of models.

The input dataset was formed on the basis of open geospatial information, which ensured the creation of a high-quality spatially structured dataset of 16 features for training deep learning models. Particular attention was paid to the selection of features that most significantly affect the prediction of accessibility of objects, including

the characteristics of the road network, infrastructure facilities, and geographic accessibility indices.

Based on the analysis of 5 architectures of deep neural networks, the choice of a recurrent neural network (RNN) is substantiated as the most suitable for prediction tasks in spatially distributed environments. The architecture was optimized by adjusting the number of layers, neurons, regularization mechanisms, and activation functions, which allowed to achieve high prediction accuracy, training stability, and computational efficiency.

Evaluation of the results showed that the created model has a high potential for integration into decision support systems in the field of spatial planning, where operational assessment of the availability of objects in a dynamic environment is required.

The obtained results provide a basis for further research aimed at scaling the approach to larger regions, integration with satellite monitoring data, and extending the functionality of the models to work with multilayer spatial systems.

Practical Significance: The obtained results are of practical importance for the tasks of spatial analysis and prediction of the availability of objects in spatially distributed systems. The proposed approach to the formation of the input feature set using OpenStreetMap geographic information data, data normalization by the Min-Max method, and the construction of five architectures of deep neural networks (FNN, CNN, RNN, GRU, LSTM) ensured high quality of model training.

Among all the architectures, the basic recurrent neural network (RNN) model demonstrated the best results in terms of the main accuracy metrics. Further optimization of the RNN architecture by improving the network structure, applying regularization techniques, and tuning hyperparameters allowed us to further improve the results, achieving a high level of agreement between predictions and actual data.

The proposed solutions can be applied in a wide range of practical tasks related to spatial planning, optimization of facility availability, and predictive analysis in management decision-making systems.

Recommendations for Further Research: Based on the findings, it is advisable to conduct further research to expand the set of spatial and infrastructural features to improve the accuracy of models for predicting the accessibility of objects. Particular attention should be paid to improving the methods of accounting for the dynamics of changes in the transport network, the availability of communication routes, and the topological features of spatial structures. It is also promising to develop approaches to integrating deep learning models with geographic information systems to ensure automated updating of transport data in real time. In addition, it is advisable to study the possibilities of using ensemble modeling methods to increase the stability and reliability of forecasts in different spatial and temporal scenarios.

Scientific Novelty: The study proposes a new approach to predicting the availability of objects in spatially distributed systems based on the integration of

deep learning and spatial data analysis methods. For the first time for this task, the optimization of the basic architecture of the recurrent neural network (RNN) is considered by introducing additional recurrent layers, applying Dropout regularization mechanisms, and increasing the number of neurons in the hidden layers. The proposed improvements allowed for significant improvement in the main accuracy indicators of the model without signs of overtraining, ensuring stability and high generalization ability in predicting complex spatial dependencies.

Prospects for Further Research: A promising direction for further research is to scale the proposed methodology to analyze large areas with various spatial and infrastructural characteristics. It is also advisable to study in depth the possibilities of using machine learning methods to automatically assess the importance of features that affect the prediction of object availability in spatially distributed systems. Particular attention should be paid to the development of new model performance metrics that will allow for a more comprehensive assessment of the quality of predictions in multidimensional spatial environments. The integration of neural network approaches with geographic information systems opens up prospects for creating adaptive analytical platforms that can quickly respond to changing spatial conditions and support real-time decision-making.

ACKNOWLEDGEMENTS

The research was carried out within the framework of the research work of the Department of Information Technologies and Electronic Communications Systems of Lviv State University of Life Safety on the topic “Information Technologies for Project Management of Regional Life Safety Systems Development” (No. 0119U002950). Also, the work was carried out in accordance with the thematic plan of comprehensive research work of the Faculty of Mechanics, Energy and Information Technologies of the Lviv National University of Environmental Management on the topic “Development of innovative information, managed, resource-saving systems, technologies and technical means for agrotechnics, industrial production and its energy supply” for 2021–2025 (No. 0121U109289), within the section “Substantiation of models, methods and approaches to the management of integrated projects and programs”.

REFERENCES

1. Siriwardena A.N., Patel G., Botan V., Rowan E., Spaight R. Community First Responders' role in the current and future rural health and care workforce: a mixed-methods study, *Health and Social Care Delivery Research*, 2024, Vol. 12(18), pp. 1–101. Mode of access: <https://doi.org/10.3310/JYRT8674>.
2. Ratushny R., Horodetskyi I., Molchak Y., Grabovets V. The configurations coordination of the projects products of development of the community fire extinguishing systems with the project environment, *CEUR Workshop*

- Proceedings*, 2021, Vol. 2851, pp. 238–248, Mode of access: <http://ceur-ws.org/Vol-2851/paper22.pdf>.
3. Hou C., Wu H. Rescuer, decision maker, and breadwinner: Women's predominant leadership across the post-Wenchuan earthquake efforts in rural areas, Sichuan, China. *Safety Science*, 2020, Vol. 125, 104623, Mode of access: <https://doi.org/10.1016/j.ssci.2020.104623>.
4. Sarmadi K., Amiri-Aref M. A distributionally robust optimisation with joint chance constraints approach for location-routing problem in urban search and rescue operations. *Computers and Operations Research*, 2025, Vol. 180, 107051, Mode of access: <https://doi.org/10.1016/j.cor.2025.107051>
5. Tryhuba A., Ratushny R., Tryhuba I., Koval N., Androshchuk I. The model of projects creation of the fire extinguishing systems in community territories. *Acta Universitatis Agriculturae et Silviculturae Mendelianae Brunensis*, 2020, Vol. 68(2), pp. 419–431, Mode of access: <https://doi.org/10.11118/actaun202068020419>
6. Makino H., Hatanaka M., Abe S., Takahashi M., Kinoshita H. Web-GIS-based emergency rescue to track triage information – System configuration and experimental results. 2012 Ubiquitous Positioning, *Indoor Navigation, and Location Based Service*, UPINLBS 2012, 2012, 6409773, Mode of access: <https://doi.org/10.1109/UPINLBS.2012.6409773>
7. Jaljolie R., Dror T., Siriba D.N., Dalyot S. Evaluating current ethical values of OpenStreetMap using value sensitive design, *Geo-Spatial Information Science*, 2023, Vol. 26(3), pp. 362–378, Mode of access: <https://doi.org/10.1080/10095020.2022.2087048>
8. Tryhuba A., Zachko O., Grabovets V., Pavlova I., Rudynets M. Examining the effect of production conditions at territorial logistic systems of milk harvesting on the parameters of a fleet of specialized road tanks, *Eastern-European Journal of Enterprise Technologies*, 2018, Vol. 5(3–95), pp. 59–69, Mode of access: <https://doi.org/10.15587/1729-4061.2018.142227>
9. Böckling M., Paulheim H., Detzler S. A Planet Scale Spatial-Temporal Knowledge Graph Based On OpenStreetMap And H3 Grid. *CEUR Workshop Proceedings*, 2024, Vol. 3743, pp. 60–72, Mode of access: <http://ceur-ws.org/Vol-3743/paper6.pdf>
10. Protaziuk G., Piątkowski R., Bembenik R. Modelling OpenStreetMap Data for Determination of the Fastest Route Under Varying Driving Conditions. *Studies in Big Data*, 2019, Vol. 40, pp. 53–71, Mode of access: https://doi.org/10.1007/978-3-319-77604-0_5
11. Tryhuba A., Ratushny R., Bashynsky O., Chubyk R., Bordun I. Planning of territorial location of fire-rescue formations in administrative territory development projects, *CEUR Workshop Proceedings*, 2020, Vol. 2565, pp. 93–105, Mode of access: <http://ceur-ws.org/Vol-2565/paper9.pdf>
12. Rani G., Siddiqui N. A., Yadav M., Ansari S. Hierarchical integrated spatial risk assessment model of fire hazard for the core city areas in India. *Land Use Policy*, 2023, Vol. 126, 106536, Mode of access: <https://doi.org/10.1016/j.landusepol.2023.106536>
13. Shahparvari S., Fadaki M., Chhetri P. Spatial accessibility of fire stations for enhancing operational response in Melbourne, *Fire Safety Journal*, 2020, Vol. 117, 103149, Mode of access: <https://doi.org/10.1016/j.firesaf.2020.103149>
14. Mao K., Chen Y., Wu G., Huang J., Yang W., Xia Z. Measuring spatial accessibility of urban fire services using historical fire incidents in Nanjing, China, *ISPRS International Journal of Geo-Information*, 2020, Vol. 9(10), 585, Mode of access: <https://doi.org/10.3390/ijgi9100585>
15. Tao Z., Cheng Y., Liu J. Hierarchical two-step floating catchment area (2SFCA) method: Measuring the spatial accessibility to hierarchical healthcare facilities in Shenzhen, China, *International Journal for Equity in Health*, 2020, Vol. 19(1), 164, Mode of access: <https://doi.org/10.1186/s12939-020-01280-7>
16. Tryhuba A., Tryhuba I., Ftoma O., Boyarchuk O. Method of quantitative evaluation of the risk of benefits for investors of fodder-producing cooperatives, *International Scientific and Technical Conference on Computer Sciences and Information Technologies*, 2019, Vol. 3, pp. 55–58, Mode of access: <https://doi.org/10.1109/STC-CSIT.2019.8929788>
17. Koval N., Kondysiuk I., Grabovets V., Onyshchuk V. Forecasting the fund of time for performance of works in hybrid projects using machine training technologies. *CEUR Workshop Proceedings*, 2021, Vol. 2917, pp. 196–206, Mode of access: <http://ceur-ws.org/Vol-2917/paper18.pdf>
18. Xu Y., Zhou C., Hu B. Measuring the accessibility of emergency shelters based on an improved two-step floating catchment area model, *International Journal of Digital Earth*, 2025, Vol. 18(1), Mode of access: <https://doi.org/10.1080/17538947.2025.2479864>
19. Chen Z., Tang L., Guo X., Zheng G. A self-supervised detection method for mixed urban functions based on trajectory temporal image. *Computers, Environment and Urban Systems*, 2024, Vol. 110, 102113, Mode of access: <https://doi.org/10.1016/j.compenvurbsys.2024.102113>
20. Chen H., Lan C., Song J., Xia J., Yokoya N. ObjFormer: Learning land-cover changes from paired OSM data and optical high-resolution imagery via object-guided transformer, *IEEE Transactions on Geoscience and Remote Sensing*, 2024, Vol. 62, 4408522, Mode of access: <https://doi.org/10.1109/TGRS.2024.3410389>
21. Liu Z. Identifying urban land use social functional units: a case study using OSM data, *International Journal of Digital Earth*, 2021, Vol. 14(12), pp. 1798–1817, Mode of access: <https://doi.org/10.1080/17538947.2021.1988161>
22. Tryhuba A., Hutsol T., Tryhuba I., Tabor S., Kwasniewski D. Risk assessment of investments in projects of production of raw materials for bioethanol, *Processes*, 2020, Vol. 9(1), pp. 1–12, Mode of access: <https://doi.org/10.3390/pr9010012>
23. Jayapriya J., Vinay M. Deep learning algorithms comparison for multiple biological sequences alignment. *Lecture Notes in Networks and Systems*, 2023, Vol. 720, P. 563–575, Mode of access: https://doi.org/10.1007/978-981-99-3761-5_50
24. Chen D. Research on network information security applications based on deep learning algorithms. 2024 5th *International Conference on Computer Vision, Image and Deep Learning (CVIDL)*, 2024, pp. 1160–1163, Mode of access: <https://doi.org/10.1109/CVIDL62147.2024.10603692>
25. Bashynsky O., Hutsol T., Rozkosz A., Prokopova O. Justification of parameters of the energy supply system of agricultural enterprises with using wind power installations, *E3S Web of Conferences*, 2020, Vol. 154, 06001, Mode of access: <https://doi.org/10.1051/e3sconf/202015406001>
26. Hutsol T., Kuboń M., Hohol T., Tomaszewska-Górecka W. Taxonomy and stakeholder risk management in integrated projects of the European Green Deal, *Energies*, 2022, Vol.

- 15(6), 2015, Mode of access: <https://doi.org/10.3390/en15062015>
27. Korstanje J. Machine Learning on Geographical Data Using Python: Introduction into Geodata with Applications and Use Cases. 2022, P. 1–230, Mode of access: <https://doi.org/10.1007/978-1-4842-8287-8>
28. Bashynsky O., Garasymchuk I., Vilchinska D., Dubik V. Research of the variable natural potential of the wind and energy in the northern strip of the Ukrainian Carpathians. *E3S Web of Conferences*, 2020, Vol. 154, 06002, Mode of access: <https://doi.org/10.1051/e3sconf/202015406002>
29. Omarkhanova D., Oralbekova Z. Interpretation of georadar data based on machine learning technologies, *EUREKA: Physics and Engineering*, 2024, (4), pp. 193–204, Mode of access: <https://doi.org/10.21303/2461-4262.2024.003289>
30. Araujo A. S., de Queiroz A. P. Spatial Characterization and Mapping of Gated Communities, *ISPRS International Journal of Geo-Information*, 2018, Vol. 7(7), 248, Mode of access: <https://doi.org/10.3390/ijgi7070248>
31. Suchenwirth L., Forster M., Lang F., Kleinschmit B. Estimation and mapping of carbon stocks in riparian forests by using a machine learning approach with multiple geodata. *Photogrammetrie, Fernerkundung, Geoinformation*, 2013, (4), pp. 333–349, Mode of access: <https://doi.org/10.1127/1432-8364/2013/0181>
32. Tryhuba A., Boyarchuk V., Tryhuba I., Pavlikha N., Kovalchuk N. Study of the impact of the volume of investments in agrarian projects on the risk of their value. *CEUR Workshop Proceedings*, 2021, Vol. 2851, P. 303–313, Mode of access: <https://ceur-ws.org/Vol-2851/paper28.pdf>
33. Tryhuba A., Boyarchuk V., Tryhuba I., Tymochko V., Bondarchuk S. Model of assessment of the risk of investing in the projects of production of biofuel raw materials. *International Scientific and Technical Conference on Computer Sciences and Information Technologies*, 2020, Vol. 2, P. 151–154, Mode of access: <https://doi.org/10.1109/CSIT49958.2020.9322024>
34. Czarnecka K., Kuchcik M., Baranowski J. Spatial development indicators as a tool to determine thermal conditions in an urban environment. *Sustainable Cities and Society*, 2024, Vol. 100, 105014, Mode of access: <https://doi.org/10.1016/j.scs.2023.105014>
35. Nemchinov D.M. The assessment of the required level of road and street network development in localities and conurbations (city agglomeration). *6th Transport Research Arena*, April 18–21, 2016, Mode of access: <https://doi.org/10.1016/j.trpro.2016.05.135>
36. Yuxin Y., Guangbin W., Yubin M., Haoxuan D. Data augmentation in training deep learning models for malware family classification. *Proceedings – International Conference on Machine Learning and Cybernetics*, 2021, December, Mode of access: <https://doi.org/10.1109/ICMLC54886.2021.9737271>
37. Ouareb S., Boutteau R., Romeo K., Ragot N., Duval F. Industrial object detection: Leveraging synthetic data for training deep learning models. *Lecture Notes in Business Information Processing*, 2024, Vol. 507, pp. 200–212, Mode of access: https://doi.org/10.1007/978-3-031-58113-7_17
38. Park J.-H., Kim Y.-S., Seo H., Cho Y.-J. Analysis of training deep learning models for PCB defect detection. *Sensors*, 2023, Vol. 23(5), 2766, Mode of access: <https://doi.org/10.3390/s23052766>
39. Tryhuba A., Kondysiuk I., Tryhuba I., Boiarchuk O., Tatomyr A. Intellectual information system for formation of portfolio projects of motor transport enterprises. *CEUR Workshop Proceedings*, 2022, Vol. 3109, pp. 44–52, Mode of access: <https://ceur-ws.org/Vol-3109/paper7.pdf>
40. Basyuk T., Vasyliuk A., Lytvyn V., Vlasenko O. Features of designing and implementing an information system for studying and determining the level of foreign language proficiency. *Proceedings of the 4th International Workshop on Modern Machine Learning Technologies and Data Science Workshop MoMLT&DS 2022*, 2022, pp. 212–225, Mode of access: <https://ceur-ws.org/Vol-3312/paper18.pdf>

Received 21.04.2025.
Accepted 30.09.2025.

УДК 004.85:528.9:614.84

ОПТИМІЗОВАНА МОДЕЛЬ ПРОГНОЗУВАННЯ ДОСТУПНОСТІ ОБ'ЄКТІВ НА ОСНОВІ ГЛИБИННОГО НАВЧАННЯ І ГЕОПРОСТОРОВИХ ОЗНАК

Тригуба А. М. – д-р техн. наук, професор, завідувач кафедри інформаційних технологій Львівського національного університету природокористування, Львів, Україна.

Ратушний Р. Т. – д-р техн. наук, професор, професор кафедри права та менеджменту у сфері цивільного захисту Львівського державного університету безпеки життєдіяльності, Львів, Україна.

Коваль Л. С. – ад'юнкт кафедри інформаційних технологій та систем електронних комунікацій Львівського державного університету безпеки життєдіяльності, Львів, Україна.

Ратушний А. Р. – ад'юнкт кафедри інформаційних технологій та систем електронних комунікацій Львівського державного університету безпеки життєдіяльності, Львів, Україна.

АНОТАЦІЯ

Актуальність. На сьогодні прогнозування доступності об'єктів у просторово-розподілених системах залишається одним із тих напрямів інформатики, який постійно привертає увагу дослідників. Причин для цього чимало. Спостерігається збільшення обсягів просторової інформації. Появляються нові типи інфраструктурних мереж, а також потреба в оперативному прийнятті рішень у мінливих умовах. Водночас, традиційні методи аналізу не завжди справляються із завданнями обробки багатовимірних даних. Особливо це актуально, коли йдеться про складні або нестабільні середовища. Це відкриває можливості для застосування методів глибинного навчання, які демонструють високу ефективність там, де класичні підходи не працюють.

Мета роботи. Дослідження спрямоване на оптимізацію моделі прогнозування доступності об'єктів у просторово-розподілених системах шляхом визначення ефективної архітектури глибинного навчання, що використовує просторові та інфраструктурні ознаки для підвищення точності прогнозування та узагальнюючої здатності.

Методи. Для досягнення поставленої мети було застосовано архітектури глибинного навчання, серед яких моделі прямого розповсюдження (FNN), згорткові нейронні мережі (CNN) та рекурентні нейронні мережі (RNN, GRU, LSTM). Під час моделювання використовувалися методи нормалізації даних, регуляризації навчання, а також комплексна система оцінювання точності прогнозів за допомогою середньої квадратичної помилки, середньої абсолютної помилки та коефіцієнта детермінації.

Результати. Під виконання дослідження було побудовано оптимізовану архітектуру рекурентної нейронної мережі, що включає комбінацію двох рекурентних шарів, шарів регуляризації Dropout та повнозв'язного шару. Проведений аналіз показав, що запропонована модель забезпечує високу точність прогнозування доступності об'єктів, демонструючи стабільність роботи на широкому спектрі просторових даних. Співставлення фактичних та прогнозованих значень підтвердило ефективність запропонованого рішення.

Висновки. Запропонований підхід до побудови оптимізованої моделі глибинного навчання для прогнозування доступності об'єктів забезпечує високий рівень узагальнення та точності, що створює передумови для її використання у системах інтелектуальної підтримки прийняття рішень у просторово-розподілених середовищах.

КЛЮЧОВІ СЛОВА: глибоке навчання, прогнозування, доступність об'єктів, геопросторові дані, оптимізація RNN, інтелектуальні системи.

ЛІТЕРАТУРА

1. Siriwardena A. N. Community First Responders' role in the current and future rural health and care workforce: a mixed-methods study / [A. N. Siriwardena, G. Patel, V. Botan et al.] // Health and Social Care Delivery Research. – 2024. – Vol. 12(18). – P. 1–101. – Mode of access: <https://doi.org/10.3310/JYRT8674>.
2. Tryhuba A. The configurations coordination of the projects products of development of the community fire extinguishing systems with the project environment / [A. Tryhuba, R. Ratushny, I. Horodetsky et al.] // CEUR Workshop Proceedings. – 2021. – Vol. 2851. – P. 238–248. – Mode of access: <http://ceur-ws.org/Vol-2851/paper22.pdf>.
3. Hou C. Rescuer, decision maker, and breadwinner: Women's predominant leadership across the post-Wenchuan earthquake efforts in rural areas, Sichuan, China / C. Hou, H. Wu // Safety Science. – 2020. – Vol. 125. – P. 104623. – Mode of access: <https://doi.org/10.1016/j.ssci.2020.104623>.
4. Sarmadi K. A distributionally robust optimisation with joint chance constraints approach for location-routing problem in urban search and rescue operations / K. Sarmadi, M. Amir-Aref // Computers and Operations Research. – 2025. – Vol. 180. – P. 107051. – Mode of access: <https://doi.org/10.1016/j.cor.2025.107051>.
5. The model of projects creation of the fire extinguishing systems in community territories / [A. Tryhuba, R. Ratushny, I. Tryhuba et al.] // Acta Universitatis Agriculturae et Silviculturae Mendelianae Brunensis. – 2020. – Vol. 68(2). – P. 419–431. – Mode of access: <https://doi.org/10.1118/actaun202068020419>.
6. Makino H. Web-GIS-based emergency rescue to track triage information – System configuration and experimental results / [H. Makino, M. Hatanaka, S. Abe et al.] // 2012 Ubiquitous Positioning, Indoor Navigation, and Location Based Service (UPINLBS). – 2012. – P. 6409773. – Mode of access: <https://doi.org/10.1109/UPINLBS.2012.6409773>.
7. Evaluating current ethical values of OpenStreetMap using value sensitive design / [R. Jaljolie, T. Dror, D. N. Siriba, S. Dalyot] // Geo-Spatial Information Science. – 2023. – Vol. 26(3). – P. 362–378. – Mode of access: <https://doi.org/10.1080/10095020.2022.2087048>.
8. Examining the effect of production conditions at territorial logistic systems of milk harvesting on the parameters of a fleet of specialized road tanks / [A. Tryhuba, O. Zachko, V. Grabovets et al.] // Eastern-European Journal of Enterprise Technologies. – 2018. – Vol. 5(3–95). – P. 59–69. – Mode of access: <https://doi.org/10.15587/1729-4061.2018.142227>.
9. Böckling M. A Planet Scale Spatial-Temporal Knowledge Graph Based On OpenStreetMap And H3 Grid / M. Böckling, H. Paulheim, S. Detzler // CEUR Workshop Proceedings. – 2024. – Vol. 3743. – P. 60–72. – Mode of access: <http://ceur-ws.org/Vol-3743/paper6.pdf>.
10. Protaziuk G. Modelling OpenStreetMap Data for Determination of the Fastest Route Under Varying Driving Conditions / G. Protaziuk, R. Piątkowski, R. Bembenik // Studies in Big Data. – 2019. – Vol. 40. – P. 53–71. – Mode of access: https://doi.org/10.1007/978-3-319-77604-0_5.
11. Planning of territorial location of fire-rescue formations in administrative territory development projects / [A. Tryhuba, R. Ratushny, O. Bashynsky et al.] // CEUR Workshop Proceedings. – 2020. – Vol. 2565. – P. 93–105. – Mode of access: <http://ceur-ws.org/Vol-2565/paper9.pdf>.
12. Hierarchical integrated spatial risk assessment model of fire hazard for the core city areas in India / [G. Rani, N. A. Siddiqui, M. Yadav, S. Ansari] // Land Use Policy. – 2023. – Vol. 126. – P. 106536. – Mode of access: <https://doi.org/10.1016/j.landusepol.2023.106536>.
13. Shahparvari S. Spatial accessibility of fire stations for enhancing operational response in Melbourne / S. Shahparvari, M. Fadaki, P. Chhetri // Fire Safety Journal. – 2020. – Vol. 117. – P. 103149. – Mode of access: <https://doi.org/10.1016/j.firesaf.2020.103149>.
14. Measuring spatial accessibility of urban fire services using historical fire incidents in Nanjing, China / [K. Mao, Y. Chen, G. Wu, J. Huang et al.] // ISPRS International Journal of Geo-Information. – 2020. – Vol. 9(10). – P. 585. – Mode of access: <https://doi.org/10.3390/ijgi9100585>.
15. Tao Z. Hierarchical two-step floating catchment area (2SFCA) method: Measuring the spatial accessibility to hierarchical healthcare facilities in Shenzhen, China / Z. Tao, Y. Cheng, J. Liu // International Journal for Equity in Health. – 2020. – Vol. 19(1). – P. 164. – Mode of access: <https://doi.org/10.1186/s12939-020-01280-7>.
16. Method of quantitative evaluation of the risk of benefits for investors of fodder-producing cooperatives / [A. Tryhuba, I. Tryhuba, O. Ftoma, O. Boyarchuk] // International Scientific and Technical Conference on Computer Sciences and Information Technologies. – 2019. – Vol. 3. – P. 55–58. – Mode of access: <https://doi.org/10.1109/STC-CSIT.2019.8929788>.
17. Forecasting the fund of time for performance of works in hybrid projects using machine training technologies / [N. Koval, I. Kondysiuk, V. Grabovets, V. Onyshchuk] // CEUR Workshop Proceedings. – 2021. – Vol. 2917. – P. 196–206. – Mode of access: <http://ceur-ws.org/Vol-2917/paper18.pdf>.

18. Xu Y. Measuring the accessibility of emergency shelters based on an improved two-step floating catchment area model / Y. Xu, C. Zhou, B. Hu // *International Journal of Digital Earth*. – 2025. – Vol. 18(1). – Mode of access: <https://doi.org/10.1080/17538947.2025.2479864>.
19. A self-supervised detection method for mixed urban functions based on trajectory temporal image / [Z. Chen, L. Tang, X. Guo, G. Zheng] // *Computers, Environment and Urban Systems*. – 2024. – Vol. 110. – P. 102113. – Mode of access: <https://doi.org/10.1016/j.compenvurbsys.2024.102113>
20. ObjFormer: Learning land-cover changes from paired OSM data and optical high-resolution imagery via object-guided transformer / [H. Chen, C. Lan, J. Song et al.], // *IEEE Transactions on Geoscience and Remote Sensing*. – 2024. – Vol. 62. – P. 4408522. – Mode of access: <https://doi.org/10.1109/TGRS.2024.3410389>.
21. Liu Z. Identifying urban land use social functional units: a case study using OSM data / Z. Liu // *International Journal of Digital Earth*. – 2021. – Vol. 14(12). – P. 1798–1817. – Mode of access: <https://doi.org/10.1080/17538947.2021.1988161>.
22. Risk assessment of investments in projects of production of raw materials for bioethanol / [A. Tryhuba, T. Hutsol, I. Tryhuba et al.] // *Processes*. – 2020. – Vol. 9(1). – P. 1–12. – Mode of access: <https://doi.org/10.3390/pr9010012>.
23. Jayapriya J. Deep learning algorithms comparison for multiple biological sequences alignment / J. Jayapriya, M. Vinay // *Lecture Notes in Networks and Systems*. – 2023. – Vol. 720. – P. 563–575. – Mode of access: https://doi.org/10.1007/978-981-99-3761-5_50.
24. Chen D. Research on network information security applications based on deep learning algorithms / D. Chen // *2024 5th International Conference on Computer Vision, Image and Deep Learning (CVIDL)*. – 2024. – P. 1160–1163. – Mode of access: <https://doi.org/10.1109/CVIDL62147.2024.10603692>.
25. Justification of parameters of the energy supply system of agricultural enterprises with using wind power installations / [O. Bashynsky, T. Hutsol, A. Rozkosz, O. Prokopova] // *E3S Web of Conferences*. – 2020. – Vol. 154. – P. 06001. – Mode of access: <https://doi.org/10.1051/e3sconf/202015406001>.
26. Taxonomy and stakeholder risk management in integrated projects of the European Green Deal / [T. Hutsol, M. Kuboń, T. Hohol, W. Tomaszewska-Górecka] // *Energies*. – 2022. – Vol. 15(6). – P. 2015. – Mode of access: <https://doi.org/10.3390/en15062015>.
27. Korstanje J. Machine Learning on Geographical Data Using Python: Introduction into Geodata with Applications and Use Cases / J. Korstanje. – 2022. – P. 1–230. – Mode of access: <https://doi.org/10.1007/978-1-4842-8287-8>.
28. Research of the variable natural potential of the wind and energy in the northern strip of the Ukrainian Carpathians / [O. Bashynsky, I. Garasymchuk, D. Vilchinska, V. Dubik] // *E3S Web of Conferences*. – 2020. – Vol. 154. – P. 06002. – Mode of access: <https://doi.org/10.1051/e3sconf/202015406002>.
29. Omarmkhanova D. Interpretation of georadar data based on machine learning technologies / D. Omarmkhanova, Z. Oralbekova // *EUREKA: Physics and Engineering*. – 2024. – (4). – P. 193–204. – Mode of access: <https://doi.org/10.21303/2461-4262.2024.003289>.
30. Araujo A.S. Spatial Characterization and Mapping of Gated Communities / A. S. Araujo, A. P. de Queiroz // *ISPRS International Journal of Geo-Information*. – 2018. – Vol. 7(7). – P. 248. – Mode of access: <https://doi.org/10.3390/ijgi7070248>.
31. Estimation and mapping of carbon stocks in riparian forests by using a machine learning approach with multiple geodata / [L. Suchenwirth, M. Forster, F. Lang, B. Kleinschmit] // *Photogrammetrie, Fernerkundung, Geoinformation*. – 2013. – (4). – P. 333–349. – Mode of access: <https://doi.org/10.1127/1432-8364/2013/0181>.
32. Study of the impact of the volume of investments in agrarian projects on the risk of their value / [A. Tryhuba, V. Boyarchuk, I. Tryhuba et al.] // *CEUR Workshop Proceedings*. – 2021. – Vol. 2851. – P. 303–313. – Mode of access: <https://ceur-ws.org/Vol-2851/paper28.pdf>.
33. Model of assessment of the risk of investing in the projects of production of biofuel raw materials / [A. Tryhuba, V. Boyarchuk, I. Tryhuba et al.] // *International Scientific and Technical Conference on Computer Sciences and Information Technologies*. – 2020. – Vol. 2. – P. 151–154. – Mode of access: <https://doi.org/10.1109/CSIT49958.2020.9322024>.
34. Czarnecka K. Spatial development indicators as a tool to determine thermal conditions in an urban environment / K. Czarnecka, M. Kuchcik, J. Baranowski // *Sustainable Cities and Society*. – 2024. – Vol. 100. – P. 105014. – Mode of access: <https://doi.org/10.1016/j.scs.2023.105014>.
35. Nemchinov D. M. The assessment of the required level of road and street network development in localities and conurbations (city agglomeration) / D. M. Nemchinov // *6th Transport Research Arena*. – 2016. – Mode of access: <https://doi.org/10.1016/j.trpro.2016.05.135>.
36. Data augmentation in training deep learning models for malware family classification / [Y. Yuxin, W. Guangbin, M. Yubin, D. Haoxuan] // *Proceedings – International Conference on Machine Learning and Cybernetics*. – 2021. – December. – Mode of access: <https://doi.org/10.1109/ICMLC54886.2021.9737271>.
37. Industrial object detection: Leveraging synthetic data for training deep learning models / [S. Ouarab, R. Bouteau, K. Romeo et al.] // *Lecture Notes in Business Information Processing*. – 2024. – Vol. 507. – P. 200–212. – Mode of access: https://doi.org/10.1007/978-3-031-58113-7_17.
38. Analysis of training deep learning models for PCB defect detection / [J.-H. Park, Y.-S. Kim, H. Seo, Y.-J. Cho] // *Sensors*. – 2023. – Vol. 23(5). – P. 2766. – Mode of access: <https://doi.org/10.3390/s23052766>.
39. Intellectual information system for formation of portfolio projects of motor transport enterprises / [A. Tryhuba, I. Kondysiuk, I. Tryhuba et al.] // *CEUR Workshop Proceedings*. – 2022. – Vol. 3109. – P. 44–52. – Mode of access: <https://ceur-ws.org/Vol-3109/paper7.pdf>.
40. Features of designing and implementing an information system for studying and determining the level of foreign language proficiency / [T. Basyuk, A. Vasyliuk, V. Lytvyn, O. Vlasenko] // *Proceedings of the 4th International Workshop on Modern Machine Learning Technologies and Data Science Workshop MoMLLeT&DS 2022*. – 2022. – P. 212–225. – Mode of access: <https://ceur-ws.org/Vol-3312/paper18.pdf>.

ПРОГРЕСИВНІ ІНФОРМАЦІЙНІ ТЕХНОЛОГІЇ

PROGRESSIVE INFORMATION TECHNOLOGIES

UDC 004.932:612.76

PHOTOGRAMMETRIC MOTION CAPTURE SUBSYSTEM FOR CHANGE OF BODY POSITION ANALYSIS IN THE FRONTAL AND SAGITTAL PLANES

Barkovska O. Y. – PhD, Associate Professor, Electronic Computers Department, Kharkiv National University of Radio Electronics, Kharkiv, Ukraine.

Kovalenko A. A. – Dr. Sc., Professor, Head of Electronic Computers Department, Kharkiv National University of Radio Electronics, Kharkiv, Ukraine.

Diachenko V. O. – Lecturer, Electronic Computers Department, Kharkiv National University of Radio Electronics, Kharkiv, Ukraine.

Bukharova L. D. – Master's student of the Department of Electronic Computers, Kharkiv National University of Radio Electronics, Kharkiv, Ukraine.

Korobko V. Y. – Master's student of the Department of Electronic Computers, Kharkiv National University of Radio Electronics, Kharkiv, Ukraine.

ABSTRACT

Context. The increase in other orthopedic injuries, particularly among military personnel, requires new innovative solutions to assess posture changes and monitor rehabilitation effectiveness. Existing systems have limitations in terms of portability, cost, and flexibility of use, which necessitates the development of hybrid systems that combine computer vision and sensory analysis methods..

Objective. To assess the effectiveness of combining non-contact computer vision and accelerometric sensors for detecting changes in human posture under different lighting, background, and movement speeds

Method. The study implemented a photogrammetric subsystem that includes MediaPipe Holistic for markerless tracking of key body points and WitMotion WT9011DCL-BT50 accelerometers for analyzing inertial motion parameters. The system model was built in IDEF0 notation. The accuracy was assessed by comparing the obtained values of the blade inclination angle and the asymmetry coefficient with the specified norms

Results. The combined use of visual and sensory data made it possible to reduce the error to 5.05% under normal conditions and ensure the stability of the results under conditions of changes in the external environment. Image modification (contrast, noise filtering) increased the accuracy of computer vision. Threshold values of the asymmetry coefficient corresponding to normal, mild and severe postural disorders were determined.

Conclusions. The proposed system demonstrates high potential effectiveness in telemedical rehabilitation support for patients with musculoskeletal disorders. Its practical significance lies in the creation of an affordable, portable, and accurate diagnostic and monitoring tool suitable for further integration into personalized medicine systems with built-in artificial intelligence modules.

KEYWORDS: rehabilitative orthopedics, motion capture, computer vision, MediaPipe Holistic, inertial sensors, postural analysis, gait monitoring, telemedicine, accelerometer.

ABBREVIATIONS

AI is artificial intelligence;
CLAHE is a contrast limited adaptive histogram equalization;
CSV is a comma-separated values file;
CP is cerebral palsy;
CT is computed tomography;
IDEF0 is an Integration Definition for Function Modeling;
IMU is an inertial measurement unit;
IPR is an individual rehabilitation plan;
MRI is a magnetic resonance imaging.
SW is a software;
US is an ultrasound.

NOMENCLATURE

θ_i is a sensor orientation in space;
 ω_a is a data acquisition frequency of the accelerometers;
 ω_b is a frame rate of the video recording;
 ω_i is an angular velocity;
 a_i is a linear acceleration;
 I_i is an input image at time t_i ;
 K_{asymm} is an asymmetry coefficient;
 K_{kyphosys} is a scapula plane tilt angle relative to the vertical axis;
 L is a tilt angle for the left side;
 P is a set of key body points;

P_a is a measurement error of the accelerometers;

P_b is an error of the markerless method;

P_{comb} is a combined measurement error;

p_j is a coordinate of the j -th key point;

R is a right side tilt angle;

t_i is a time moment corresponding to image I_i ;

Y is a horizontal component of accelerometer position;

Z is a vertical component of the accelerometer position.

INTRODUCTION

Rehabilitation orthopedics is an important area of modern medicine that concerned with the treatment and restoration of musculoskeletal functions in patients who have suffered injuries or have either congenital or acquired orthopedic problems. The primary objective of rehabilitation orthopedics is to help the patient return to a normal lifestyle, reduce mobility restrictions and improve the quality of life [1]. In modern conditions, rehabilitation

orthopedics faces challenges related to the accuracy of diagnostics, monitoring the progress of treatment and ensuring patient safety during therapy. One of the promising areas that can contribute to problem-solving is the implementation of computer vision systems and other machine learning methods [2].

As a result of computer vision techniques application in the practice of rehabilitation orthopedics, it becomes possible to ensure accurate monitoring of patient movements, analysis of their biomechanics and real-time feedback provision, which is important for rehabilitation processes [3]. In addition, the use of computer vision technologies can partially eliminate the physical presence of the orthopedic doctor, providing separate monitoring of the recovery process and supporting feedback to achieve optimal results in the recovery of patients [4].

Rehabilitation orthopedics is aimed at solving a wide range of musculoskeletal disorders (Table 1), each of which requires an individual approach and the use of different methods to achieve optimal results.

Table 1 – Key conditions addressed by rehabilitation orthopedics

Injury class	Injury example	Rehabilitation direction
1	2	3
Injuries to bones and joints	fractures, dislocations and sprains.	strengthening the muscles that support the joint.
Degenerative joint diseases	– osteoarthritis (arthrosis) – osteoporosis.	– improving mobility and muscle strength around the affected joint, physiotherapy, and reducing the load on the joint; – improving balance and coordination to reduce the risk of falls and injuries.
Spinal injuries and posture disorders.	– damage to the intervertebral discs (herniation, disc protrusion); – scoliosis, kyphosis, lordosis	– reduction of pain syndrome, restoring mobility and strengthening back muscles; – posture correction, strengthening the corset muscle and improving coordination.
Limb amputations	– rehabilitation after amputation; – complications prevention; – contracture.	– physical therapy and psychological support; – maintenance of joint mobility and limb functionality
Muscle and tendon injuries and diseases	– muscle tears and tendon injuries; – tendonitis and bursitis.	– restoring muscle strength, flexibility and endurance, which reduces the risk of reinjury; – stretching and strengthening of the affected areas exercises, as well as physiotherapy procedures.
Post-surgery rehabilitation	– rehabilitation after arthroplasty; – rehabilitation after osteosynthesis.	– restore limb functions, adapt to new movement conditions and return to an active life; – preserve the mobility of the surgical site and strengthen the surrounding muscles.
Neurological conditions with impaired musculoskeletal function	– strokes and paralysis – CP (cerebral palsy)	– coordination training, restoring balance and mobility, strengthening muscles and learning new movement skills; – rehabilitation of children with cerebral palsy includes special exercises to improve movement control and motility development.

Sadly, there is a need for affordable and effective rehabilitation tools in Ukraine, especially for combat veterans and people who have been injured as a result of military conflicts or accidents [5]. Computer vision systems allow for remote patient monitoring, which is important in conditions where many people do not have access to specialized rehabilitation centers. Therefore, the development and implementation of such technologies is not only relevant but also the most demanded task for the medical industry of Ukraine.

Thus, research in the field of computer vision for rehabilitation orthopedics is an extremely relevant and promising direction, especially for Ukraine [6–8]. The

implementation of such systems can significantly improve the quality of rehabilitation, reduce the workload on doctors and facilitate the recovery process for patients.

The idea of developing medical care technology that can be applied at the rehabilitation stage of military personnel and civilians who received surgical pathologies because of military actions in our country is justified by the unfortunate statistics that show that the number of orthoprosthetic patients and those in need of prosthetics and further physical rehabilitation has increased significantly recently (as of June 2023, 58,852 people applied for rehabilitation aids, of which 20,737 people received prosthetics).

Blast injuries, shrapnel and gunshot wounds are causes of surgical pathologies, which include pathologies of joints and bones that require surgical intervention (for example, joint endoprosthetics, orthoprosthetics of affected limbs) in order to restore the function of the affected limb and return the person to motor activity, improve the quality of life, psycho-emotional state, and become more independent in everyday life [9, 10].

A mandatory stage after such operations is rehabilitation and adaptation to the new "limb." Rehabilitation is a long, continuous, multi-stage process [11].

A common rehabilitation algorithm is:

- initial examination and assessment of the condition, review of medical history;
- agreeing on the treatment strategy, prognosis and client expectations;
- composition and implementation of an individual rehabilitation plan (IRP);
- results monitoring and the IPR adjustments if necessary;
- after achieving maximum results – support and guidance during the post-rehabilitation period, which involves psychological and physical adaptation, and the patient's return to a normal and quality life.

Despite the advantages, existing solutions still have such disadvantages as high cost, complexity in installation and use, focus on clinical settings, and limited data integration, which confirms the relevance and necessity of developing portable systems available for independent use, using the principles of telemedicine to monitor the rehabilitation process.

The aim of the work is to investigate the effectiveness of using a computer vision system in combination with electronic sensors in a photogrammetric motion capture subsystem for analyzing changes in body position relative to the frontal and sagittal planes in real time.

To achieve the objective, **the following tasks** must be solved:

- develop a prototype of a system that can track and analyze patient movements;
- to evaluate the effectiveness of using computer vision methods under different environmental conditions to determine the condition of joints during walking;
- to evaluate the effectiveness of using electronic sensors to determine the condition of joints during walking;
- develop a combined method for monitoring the position of the patient's joints during gait.

1 PROBLEM STATEMENT

Considered the problem of determining deviations in the position of a human body relative to the frontal and sagittal planes during movement or in statics based on a combination of computer vision data and inertial sensors.

Let $I = \{I_1, I_2, \dots, I_n\}$ be an ordered set of input images, where each image I_i contains an image of the patient in motion or in a static pose. For each image, using

computer vision algorithms, a set of coordinates of key body points $P_i = \{p_1, p_2, \dots, p_k\}$ is determined, where $p_j = (x_j, y_j) \in R^2$ is the projection of an anatomically significant point onto the image plane.

In parallel to each time point t_i corresponding to I_i , a vector of inertial sensor readings $S_i = (a_i, \omega_i, \theta_i)$, where $a_i \in R^3$ is the linear acceleration, $\omega_i \in R^3$ is the angular velocity, $\theta_i \in R^3$ is the sensor orientation in space, is received. Let $K_{\text{kyphosys}} \in R^3$ be the calculated angle of inclination of the scapula plane to the vertical axis in frame i , determined using the key points of the shoulder joints and the vertical vector. When $K_{\text{kyphosys}} \in [15^0, 30^0]$, the patient is considered to be in a normal state; exceeding this range is interpreted as the degree of kyphosis.

The asymmetry coefficient K_{asymm} is determined, which characterizes the deviation of the symmetry of the location of the joints of the upper or lower extremities at time t_i , as a function of the Euclidean distance between paired points, normalized to the anatomical width of the shoulders or pelvis. With values of $K_{\text{asymm}} < 5\%$, symmetry is considered normal; for $5\% \leq K_{\text{asymm}} < 15\%$, a mild form of scoliosis is recorded; and with $K_{\text{asymm}} \geq 15\%$ – pronounced asymmetry requires clinical intervention. To increase the accuracy of determining the deviation of the body position from the norm, the work proposes a combination of the two aforementioned approaches – optical and based on inertial sensors, which involves the use of a weighted combination of errors of both sources of information P_{comb} .

Thus, the input variables of the problem are the coordinates of key points of the body obtained using a computer vision algorithm; acceleration and orientation vectors from inertial sensors (accelerometers); environmental parameters: lighting level, background type, speed of movement.

The output variables are the numerical values of body position deviations, which are determined by the angular inclination of the shoulder blades relative to the vertical axis – K_{kyphosys} and the asymmetry coefficient (degree of scoliosis) – K_{asymm} based on accelerometers and the error in detecting deviations of the musculoskeletal system from the norm based on the optical marker-free method and the combined method of monitoring the position of the patient's joints during gait.

The problem has such limitations as fixed resolution and frame rate, a limited number of accelerometers, the presence of interference in the images, such as a change in lighting, the presence of a noisy background, and uneven gait.

2 REVIEW OF THE LITERATURE

Today, there are several approaches of using innovative technologies and artificial intelligence methods in rehabilitation orthopedics in the world. Among the most common are:

- camera-based systems – this approach involves using cameras to read the patient's movements. Specialized software analyses the video and determines the movement parameters. Well-known examples are technologies using Microsoft Kinect cameras or motion capture systems, which are often used in sports rehabilitation [12, 13];

- sensor and accelerometer systems – this approach involves the use of special sensors that are attached to the patient's body and track his movements. This method is less sensitive to external interference, but requires specialized equipment [14];

- integrated systems using artificial intelligence (AI) – such systems use machine learning algorithms to analyze and predict patient movements, allowing for an individual rehabilitation approach. AI can quickly process large amounts of data and draw conclusions that contribute to improving the quality of therapy [15];

- rehabilitation simulators with feedback systems – in some cases, special simulators with built-in cameras or sensors are used, which provide feedback to the patient about the exercise's accuracy. This allows the patient's movements to be corrected in real-time [16].

Each of these approaches has its advantages and disadvantages, but they all aim to improve the rehabilitation process, increase diagnostic accuracy, and create individual programs for each patient (Table 2).

Table 2 – Table describing the advantages and disadvantages of the methods

Method	Advantages	Disadvantages
Optical methods	High accuracy in marker systems, contactless data collection in markerless	Dependence on lighting, space limitations
Inertial Measurement Unit (IMU)	Compact, independent of lighting, effective in large spaces	Error accumulates over time, especially with long-term tracking
Magnetic methods.	Visibility-independent, suitable for indoor environments	The occurrence of obstacles in the environment
Ultrasonic methods	High accuracy in small spaces, no need for cameras	Dependence on acoustic conditions, limitations in large rooms
Mechanical methods	Precision joint tracking, suitable for medical and industrial applications	Uncomfortable for prolonged use, restrict natural movements

Among optical methods, marker and markerless methods can be distinguished. Marker optical methods use reflective or active markers placed on the body of the object. Cameras record the position of the markers and transmit this data for further processing. Markerless optical methods use computer vision algorithms to recognize characteristic points on the body without markers, in particular in three-dimensional space, which is more convenient and does not depend on markers. They are used in rehabilitation, virtual reality and sports.

The combination of inertial sensors and optical methods provides a visual analysis of motor activity and biomechanics of movements. It helps determine acceleration, angular displacements and limb positions, transferring them to a computer to build an accurate movement model. This allows the system to be used both indoors and outdoors without restrictions. The advantages are independence from special lighting and premises, the possibility of real-time use in open spaces, and support for accurate capture of even complex body movements. The disadvantages are higher cost than in conventional systems based on optical methods alone.

Such systems can significantly facilitate the work of medical professionals, increasing the accuracy of diagnostic and rehabilitation procedures, as well as reducing the time and cost of treating patients.

The combination of various analysis and monitoring of gait and posture methods in a single system allows for the work's high diagnostic and prognostic value.

3 MATERIALS AND METHODS

To ensure a clear understanding of the research process, it is necessary to describe the model at a basic and

detailed level, which will allow you to structure and divide the entire process into different stages for more detailed analysis. Describing the model in IDEF0 notation will help describe the resulting process in the form of a “black box” and a detailed sequence of sub-processes.

Consider the top level of the photogrammetric motion capture subsystem model for analyzing changes in body position relative to the frontal and sagittal planes (Fig. 1). This model takes into consideration aspects of input data, constraints and controlling influences, performers or participants, and output data.

The input data are video recordings of patients' body movements and accelerometer data on the patient's body during movement. The controlling influences and constraints are the evaluation rules for determining deviations from the norm and the external conditions of the experiment (lighting, background, patient's movement speed). The performer or participant of the process is the patient, and the output data or result is the patient receiving recommendations for correcting the deviation of one's musculoskeletal system. By decomposing the base-level model, we can obtain a detailed model level (Fig. 2), in which we can see that the primary process has been divided into four sub-processes, which will be elaborated in more detail.

The decomposed model of the photogrammetric motion capture subsystem for analyzing changes in body position relative to the frontal and sagittal planes includes the following modules:

- video processing module using MediaPipe Holistics, for which the patient is a participant, the controlling influences and constraints are external conditions (light-

ing, background, walking speed) and numerical values for determining deviations between key points, the output data is the processed video recording with detected deviations or their absence.

– module for processing data from accelerometers, where the patient is a participant, the controlling influences and limitations are external conditions (gait speed and location of accelerometers) and numerical values for measuring deviations from the norm due to the difference in acceleration indicators from accelerometers, the initial data is a sample of data from the movement of points on the patient's body during movement with a conclusion as to whether there are deviations or not;

– the combined data analysis module uses processed video and accelerometer data to improve the accuracy of patient-participated experiment data; the governing influ-

ences and limitations are numerical values for the combined data comparison to detect abnormalities, the output is the final conclusion, whether the patient has musculoskeletal abnormalities;

– module for recommendations provision based on the obtained conclusion of the detected deviations from the patient's norm, if any were detected, the initial data is an expert conclusion with recommendations for further actions, which is built based on the obtained results of automatic analysis and expert analysis by a doctor.

It is important to note that within the framework of the study of the motion capture subsystem for analyzing changes in body position, only a specialist can provide an expert opinion and further recommendations. Therefore, the main goal will be to describe and investigate the first three subprocesses of the model.

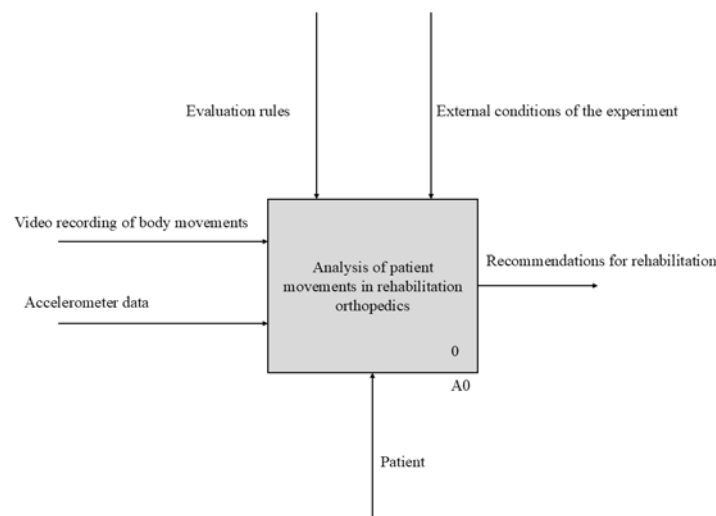


Figure 1 – Model of a photogrammetric motion capture subsystem for analyzing changes in body position relative to the frontal and sagittal planes. Top-Level (A-0)

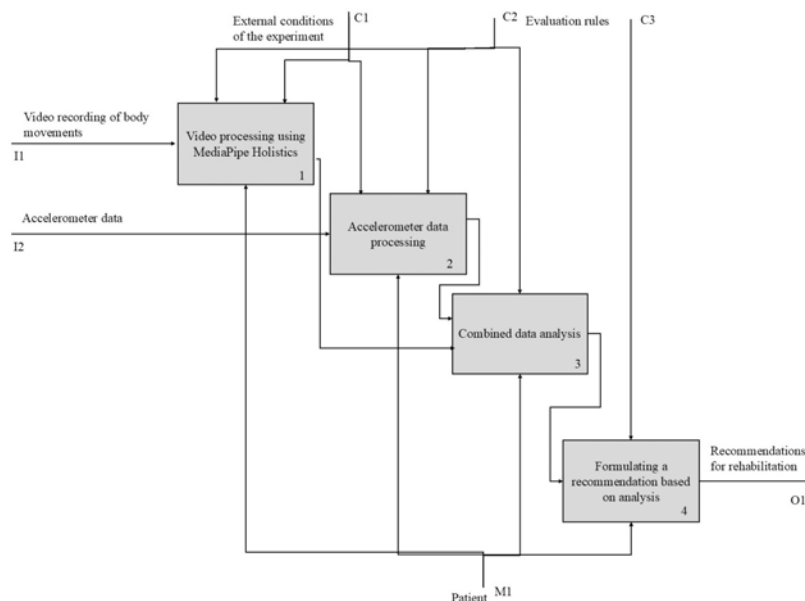


Figure 2 – Decomposition of the photogrammetric motion capture subsystem model for analyzing changes in body position relative to the frontal and sagittal planes (A-1)

To detect deviations from the norm of the human musculoskeletal system using computer vision methods and accelerometers, it is necessary to determine key indicators by which deviations from a healthy state can be distinguished. As key indicators, you can use:

- displacement of the body's center of mass from the vertical (determined based on key points of the body and allows assessing the patient's stability during movement);
- the angle of inclination of the body, shoulder blades and hip joints (evaluating the angle of inclination relative to the horizontal for each group of joints allows you to identify asymmetries and incorrect posture);
- symmetry and smoothness of limb movements (measured by comparing the amplitudes of movements of the right and left limbs, determined by analyzing the trajectories of movement of key points).

When working with computer vision methods, external conditions are important factors that can affect the accuracy of assessing a person's musculoskeletal system. The study focused on the following external factors to determine their impact on assessing the state of the musculoskeletal system: lighting, background type, and walking speed.

The hypothesis of the study is that the combination of MediaPipe Holistics and accelerometers will help to achieve greater accuracy and reliability of the analysis, but for the correct interpretation of the data, the results need to be synchronized. Depending on whether the recording is real-time or previously prepared and data sampling from the accelerometers, a synchronization tool is needed that will work in both cases. A unified timestamp system can be used to reconcile the data both pre-recorded and when working in real time.

The general algorithm for determining deviations from the norm of the human musculoskeletal system can be depicted in the form of a flowchart (Fig. 3).

The studied methods for determining the degree of deviation of the body position from the norm of the human musculoskeletal system relative to the frontal and sagittal planes are:

- using a markerless method of motion analysis by means of computer vision;
- using accelerometers to monitor acceleration and body tilt during movement;
- the combination of the markerless method and the use of accelerometers to analyze changes in accuracy rates.

As part of the research on the markerless method, the main attention was focused on the accuracy of detecting deviations from the norm and the time of data processing under different environmental conditions. The patient's movements were analyzed to detect kyphosis or scoliosis; for each deviation, separate groups of experiments were conducted with different input environmental conditions.

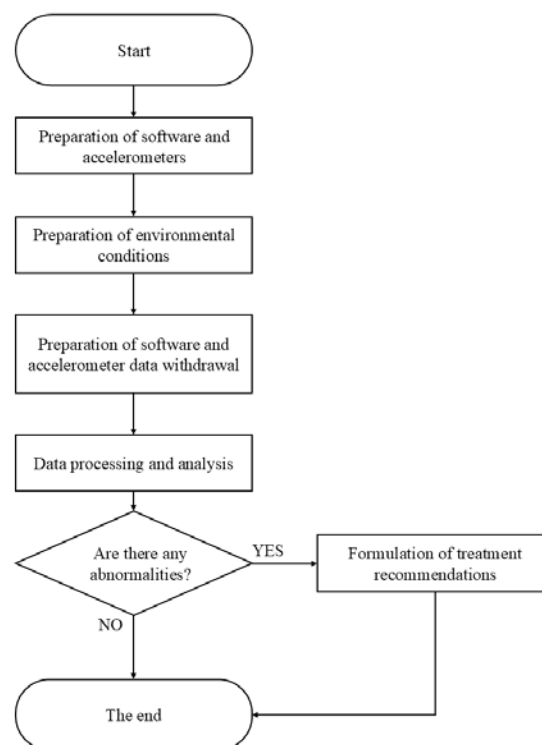


Figure 3 – Flowchart of the algorithm for determining musculoskeletal abnormalities

To determine the presence of kyphosis, it is necessary to find the angle of inclination of the shoulder blades (scapulas) relative to the vertical axis using the formula (1):

$$K_{\text{kyphosys}} = \arctan\left(\frac{Z}{Y}\right). \quad (1)$$

The normal angle of inclination is from 15 to 30 degrees; if the angle of inclination is greater, then different degrees of kyphosis are diagnosed.

The same accelerometer data will be used to conduct experiments to detect scoliosis, namely angle data that can be used to determine the degree of scoliosis or the asymmetry coefficient (2):

$$K_{\text{asymm}} = \frac{L - R}{\frac{L + R}{2}} \times 100\%. \quad (2)$$

For the joint use of the markerless method and accelerometers, it is necessary to synchronize the received data, which can be done using time information. The data set obtained from the accelerometers also contains data on the time of fixation of the indicators of each accelerometer, respectively, it is possible to bind to a specific frame of the shooting from this time data. Now you can calculate the error when using the combined analysis using the markerless method and the accelerometer (3):

$$P_{\text{comb}} = \frac{\omega_a \times P_a + \omega_b \times P_b}{\omega_a + \omega_b}. \quad (3)$$

4 EXPERIMENTS

The work considers good and dim lighting:

– good lightning is from 300 lux or 300 lumens per m²;

– dim lighting is up to 100 lux or 100 lumens per m².

The background is compared for uniformity:

– a solid background is an ideal condition for working with computer vision methods;

– a noisy background or a multi-colored background with high contrast can create interference when identifying key points of the body.

Walking speed – a variable condition that can affect the result in different ways, the following approximate indicators were chosen:

– slow, up to 50 steps per minute;

– fast, from 100 steps per minute.

It is also important to define metrics for evaluating the success and efficiency of the program, which will help to weigh in on the correct decisions for software implementation and successful conditions:

– error in detecting inclination angles relative to detected key points of the body;

– delay of data processing of one frame in milliseconds.

Using the selected indicators, you can not only qualitatively present the results, but also justify improvements to the program or process of identifying deviations from the norm of the human musculoskeletal system.

The research tool is a developed software application created using the Python programming language and the MediaPipe library for markerless movement analysis based on the identification of key points of the body and the WitMotion WT9011DCL-BT50 accelerometer (Fig. 4) together with software from manufacturers for collecting inertial data (Fig. 5).

The study used the MediaPipe Holistic model, which allows tracking key points of the body, hands, and face in real time.

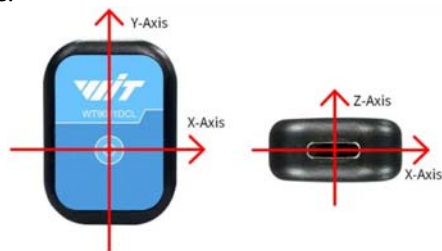


Figure 4 – WitMotion WT9011DCL-BT50 sensors

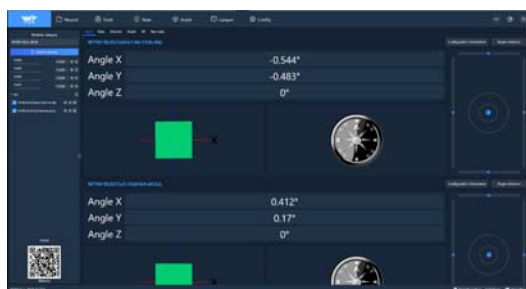


Figure 5 – Software from WitMotion

Accelerometers are attached to the patient's shoulder blades to monitor movements, and then WitMotion software is used for initial data recording in the form of a CSV file. The final stage is the analysis of the obtained data in a developed Python program.

The first stage is the preparation of software and configuration of equipment (camera, accelerometers) for data collection. Before the start of the experiments, the system was tested for the correctness of determining key points (Fig. 6) and the accelerometers were calibrated to obtain correct data (Fig. 5).



Figure 6 – Correctness of key point recognition test using MediaPipe Holistics

External conditions were adjusted to assess the accuracy of the methods in different scenarios:

– two lighting modes are organized: natural daylight and artificial light of varying intensity;

– a solid background is provided using a canvas and a noisy background with furniture, decorative elements and moving objects;

– speed measurement was performed using a metronome and stopwatch for patients' stable pace compliance.

To detect kyphosis in a patient, he needs to move from side to side with a normal gait so that the back is visible from the side (Fig. 7a, Fig. 7b). The results of the developed program for detecting kyphosis are presented in Figure 8a.

The results of this experiment indicate that poor lighting conditions do indeed negatively affect the accuracy of detecting musculoskeletal abnormalities.

Also, the results of studies on the detection of kyphosis based on optical markerless methods showed a large error in poor lighting and noisy background. With increasing patient movement speed, the error decreased but was too large to provide an accurate result. It is also worth noting that poor environmental conditions increase the time for processing one video frame, from which we can conclude that for a satisfactory analysis result and program speed, it is recommended to create conditions that will not interfere with data processing and reduce the overall error when detecting deviations from the norm.

For the scoliosis detection experiments, the input data did not change except for the patient's direction of movement; to detect scoliosis, the patient needs to move towards or away from the camera so that the back is in full view. The results of the developed scoliosis detection program (Fig. 7b) are presented in Figure 8a.

Good lighting and a solid background have the greatest impact on reducing the error in determining the deviation, which can be seen in Figures 8a and 8b.

The results of studies using markerless methods have shown the ease of use and speed of analysis, but poor lighting and a non-uniform background with the presence of objects in the final frame increase the processing time of each frame and error in determining the presence of deviations from the norm in the patient. To reduce the error, it is necessary to improve image processing methods and/or use the markerless method together with others.

In the study of detecting musculoskeletal abnormalities using accelerometers, there are two parameters that affect the accuracy of identifying the degree of joint abnormalities: the placement of the accelerometers on the body and the patient's walking speed. The accelerometers will be placed at the beginning and middle of the thoracic vertebrae to detect kyphosis (Fig.7c), and symmetrically on each scapula to detect scoliosis (Fig. 7d).

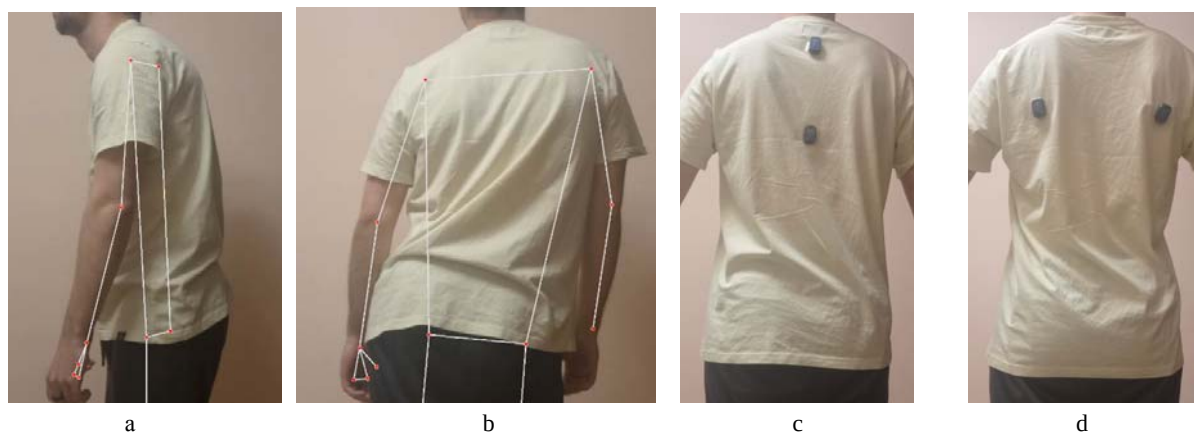


Figure 7 – Joint position and additional sensors to determine kyphosis and scoliosis:

a – the position of the back in which kyphosis is determined based on the optical markerless method, b – the position of the back in which scoliosis is detected based on the optical markerless method, c – placement of accelerometers to determine kyphosis, d – placement of accelerometers to determine scoliosis

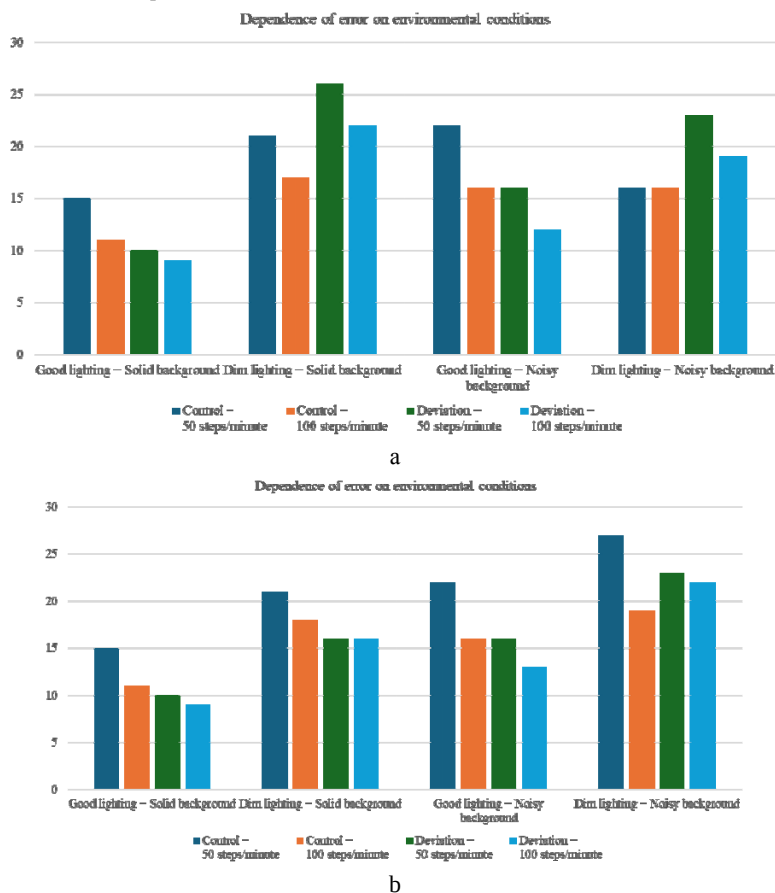


Figure 8 – Dependence graphs of the accuracy of determining the body position deviation on external conditions: a – dependence graph of the error in determining kyphosis on external conditions; b – dependence graph of the error in determining scoliosis on external conditions

The results of processing data from accelerometers for determining kyphosis are given in Table 3.

Table 3 – Results of using accelerometers to detect kyphosis

	Angle of inclination	
	Walking speed up to 50 steps per minute	Walking speed up to 100 steps per minute
Control experiment (normal body position)	20°	30°
Experiment with deflection (with a case with a tilt)	41°	54°

The results of studies on detecting kyphosis using accelerometers show high accuracy rates in identifying the presence or absence of deviations in patients, as well as the dependence of the increase in the angle of inclination on the patient's movement speed, which must be considered when analyzing the data.

The results of processing data from accelerometers for determining scoliosis are given in Table 4.

Table 4 – Results of using accelerometers to detect scoliosis

	Asymmetry coefficient	
	Walking speed up to 50 steps per minute	Walking speed up to 100 steps per minute
Control experiment (normal body position)	1.5%	3%
Deflection experiment (with a tilted body)	8.2%	13.5%

From the presented results of studies on the detection of scoliosis, it is clear that the use of accelerometers determines with high accuracy whether a deviation is present or not. As in the experiments on the detection of kyphosis, the dependence of the resulting data on the speed of the patient's movement is visible, therefore, in future analysis, it is also necessary to take into account the increase in the asymmetry coefficient due to the speed of the patient's movement.

The use of accelerometers has shown that their use provides sufficiently accurate data with little dependence on environmental conditions, but to obtain such an analysis requires a number of conditions, namely: the presence of at least a pair of accelerometers, their precise installation at the specified points of the patient's body, preliminary data collection without the possibility of obtaining the result immediately. Despite the disadvantages of using accelerometers, the high accuracy of the obtained analysis allows it to be used not only as an independent method for detecting deviations of the musculoskeletal system in a patient, but also to increase the accuracy of the results of visual methods of data acquisition and analysis.

5 RESULTS

The results of the proposed combined method of monitoring the position of the patient's joint during gait are presented in this section. As the research results showed (Fig. 8, Table 3, Table 4), each of the methods has its strengths and weaknesses, but the key features are

fast processing in the markerless method and high accuracy in accelerometers, regardless of conditions such as lighting intensity and monotony of the video background. If these two methods are combined, this will ensure increased accuracy of measurement and analysis and resistance to external factors. But when used together, the final analysis can be obtained only after processing data from the video recording and accelerometers, which means the impossibility of obtaining a result in real time.

Returning to the results of the markerless method, we can see that the smallest error that was achieved was 9% (Fig. 8). Since the MediaPipe framework works with color images, in addition to improving the shooting conditions, you can use the methods of contrast enhancement (CLAHE) and noise filtering (Gaussian Blur). To check how much these modifications will improve the accuracy of the analysis, we will conduct an experiment in the best lighting conditions, a monochromatic background, at a patient speed of up to 100 steps per minute, the results of kyphosis detection are given in Table 5.

Table 5 – Results of using a modified markerless method for detecting kyphosis

	Processing time of one frame, ms	Number of correctly identified frames	Error
Control experiment	48	141	6%
Deflection experiment	48	142	6%

The results of experiments on determining scoliosis using the modified program are given in Table 6.

Table 6 – Results of using a modified markerless method for detecting scoliosis

	Processing time of one frame, ms	Number of correctly identified frames	Error
Control experiment	48	142	6%
Deflection experiment	48	140	6%

From the results of the experiments, the error in determining musculoskeletal abnormalities when using marker-free methods has decreased to 6%, which is a significant improvement from the previous best result.

Video recording was conducted at a frequency of 30 frames per second, data collection by accelerometers occurred on average every 140 milliseconds (7 samples per second). Accelerometers cannot work with 100% accuracy, so assuming that their error will be 1%, we can calculate the total error (4):

$$P_{\text{comb}} = \frac{7 \times 1 + 30 \times 6}{7 + 30} \approx 5.05\%. \quad (4)$$

Thus, combining the use of a marker-free method and accelerometers allowed us to reduce the error in detecting deviations of the musculoskeletal system from the norm to 5%, increasing resistance to external factors.

6 DISCUSSION

Thus, the combined approach demonstrates high efficiency for the analysis of movements in rehabilitation orthopedics, which allows for expanding the possibilities for both diagnostics and monitoring of the treatment process of patients. The results of the experiments confirm the feasibility of its further improvement and implementation in practice. However, the markerless method of analysis and the use of accelerometers may have a reasonable separate use due to their advantages (Table 7).

Table 7 – Substantiation of the effectiveness of combined processing of data on the condition of patients' joints

Method	Advantages	Disadvantages	Recommended usage
Markerless method	Does not require special equipment; Analysis speed; Minimal implementation costs.	High dependence on lighting and background.	For initial analysis in good lighting conditions.
Using accelerometers	High measurement accuracy; Independence from lighting and background.	Does not work in real time; Difficult to install and calibrate.	For monitoring movements in difficult conditions (poor lighting, noisy background).
Combined method	High measurement accuracy; High resistance to external factors. Comprehensive analysis of movements.	Does not work in real time; Difficulty in data synchronization.	For precise analysis of complex cases and creation of individual programs.

Re-examining the advantages and disadvantages of the methods for analyzing musculoskeletal abnormalities has allowed us to determine the conditions under which each of them can be used. The combined use of methods is most effective for rehabilitation orthopedics, which opens up opportunities for the development of complex systems for analyzing movements that can provide high diagnostic accuracy, adaptation to individual patient needs, and the possibility of remote monitoring. Further research can be aimed at implementing machine learning algorithms to automate the analysis and adapt the systems to work in real time.

CONCLUSIONS

The scientific novelty of the proposed study lies in the development of a hybrid method for real-time postural analysis using a combination of markerless computer vision (MediaPipe Holistic) and inertial accelerometric data. Within the framework of this work, a combined method for monitoring the position of the patient's joints

during gait has been proposed, which includes error correction based on a weighted combination of optical and sensor-based approaches. For the first time, a formal definition of postural deviation metrics has been introduced, enabling the classification of symmetry states, including mild and pronounced scoliosis, based on an asymmetry coefficient.

The practical significance of the work consists in the creation of an affordable, portable, and customizable tool for remote rehabilitation diagnostics that can be used outside clinical settings. The system supports individualized monitoring of patients with musculoskeletal disorders, improves the accuracy of motion analysis, and enhances the quality of feedback for physiotherapists and healthcare professionals. It ensures accurate detection of postural disorders under varying external conditions, which is critically important for scalable telemedicine applications.

Prospects for further research include the implementation of machine learning algorithms for automatic classification and the generation of adaptive rehabilitation recommendations. Future system improvements will focus on real-time synchronization, integration with electromyographic sensors, and the personalization of recovery strategies based on long-term monitoring. Additionally, the expansion of functionality is planned through artificial intelligence modules for predictive analysis and anomaly detection in gait patterns.

ACKNOWLEDGEMENTS

The work was supported by the state-funded research project of Kharkiv National University of Radio Electronics entitled "Integrated technological solutions for accelerating physical and socio-psychological rehabilitation of prosthetic patients" (state registration number 0125U001654).

REFERENCES

1. Kuroda Yu., Young M., Shoman H. et al. Advanced rehabilitation technology in orthopaedics – a narrative review, *International Orthopaedics*, 2021, Vol. 45, pp. 1933–1940. DOI: 10.1007/s00264-020-04814-4.
2. Santilli V., Mangone M., Diko A. et al. The Use of Machine Learning for Inferencing the Effectiveness of a Rehabilitation Program for Orthopedic and Neurological Patients, *International Journal of Environmental Research and Public Health*, 2023, Vol. 20, Article ID 5575. DOI: 10.3390/ijerph20085575.
3. Barkovska O., Oliinyk D., Sorokin A. et al. A system for monitoring the progress of rehabilitation of patients with musculoskeletal disorder, *Advanced Information Systems*, 2024, Vol. 8, Issue 3, pp. 13–24. DOI: 10.20998/2522-9052.2024.3.02.
4. Phuphanich M. E., Sinha K. R., Truong M. et al. Telemedicine for musculoskeletal rehabilitation and orthopedic post-operative rehabilitation, *Physical Medicine and Rehabilitation Clinics of North America*, 2021, Vol. 32, Issue 2, pp. 319–353. DOI: 10.1016/j.pmr.2020.12.004.
5. Kiro L., Urbanovych A., Zak M. Intervention impact on quality of life in Ukrainians with post-traumatic stress disorder.

- der, *BMC Psychology*, 2024, Vol. 12, Issue 1, pp. 601. DOI: 10.1186/s40359-024-02109-6.
6. Ko S., Pareek A., Ro D. H. et al. Artificial intelligence in orthopedics: three strategies for deep learning with orthopedic specific imaging, *Knee Surgery, Sports Traumatology, Arthroscopy*, 2022, Vol. 30, Issue 3, pp. 758–761. DOI: 10.1007/s00167-021-06838-8.
7. Prijs J., Liao Z., Ashkani-Esfahani S. et al. Artificial intelligence and computer vision in orthopaedic trauma: the why, what, and how, *The Bone & Joint Journal*, 2022, Vol. 104-B, Issue 8, pp. 911–914. DOI: 10.1302/0301-620X.104B8.BJJ-2022-0119.R1.
8. Fan X., Zhu Q., Tu P. et al. A review of advances in image-guided orthopedic surgery, *Physics in Medicine & Biology*, 2023, Vol. 68, Issue 2. DOI: 10.1088/1361-6560/aca9e9.
9. Goto R., Pinchuk I., Kolodezhny O. et al. Mental health of adolescents exposed to the war in Ukraine, *JAMA Pediatrics*, 2024, Vol. 178, Issue 5, pp. 480–488. DOI: 10.1001/jamapediatrics.2024.0295.
10. Dzhus M., Golovach I. Impact of Ukrainian-Russian war on health care and humanitarian crisis, *Disaster Medicine and Public Health Preparedness*, 2023, Vol. 17, pp. e340. DOI: 10.1017/dmp.2022.265.
11. Prill R., Królikowska A., de Girolamo L. et al. Checklists, risk of bias tools, and reporting guidelines for research in orthopedics, sports medicine, and rehabilitation, *Knee Surgery, Sports Traumatology, Arthroscopy*, 2023, Vol. 31, Issue 8, pp. 3029–3033. DOI: 10.1007/s00167-023-07442-8.
12. Bradshaw T. J., Huemann Z., Hu J. et al. A guide to cross-validation for artificial intelligence in medical imaging, *Radiology: Artificial Intelligence*, 2023, Vol. 5, Issue 4, pp. e220232. DOI: 10.1148/ryai.220232.
13. Esteva A., Chou K., Yeung S. et al. Deep learning-enabled medical computer vision, *NPJ Digital Medicine*, 2021, Vol. 4, P. 5. DOI: 10.1038/s41746-020-00376-2.
14. Braun B. J., Grimm B., Hanflik A. M. et al. Wearable technology in orthopedic trauma surgery – an AO trauma survey and review of current and future applications, *Injury*, 2022, Vol. 53, Issue 6, pp. 1961–1965. DOI: 10.1016/j.injury.2022.03.026.
15. Elyan E., Vuttipittayamongkol P., Johnston P. et al. Computer vision and machine learning for medical image analysis: recent advances, challenges, and way forward, *Artificial Intelligence Surgery*, 2022, Vol. 2, pp. 24–45. DOI: 10.20517/ais.2021.15.
16. Lee H. Y., Park J. H., Kim T. W. Comparisons between Locomat and Walkbot robotic gait training regarding balance and lower extremity function among non-ambulatory chronic acquired brain injury survivors, *Medicine (Baltimore)*, 2021, Vol. 100, Issue 18, pp. e25125. DOI: 10.1097/MD.00000000000025125.

Received 07.04.2025.
Accepted 17.09.2025.

УДК 004.932:612.76

ФОТОГРАМЕТРИЧНА ПІДСИСТЕМА ЗАХОПЛЕННЯ РУХУ ДЛЯ АНАЛІЗУ ЗМІН ПОЛОЖЕННЯ ТІЛА ВІДНОСНО ФРОНТАЛЬНОЇ ТА САГІТАЛЬНОЇ ПЛОЩИН

Барковська О. Ю. – канд. техн. наук, доцент, кафедра електронних обчислювальних машин, Харківський національний університет радіоелектроніки, Харків, Україна.

Коваленко А. А. – д-р техн. наук, професор, завідувач кафедри електронних обчислювальних машин, Харківський національний університет радіоелектроніки, Харків, Україна.

Дяченко В. О. – ст. викл. кафедри електронних обчислювальних машин, Харківський національний університет радіоелектроніки, Харків, Україна.

Бухарова Л. Д. – магістрант кафедри електронних обчислювальних машин, Харківський національний університет радіоелектроніки, Харків, Україна.

Коробко В. Ю. – магістрант кафедри електронних обчислювальних машин, Харківський національний університет радіоелектроніки, Харків, Україна.

АНОТАЦІЯ

Актуальність. Зростання кількості ортопедичних травм, зокрема серед військовослужбовців, потребує нових інноваційних рішень для оцінки змін постави та моніторингу ефективності реабілітації. Існуючі системи мають обмеження щодо портативності, вартості та гнучкості у використанні, що зумовлює потребу в розробці гібридних систем, які поєднують методи комп'ютерного зору та сенсорного аналізу.

Мета роботи – оцінити ефективність поєднання безконтактного комп'ютерного зору та акселерометричних сенсорів для виявлення змін постави людини в умовах різного освітлення, фону та швидкості руху.

Метод. У дослідженні реалізовано фотограмметричну підсистему, що включає MediaPipe Holistic для безмаркерного відстеження ключових точок тіла та акселерометри WitMotion WT9011DCL-BT50 для аналізу інерційних параметрів руху. Модель системи побудовано у нотатції IDEF0. Оцінку точності виконано шляхом порівняння отриманих значень кута нахилу лопаток та коефіцієнта асиметрії із заданими нормами.

Результати. Комбіноване використання візуальних і сенсорних даних дало змогу зменшити похибку до 5,05% при нормальних умовах та забезпечити стабільність результатів за умов змін зовнішнього середовища. Модифікація зображень (контраст, фільтрація шуму) підвищила точність комп'ютерного зору. Визначено порогові значення коефіцієнта асиметрії, що відповідають нормі, легким та важким порушенням постави.

Висновки. Запропонована система демонструє високу потенційну ефективність у телемедичному супроводі реабілітації пацієнтів із порушеннями опорно-рухового апарату. Її практична значущість полягає у створенні доступного, портативного та точного засобу діагностики та контролю, придатного до подальшої інтеграції у системи персоналізованої медицини з вбудованими модулями штучного інтелекту.

КЛЮЧОВІ СЛОВА: реабілітаційна ортопедія, захоплення руху, комп'ютерний зір, MediaPipe Holistic, інерційні сенсори, аналіз постави, моніторинг ходи, телемедицина, акселерометр.

ЛІТЕРАТУРА

1. Advanced rehabilitation technology in orthopaedics – a narrative review / [Yu. Kuroda, M. Young, H. Shoman et al.] // *International Orthopaedics*. – 2021. – Vol. 45. – P. 1933–1940. DOI: 10.1007/s00264-020-04814-4.
2. The Use of Machine Learning for Inferencing the Effectiveness of a Rehabilitation Program for Orthopedic and Neurological Patients / [V. Santilli, M. Mangone, A. Diko et al.] // *International Journal of Environmental Research and Public Health*. – 2023. – Vol. 20. – Article ID 5575. DOI: 10.3390/ijerph20085575.
3. A system for monitoring the progress of rehabilitation of patients with musculoskeletal disorder / [O. Barkovska, D. Oliynyk, A. Sorokin, et al.] // *Advanced Information Systems*. – 2024. – Vol. 8, № 3. – P. 13–24. DOI: 10.20998/2522-9052.2024.3.02.
4. Telemedicine for musculoskeletal rehabilitation and orthopedic postoperative rehabilitation / [M. E. Phuphanich, K. R. Sinha, M. Truong et al.] // *Physical Medicine and Rehabilitation Clinics of North America*. – 2021. – Vol. 32, № 2. – P. 319–353. DOI: 10.1016/j.pmr.2020.12.004.
5. Kiro L. Intervention impact on quality of life in Ukrainians with post-traumatic stress disorder / L. Kiro, A. Urbanovych, M. Zak // *BMC Psychology*. – 2024. – Vol. 12, Issue 1. – P. 601. DOI: 10.1186/s40359-024-02109-6.
6. Artificial intelligence in orthopedics: three strategies for deep learning with orthopedic specific imaging / [S. Ko, A. Pareek, D. H. Ro et al.] // *Knee Surgery, Sports Traumatology, Arthroscopy*. – 2022. – Vol. 30, № 3. – P. 758–761. DOI: 10.1007/s00167-021-06838-8.
7. Artificial intelligence and computer vision in orthopaedic trauma: the why, what, and how / [J. Prijs, Z. Liao, S. Ashkani-Esfahani et al.] // *The Bone & Joint Journal*. – 2022. – Vol. 104-B, № 8. – P. 911–914. DOI: 10.1302/0301-620X.104B8.BJJ-2022-0119.R1.
8. A review of advances in image-guided orthopedic surgery / [X. Fan, Q. Zhu, P. Tu et al.] // *Physics in Medicine & Biology*. – 2023. – Vol. 68, № 2. DOI: 10.1088/1361-6560/acaae9.
9. Mental health of adolescents exposed to the war in Ukraine / [R. Goto, I. Pinchuk, O. Kolodezhny et al.] // *JAMA Pediatrics*. – 2024. – Vol. 178, № 5. – P. 480–488. DOI: 10.1001/jamapediatrics.2024.0295.
10. Dzhus M. Impact of Ukrainian-Russian war on health care and humanitarian crisis / M. Dzhus, I. Golovach // *Disaster Medicine and Public Health Preparedness*. – 2023. – Vol. 17. – P. e340. DOI: 10.1017/dmp.2022.265.
11. Checklists, risk of bias tools, and reporting guidelines for research in orthopedics, sports medicine, and rehabilitation / [R. Prill, A. Królikowska, L. de Girolamo et al.] // *Knee Surgery, Sports Traumatology, Arthroscopy*. – 2023. – Vol. 31, № 8. – P. 3029–3033. DOI: 10.1007/s00167-023-07442-8.
12. A guide to cross-validation for artificial intelligence in medical imaging / [T. J. Bradshaw, Z. Huemann, J. Hu et al.] // *Radiology: Artificial Intelligence*. – 2023. – Vol. 5, № 4. – P. e220232. DOI: 10.1148/ryai.220232.
13. Deep learning-enabled medical computer vision / [A. Esteva, K. Chou, S. Yeung et al.] // *NPJ Digital Medicine*. – 2021. – Vol. 4. – P. 5. DOI: 10.1038/s41746-020-00376-2.
14. Wearable technology in orthopedic trauma surgery – an AO trauma survey and review of current and future applications / [B. J. Braun, B. Grimm, A. M. Hanflik et al.] // *Injury*. – 2022. – Vol. 53, № 6. – P. 1961–1965. DOI: 10.1016/j.injury.2022.03.026.
15. Computer vision and machine learning for medical image analysis: recent advances, challenges, and way forward / [E. Elyan, P. Vuttipittayamongkol, P. Johnston et al.] // *Artificial Intelligence Surgery*. – 2022. – Vol. 2. – P. 24–45. DOI: 10.20517/ais.2021.15.
16. Lee H. Y. Comparisons between Locomat and Walkbot robotic gait training regarding balance and lower extremity function among non-ambulatory chronic acquired brain injury survivors / H. Y. Lee, J. H. Park, T. W. Kim // *Medicine (Baltimore)*. – 2021. – Vol. 100, Issue 18. – P. e25125. DOI: 10.1097/MD.00000000000025125.

USE OF GENETIC ALGORITHMS IN ADAPTIVE COMPILERS FOR CROSS-PLATFORM OPTIMIZATION

Berdnyk M. G. – Dr. Sc., Associate Professor, Professor of the Department of Computer Systems Software, National Technical University “Dnipro Polytechnic”, Dnipro, Ukraine.

Starodubskiy I. P. – Post-graduate student, Department of Computer Systems Software, National Technical University “Dnipro Polytechnic”, Dnipro, Ukraine.

ABSTRACT

Context. Modern software is developed under conditions of continuously increasing complexity of computing systems. Today, developers must consider a vast diversity of platforms, ranging from resource-constrained mobile devices to servers with high-performance processors and specialized architectures such as GPUs, FPGAs and even quantum computers. This heterogeneity requires software to operate efficiently across various hardware platforms. However, portability remains one of the most challenging tasks, especially when high performance is required.

One of the promising directions to address this problem is the use of adaptive compilers that can automatically optimize code for different architectures. This approach allows developers to focus on the functional part of the software, minimizing the effort spent on optimization and configuration. Genetic algorithms (GAs) play a special role among the methods used to create adaptive compilers. These are powerful evolutionary techniques that allow finding optimal solutions in complex and multidimensional parameter spaces.

Objective. The objective of this research is to apply genetic algorithms in the process of adaptive compilation to enable automatic optimization of software across different hardware platforms.

Method. The approach is based on genetic algorithms to automate the compilation process. The key stages include: population initialization – creation of an initial set of compilation parameters; fitness function evaluation – assessment of the efficiency of each parameter combination; evolutionary operations – applying crossover, mutation and selection to generate the next generation of parameters; termination criteria – stopping the iterative process upon reaching an optimal result or stabilization of metrics.

Results. The developed algorithm was implemented in Python using the numpy, multiprocessing and subprocess libraries. Performance evaluation of the algorithm was carried out using execution time, energy consumption and memory usage metrics.

Conclusions. The scientific novelty of the study lies in the development of: an innovative approach to automatic compilation parameter optimization based on genetic algorithms; a method for dynamic selection of optimization strategies based on performance metrics for different architectures; integration of GAs with modern compilers such as LLVM for automatic analysis of intermediate representation and code optimization; and methods for applying adaptive compilers to solve cross-platform optimization problems. The practical significance is determined by the use of genetic algorithms in adaptive compilers, which significantly improves the efficiency of the compilation process by automating the selection of optimal parameters for various architectures. The proposed approach can be successfully applied in the fields of mobile computing, cloud technologies and high-performance systems.

KEYWORDS: genetic algorithms, adaptive compilers, cross-platform optimization, automation.

ABBREVIATIONS

GA – Genetic Algorithm;

IR – Intermediate Representation (a form of program representation used during compilation for analysis and optimization);

LLVM – Low Level Virtual Machine (a toolkit for compiler development that provides IR for program analysis and optimization, as well as machine code generation for various architectures);

Clang – a high-performance compiler for C, C++ and Objective-C that uses LLVM as a backend for code optimization and generation;

LLVM Pass Manager – a mechanism in LLVM for managing code optimizations and transformations, allowing step-by-step analysis and modification of IR;

SPEC CPU2017 – a standardized benchmark suite for evaluating processor and compiler performance, including real-world workloads from domains such as computational math, artificial intelligence and physics;

429.mcf – a benchmark from the SPEC CPU2017 suite that simulates a network flow problem, used to test system performance under heavy computational loads;

456.hmmr – a benchmark from the SPEC CPU2017 suite used to evaluate performance in analyzing biological sequences using Hidden Markov Model (HMM) algorithms.

NOMENCLATURE

C is a set of compilation parameters;

C_i is a i -th chromosome (a vector of compilation parameters) in the current population;

C_{best} is a the best solution that has the maximum value of the fitness function;

C_{new} is a new chromosome obtained after crossover or mutation operations;

E is a set of graph edges representing dependencies between instructions in the form of data or control flow relationships;

$F^{(t)}$ is a vector of fitness function values for the population at generation t ;

G_{IR} is a directed graph of intermediate representation (Intermediate Representation, IR) used for analyzing code structure during compilation;

$inst_i$ is an i -th instruction in the intermediate representation;

N is a normal distribution with zero mean and variance σ^2 ;

N_{elite} is a number of elite chromosomes preserved for the next generation;

P is a population of chromosomes used in the genetic algorithm; each chromosome represents a possible configuration of compilation parameters;

P_{next} is a next generation population;

$P^{(0)}$ is an initial (zero-th) population;

$P^{(t)}$ is a population at iteration t ;

$P_{crossed}^{(t)}$ is a set of chromosomes obtained through crossover at generation t ;

$P_{elite}^{(t)}$ is an elite subset of the population at generation t ;

$P_{mutated}$ is a subset of the population obtained after mutation;

$P_{mutated}^{(t)}$ is a set of chromosomes obtained through mutation at generation t ;

p_i is a value of a compilation parameter that reflects whether a specific compilation optimization is applied;

$p_{new,i}$ is a new i -th chromosome after applying genetic operators;

V is a set of graph vertices, each representing a separate instruction of the intermediate representation;

w_i is a i -th weight coefficients that determine the priority of the corresponding metrics;

δ is a delta value sampled from a normal distribution with zero mean and variance σ^2 ;

σ is a standard deviation used in the mutation process to sample random variations.

INTRODUCTION

Modern adaptive compilers aim to solve the problem of automatic software optimization for various computing platforms. This is achieved through algorithms capable of analyzing the characteristics of target platforms, adapting compilation parameters and ensuring maximum performance. One of the most promising approaches is the use of genetic algorithms (GA), which provide powerful tools for searching optimal solutions in complex multidimensional parameter spaces.

Genetic algorithms are based on the principles of biological evolution. Their core idea lies in using processes similar to natural selection, mutation and crossover to improve solutions to complex problems. In the context of compilation, each possible solution, represented as a “chromosome”, contains a set of parameters that define the optimization strategy. These parameters include elements such as loop unrolling, instruction selection, register allocation and memory management.

The population of solutions goes through iterations called “generations”. At each step, the algorithm evaluates the effectiveness of the solutions based on a fitness function calculated using given performance metrics. The most successful solutions are preserved and used to create new, improved generations. This process continues until an optimal solution is found or until predefined termination criteria, such as a set number of iterations, are met.

To apply genetic algorithms in adaptive compilers, it is necessary to integrate them with existing tools for code analysis and generation.

The object of this research is the process of integrating genetic algorithms into adaptive compilation for various hardware platforms.

The subject of this research is the models and methods for developing and selecting genetic algorithms to build adaptive compilers.

The purpose of this research is to develop a model for using genetic algorithms in the process of adaptive compilation to ensure automatic optimization of software for different platforms.

1 PROBLEM STATEMENT

To apply genetic algorithms in adaptive compilers, it is necessary to integrate them with the existing LLVM code analysis and generation tool, which provides powerful capabilities for working with intermediate representation (IR). Genetic algorithms should utilize metrics collected during the IR analysis stage, such as the number of instructions, loop depth, register usage, memory access frequency, vectorization level and the size of the IR to evaluate the efficiency of each solution.

For the formalization of the problem, we assume that a set of compilation parameters $P = \{p_1, p_2, \dots, p_k\}$ is given, where each parameter p_i takes a value of 0 or 1, corresponding to its disabled or enabled state. The compilation parameters may include: -O0, -O1, -O2, -O3, -Ofast, -Og, -funroll-loops, -fno-unroll-loops, -ftree-vectorize, -fno-tree-vectorize, -fomit-frame-pointer, -fno-omit-frame-pointer, -march=native, -march=armv8-a, -mtune=cortex-a73, -mtune=generic, -flto, -fno-lto, and others.

Additionally, a set of possible compilation configurations C is given, where each configuration $C \subseteq P$ is a feature vector describing the selected optimization parameters.

The performance metrics for compilation with parameters C include the program execution time $T(C)$, energy consumption $E(C)$ and memory usage $M(C)$.

It is required to find such a configuration of compilation parameters C^* that minimizes the objective function:

$$f(C) = w_1 \cdot T(C) + w_2 \cdot E(C) + w_3 \cdot M(C). \quad (1)$$

The integration process should include the following stages:

1. Initial population generation. Initial compilation parameters are formed based on standard strategies or generated randomly. These may include loop unrolling, optimization flags and other parameters.

2. Fitness evaluation. Each combination of compilation parameters is tested on the target platform, and performance data is collected. This includes program execution time, energy consumption or memory usage.

3. Application of GA operators. The mutation operator randomly alters compilation parameters, adding diversity to the population. The crossover operator combines parameters from two successful solutions to produce a potentially more effective one.

4. Selection of the best solutions. Solutions with the highest fitness scores are retained for the next generation.

It is necessary to design application scenarios such as:

1. Optimization for mobile devices. Genetic algorithms can select parameters that minimize energy consumption, such as reducing input-output operations or minimizing memory access [1].

2. Server applications. For servers, GAs may focus on optimizing instructions for parallel execution, which is critical for high-performance computing.

Heterogeneous systems. In systems that use both CPUs and GPUs, GAs can find a balance between tasks executed on each architecture to minimize latency and improve overall performance.

2 REVIEW OF THE LITERATURE

Genetic algorithms (GAs) have long been used to search for optimal compiler parameters due to their ability to efficiently explore large solution spaces [2]. Modern compilers contain dozens or even hundreds of options and optimization phases, and a fixed optimization sequence (e.g., standard levels -O1 to -O3) is not universally optimal for all programs and platforms. One basic solution to this issue is iterative compilation, where the code is compiled repeatedly with different combinations of options, and the results are evaluated to select the best configuration. In this context, GAs have proven to be an effective global search method: evolutionary operators such as selection, crossover and mutation gradually improve generations of solutions and discover combinations of flags or compiler pass sequences that yield better program performance.

For example, a 2022 study proposed a parallel GA for selecting optimization sequences in LLVM [3]. The authors clustered programs into subsets and ran three GAs in parallel (with solution sharing between them), allowing exploration of different regions of the solution space. As a result, this approach outperformed the standard -O2 level by an average of 87% in execution speed and exceeded the performance of a sequential (single-threaded) GA by 72–75% [3].

Another 2021 study focused on autotuning GCC compiler flags using a GA. The proposed framework FOGA (Flag Optimization with Genetic Algorithm)

implemented dynamic tuning of the GA's own hyperparameters and a stopping criterion when no improvements were observed, resulting in a significant speedup of C++ program execution compared to the then-leading OpenTuner toolkit [4].

These studies confirm that evolutionary algorithms effectively solve both the problem of selecting compiler parameters and the phase-ordering problem in cases where exhaustive search is practically infeasible.

Adaptive compilation approaches can generally be divided into two major categories: search-based methods (such as GAs and other metaheuristics) and learning-based methods (machine learning). Iterative search methods (including GAs, evolutionary strategies, simulated annealing and others) perform black-box optimization by directly measuring the performance of the generated code [5]. Their advantage lies in discovering high-quality solutions without requiring an a priori model, but their main drawback is high computational cost – a compiler may need to run hundreds or thousands of times for a single program. This makes such approaches difficult to apply in real-world conditions, especially when just-in-time (JIT) optimization is needed.

An alternative is to train a model capable of predicting optimization decisions. In this case, the cost is shifted to the training stage (offline), after which the model quickly produces optimization sets for new programs. In recent years [6], there has been growing interest in integrating machine learning into compilers instead of relying on fixed heuristics. For example, in the MLGO (Machine Learning Guided Optimization) system, Google researchers replaced the heuristic function inlining algorithm in LLVM with a model trained using reinforcement learning [2].

This approach reduced the size of the executable code by up to 7% compared to the standard -Oz mode in LLVM, and the same model worked successfully across various program sets and hardware platforms [2].

The combination of machine learning (ML) and evolutionary algorithms opens new possibilities for code optimization. Evolutionary algorithms, including GAs, are good at exploring extreme solutions but can be complemented by ML models to accelerate convergence [6].

One of the emerging directions is predictive compilation optimization, where an ML model is trained on a dataset of programs and corresponding optimization settings to predict which optimizations will be most effective for a new program. By applying a k-nearest neighbors (KNN) algorithm along with a filtering stage to remove non-beneficial transformations, the system automatically selects an individualized sequence of passes for each program instead of using the standard -O2/-O3 levels [7].

Experimental results show that this method outperforms LLVM's fixed -O2 level, achieving an average of ~21% speedup, and with filtering included, the improvement reaches up to 23% [7].

Despite success in specific scenarios, deploying these approaches in production compilers remains challenging. First, the computational cost of iterative methods is still high. Second, machine learning models risk losing effectiveness on programs significantly different from the training set – the generalization problem remains an active area of research.

Another challenge is integrating these methods into existing compilation pipelines. Compatibility with existing optimization passes and programming languages must be ensured, along with the predictability of results.

Future prospects for adaptive compilers are linked to the continued growth of hardware capabilities and advances in artificial intelligence algorithms. It is expected that compilers will be able to dynamically adapt to the execution environment – for example, performing on-the-fly recompilation based on actual program profiling data.

Recent studies clearly demonstrate the advantages of adaptive compilation, particularly involving genetic algorithms and machine learning methods, for cross-platform optimization. The main challenges today include improving the efficiency and speed of such approaches, ensuring their reliability under various conditions and integrating them into real-world development tools.

3 MATERIALS AND METHODS

Adaptive compilers that utilize genetic algorithms (GAs) are complex modular software systems that provide automatic code optimization for various computing platforms. The architecture of such a compiler includes several interrelated components that sequentially perform analysis, optimization and code generation, enabling dynamic adjustment of compilation parameters based on target platform data.

The architecture of a GA-based compiler consists of the following core modules [8]:

1. Input code analysis module: performs syntactic and semantic analysis, transforming the source code into intermediate representation (IR);

2. Performance metrics collection module: generates numerical data such as instruction count, loops, register usage and memory footprint to evaluate the efficiency of different compilation strategies;

3. Genetic optimizer: responsible for evolutionary improvement of compilation parameters, including mutation, crossover and selection operators;

4. LLVM Pass Manager control module: applies the generated optimization parameters to the IR to enhance performance;

5. Machine code generator: creates executable code adapted to the target platform.

At the first stage, the source code is compiled into IR using LLVM tools such as Clang. IR is a low-level abstraction of the program that includes detailed information about instructions and control flow. Formally, IR can be represented as a dependency graph

where vertices V correspond to instructions and edges E represent dependencies between them [3]:

$$G_{IR} = (V, E), V = \{inst_1, inst_2, \dots, inst_n\}, E \subseteq V \times V.$$

At this stage, basic performance indicators are also collected. Total number of instructions, number of loops in the program, register usage.

Each chromosome represents a vector of compilation parameters:

$$C = \{p_1, p_2, \dots, p_k\}, p_i \in \{0, 1\},$$

where $p_i = 1$ indicates the activation of a specific optimization pass (such as loop unrolling, function inlining, or instruction combination), and $p_i = 0$ indicates its deactivation.

To evaluate the effectiveness of each chromosome, a fitness function $F(C)$ is used, which is calculated based on performance metrics:

$$F(C) = w_1 \cdot T(C)^{-1} + w_2 \cdot E(C)^{-1} + w_3 \cdot M(C)^{-1}.$$

During the training phase, $F(C)$ is optimized using the collected performance data. The mutation operator is defined as follows, where a random value δ is sampled from a normal distribution with zero mean and variance σ^2 :

$$p'_i = p_i + \delta, \delta \sim N(0, \sigma^2).$$

Let us define the crossover for two chromosomes C_1 and C_2 as follows:

$$C_{new} = \alpha \cdot C_1 + (1 - \alpha) \cdot C_2, \alpha \in \{0, 1\}.$$

For selection, an elitist approach is used, in which the best chromosomes of the current generation are preserved for the next one:

$$P_{next} = Elite(P) \cup Crossover(P) \cup Mutation(P),$$

where, $Elite(P)$ is the set of the best (elite) individuals from the current population P , which have the highest fitness and are passed unchanged to the next generation, preserving the quality of already found solutions and preventing the loss of top-performing chromosomes; $Crossover(P)$ is the set of new individuals obtained as a result of crossover between pairs of parent chromosomes from the population P , allowing the combination of successful traits from different solutions and the exploration of new areas in the parameter space; $Mutation(P)$ is the set of individuals created through mutation of existing chromosomes from the population

P , introducing random variations that provide genetic diversity and help avoid local maxima.

At the first stage, the initial population $P^{(0)}$ is generated, consisting of N chromosomes. Each chromosome is represented as a parameter vector C , which encodes a possible solution to the optimization problem.

The population can be initialized randomly:

$$P^{(0)} = \{C_1, C_2, \dots, C_N\}, C_j = \{p_{j1}, p_{j2}, \dots, p_{jk}\}, \\ p_{ji} \in \{0,1\}.$$

Each chromosome is evaluated using the fitness function $F(C)$, which reflects the quality of the solution. The task is to minimize a certain objective function $F(C)$, and the fitness function can be defined as [3]:

$$F(C) = \frac{1}{1 + f(C)}.$$

$f(C)$ is calculated using a standard benchmark suite such as SPEC CPU2017 [9].

For maximization problems, the fitness function can be written as:

$$F(C) = f(C).$$

The objective function for compilation optimization is as follows: (1).

For all chromosomes in the current population $P^{(t)}$, a vector of fitness function values is computed:

$$F^{(t)} = \{F(C_1), F(C_2), \dots, F(C_N)\}.$$

Based on the fitness function values, chromosomes are selected to participate in the next generation. One of the most popular methods is tournament selection. K random chromosomes are chosen from the current population, and the winner of the tournament is the chromosome with the highest fitness value.:

$$C_{best} = \arg \max_{C \in K} F(C).$$

Thus, an intermediate population $P_{selected}^{(t)}$ is formed. For each pair of chromosomes C_1 and C_2 from $P_{selected}^{(t)}$, a crossover operation is performed. Single-point crossover is defined as follows:

$$C_{new} = \{p_{11}, p_{12}, \dots, p_{1m}, p_{2(m+1)}, \dots, p_{2k}\},$$

where the crossover point m is chosen randomly, $m \in [1, k - 1]$. This can also be written as:

$$p_{new,i} = \begin{cases} p_{1,i}, & i \leq m, \\ p_{2,i}, & i > m. \end{cases}$$

The mutation operation modifies the values of individual genes in a chromosome with a certain probability $P_{mutated}$.

The new generation $P^{(t+1)}$ is formed by adding elite chromosomes (elitism):

$$P_{elite}^{(t+1)} = \arg \max_{C \in P^{(t)}} F(C), |P_{elite}^{(t+1)}| = N_{elite}.$$

Crossover and mutation for the remaining chromosomes are performed as follows:

$$P^{(t+1)} = P_{elite}^{(t+1)} \cup P_{crossed}^{(t+1)} \cup P_{mutated}^{t+1}.$$

The iterative process terminates when one of the following criteria is met:

1. The maximum number of generations has been reached.
2. The fitness function value does not change over a stable period of time.
3. The target fitness value has been achieved.

The program represents an implementation of a system for automatic optimization of compilation parameters using a genetic algorithm, developed in the Python programming language. The system consists of four interconnected modules: `genetic_algorithm.py`, `compiler_wrapper.py`, `fitness_evaluator.py` and `main.py`, each performing a distinct functional role within the overall architecture. Python 3 was chosen as the primary development language due to its expressiveness and extensive standard library.

The following libraries are used across the modules: `numpy`, which is employed for generating the initial population, performing array operations, and implementing tournament selection; `random`, used for the stochastic components of the algorithm, including mutation and crossover; `logging`, which serves to systematically log the execution process and diagnostic messages; `concurrent.futures.ThreadPoolExecutor`, which enables multithreaded evaluation of population fitness in order to improve performance; `subprocess`, which provides an interface for invoking the external gcc compiler and executing compiled binaries; and `time`, which is used to measure the execution time of program components and the target executables.

The following compiler input parameters were used to implement the compilation optimization:

- O0, -O1, -O2, -O3, -Ofast, -Og – compiler optimization levels;
- funroll-loops – loop unrolling to reduce the number of iterations and improve performance;

- fno-unroll-loops – disables loop unrolling;
- ftree-vectorize – automatic loop vectorization for using processor SIMD instructions;
- fno-tree-vectorize – disables automatic vectorization;
- fomit-frame-pointer – removes the use of the frame pointer to improve performance, but may complicate debugging;
- fno-omit-frame-pointer – retains the frame pointer, which is useful for debugging;
- march=native – applies optimizations for the current processor architecture;
- march=armv8-a – compiles for the ARMv8-A architecture;
- mtune=cortex-a73 – code optimization for Cortex-A73 processors;
- mtune=generic – general optimizations without targeting a specific architecture;
- flto – Link-Time Optimization (applies optimizations during the linking stage);
- fno-lto – disables link-time optimizations;
- funsafe-math-optimizations – allows potentially unsafe math optimizations to improve performance (e.g., changing the order of floating-point operations);
- fstrict-aliasing – enables aggressive memory access optimizations based on assumptions about pointer aliasing (can improve performance but may cause unpredictable behavior);
- fno-strict-aliasing – disables these optimizations;
- fipa-pta – improved pointer analysis for whole-program optimization;
- fno-ipa-pta – disables this analysis;
- fexpensive-optimizations – enables complex and resource-intensive optimizations;
- fno-expensive-optimizations – disables such optimizations.

This expansion of the parameter space allows exploring a wide range of combinations, including architecture-specific optimizations (-march, -mtune), as well as parameters that affect memory management, loops and computations. Thus, the genetic algorithm minimizes the objective function (1).

The population initialization is performed as follows: a population of 100 chromosomes is created. Each chromosome encodes a combination of compilation parameters as a binary vector {0,1}, where 1 indicates that a parameter is enabled and 0 means it is disabled.

Selection is performed using the tournament method. Crossover is defined as two-point, with a probability of $P_{cross}=0.8$. The probability of gene mutation is $P_{mut}=0.05$. Termination criteria are defined as reaching 100 generations and no improvements for 10 consecutive generations.

The use of an extended parameter space makes it possible to account for the unique characteristics of the platform and application scenario, making the compilation process adaptive and highly efficient.

4 EXPERIMENTS

The experiment consists in validating the proposed algorithm for improving the efficiency of the compilation process by automating the selection of optimal parameters for different architectures.

To assess the effectiveness of the algorithm, the following metrics were used: execution time, energy consumption, and memory usage. Specialized hardware counters (Performance Monitoring Unit, PMU) were used to collect these metrics.

Testing was conducted on three distinct hardware platforms covering a wide range of usage scenarios:

- Mobile processor ARM Cortex-A73. Specifications: quad-core processor with a clock frequency of 2.2 GHz, manufactured using a 10 nm process technology. Environment: Android 11 with built-in energy sensors for measuring power consumption. The SPEC CPU2017 benchmarks (429.mcf, 456.hammer) and mobile workload simulations were used for testing.

- Server-grade processor Intel Xeon Platinum. Specifications: 24-core processor with a clock frequency of 2.3 GHz, supporting 48 threads and a TDP of 165 W. Environment: Linux (kernel 5.15), GCC version 11.2 configured for multi-core optimization.

- Graphics processor NVIDIA RTX 3080 GPU. Specifications: Ampere architecture, 8,704 CUDA cores, 10 GB of GDDR6X memory. Environment: CUDA 11.5 with Python API used for running large-scale data processing tests.

The optimal compilation parameter configurations for the C program were identified for each of the three tested hardware platforms as follows:

- ARM Cortex-A73: -funroll-loops -ftree-vectorize -march=armv8-a -mtune=cortex-a73 -flto -fstrict-aliasing.

- Intel Xeon Platinum: -O3 -funroll-loops -ftree-vectorize -march=native -mtune=generic -flto -fstrict-aliasing -fipa-pta -fexpensive-optimizations.

- NVIDIA RTX 3080 GPU: -Ofast -funroll-loops -ftree-vectorize -flto -fomit-frame-pointer -funsafe-math-optimizations.

As a result of the optimization, the following improvements in three key metrics were achieved for each hardware platform:

- ARM Cortex-A73: execution time reduced by 35%, energy consumption by 50%, and memory usage by 25%;

- Intel Xeon Platinum: execution time reduced by 30%, energy consumption by 35%, and memory usage by 35%;

- NVIDIA RTX 3080: execution time reduced by 25%, energy consumption by 35%, and memory usage by 24%.

Figure 1 presents a diagram of the average performance across the three evaluated hardware platforms when applying the genetic algorithm.

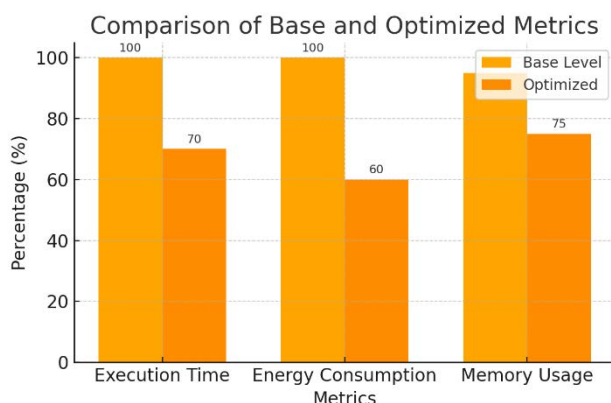


Figure 1 – Performance diagram when using the GA

5 RESULTS

In this work, alongside the theoretical presentation of applying a genetic algorithm to determine optimal compiler parameters, some practical testing results of the developed software were also partially presented for three different hardware platforms.

Testing demonstrated (Figure 1):

- a reduction in execution time by up to 30% on average across platforms, with the most significant improvement observed on the GPU due to parallelization;
- an average energy consumption decrease of 40% across platforms, especially noticeable on the ARM Cortex-A73 due to memory access optimization;
- an average memory usage improvement of 25%, achieved by eliminating redundant operations with memory and registers.

6 DISCUSSION

Genetic algorithms are particularly well suited for tasks related to adaptive compilation for several reasons:

1. They can efficiently explore large solution spaces, even when those spaces contain many local minima. This is especially important in compilation, where various parameters can interact in complex ways, affecting program performance.

2. GAs can be tuned to solve different optimization problems, including minimizing execution time, energy consumption or memory usage. In mobile devices, energy efficiency is the top priority, whereas in server systems, maximum performance takes precedence.

3. GAs can dynamically adjust compilation parameters based on an analysis of the target platform's characteristics. For programs running on GPUs [10], the algorithms can identify optimal loop unrolling and thread distribution parameters. For ARM processors, GAs may propose strategies that minimize energy consumption by reducing memory operations.

Despite their numerous advantages, using genetic algorithms in compilers also presents some limitations. First, the computational complexity of GAs can be significant, especially when the parameter space is large. This can slow down the compilation process, which is particularly critical in CI/CD pipelines. Second, the

quality of the discovered solution heavily depends on the fitness function. If the performance metrics do not fully reflect the application's needs, GAs may lead to suboptimal outcomes.

CONCLUSIONS

The use of genetic algorithms in adaptive compilers significantly improves the efficiency of the compilation process by automating the selection of optimal parameters for different architectures. The study results demonstrated:

- a reduction in execution time by up to 30% on average across platforms, with the greatest improvement observed on GPU due to parallelism;
- a reduction in energy consumption by 40% on average across platforms, especially evident on ARM Cortex-A73 thanks to memory access optimization;
- energy consumption decreased by 18% on average for mobile platforms;
- manual optimization was replaced with automatic tuning, reducing development time.

The scientific novelty lies in the development of a new approach to adaptive compilation based on the use of genetic algorithms to optimize compilation parameters depending on the architectural features of the hardware platform. Within the scope of this work, the following were proposed:

- an innovative method for dynamic tuning of compilation parameters using genetic algorithms that enables automatic adaptation to heterogeneous computing environments;
- a method for selecting optimization pass sequences for compilers based on evolutionary algorithms, which improves the performance of the generated code compared to static optimization levels (-O2, -O3);
- a mechanism for integrating genetic algorithms with modern compilers (LLVM, GCC) to analyze and adapt intermediate representation (IR) during compilation;
- a multi-objective optimization method for balancing trade-offs between performance, energy consumption and memory usage;
- an automated process for discovering optimal compilation configurations, reducing the need for manual tuning and accelerating software development.

The practical significance lies in implementing genetic algorithms in compiler technologies to increase the efficiency of cross-platform programming. Key aspects of practical application include:

- automation of the compilation optimization process, eliminating the need to manually select parameters for different architectures and reducing code tuning time;
- increased software performance, reducing code execution time by 30% compared to traditional optimization levels (-O2, -O3);
- 25% memory optimization achieved by eliminating redundant memory and register operations;
- reduced energy consumption of computing systems, which is critical for mobile devices and energy-efficient servers (up to 40% energy savings through memory access optimization);

– optimized use of hardware resources for both CPU and GPU, which is crucial in high-performance computing (HPC) and cloud technologies;

– flexibility and adaptability of the approach, making it suitable for a wide range of environments from embedded systems to large data centers.

Future research in this field may focus on:

– expanding algorithmic optimization methods: using hybrid approaches that combine genetic algorithms with machine learning to predict optimal compilation parameters;

– optimizing runtime performance: integrating approaches for Just-In-Time (JIT) compilation to adapt code execution to runtime environment characteristics;

– modeling adaptive strategies for heterogeneous systems: exploring effective mechanisms for distributing computing resources across CPU, GPU and FPGA;

– application in continuous integration and deployment (CI/CD): developing optimization methods that can be embedded into modern software development pipelines for automatic adaptation of code to new platforms;

– reducing the computational complexity of the optimization process: applying heuristic approaches to decrease the number of algorithm iterations without sacrificing optimization quality.

ACKNOWLEDGEMENTS

The research was conducted at the Department of Software Engineering of the National Technical University “Dnipro Polytechnic” as part of the state-funded scientific project: “Methods, Models and Technologies for Data Processing and Analysis in Computerized Systems” (State Registration No. 0124U004583, Project No. E-346).

REFERENCES

1. Tagtekin B., Höke B., Sezer M. K., Öztürk M. U. FOGA: Flag Optimization with Genetic Algorithm, *Proc. IEEE Int. Conf. on Innovations in Intelligent Systems and Applications (INISTA 2021)*, 29 June–1 July 2021, Kocaeli, Turkey, IEEE, 2021, pp. 1–6. DOI: 10.1109/INISTA52462.2021.9456364.
2. Júnior N. L. Q., Rodriguez L. G. A., Silva A. F. da Combining Machine Learning with a Genetic Algorithm to Find Good Compiler Optimizations Sequences, *Proc. 19th Int. Conf. on Enterprise Information Systems (ICEIS 2017)*, Vol. 1. Porto, Portugal, 26–29 Apr. 2017, pp. 397–404. DOI: 10.5220/0006270403970404.
3. Peeler H., Li S., Sloss A., Reid K., Yuan Y., Banzhaf W. Optimizing LLVM Pass Sequences with Shackleton: A Linear Genetic Programming Framework, *Companion Proc. Genetic and Evolutionary Computation Conf. (GECCO '22)*, 9–13 July 2022. Boston, USA, ACM, 2022, pp. 578–581. DOI: 10.1145/3520304.3528945.
4. Ashouri A. H., Killian W., Cavazos J., Palermo G., Silvano C. A Survey on Compiler Autotuning Using Machine Learning, *ACM Computing Surveys*, 2018, Vol. 51, № 5, Art. 96, 42 p. DOI: 10.1145/3197978.
5. Rosario V. M. do, Borin E., Nacul A., Breternitz M., Collange S., Barrett E. Smart Selection of Optimizations in Dynamic Compilers, *Concurrency and Computation: Practice and Experience*, 2021, Vol. 33, № 10, e6089. DOI: 10.1002/cpe.6089.
6. Cummins C., Fisches Z. V., Ben-Nun T., Hoefler T., O’Boyle M. F. P., Leather H. ProGraML: A Graph-Based Program Representation for Data Flow Analysis and Compiler Optimizations, *Proc. 38th Int. Conf. on Machine Learning (ICML 2021)*, 18–24 July 2021, Virtual Event. PMLR, 2021, Vol. 139, pp. 2244–2253.
7. Nugteren C., Codreanu V. CLTune: A Generic Auto-Tuner for OpenCL Kernels, *Proc. IEEE 9th Int. Symp. on Embedded Multicore/Many-core Systems-on-Chip (MCSoc 2015)*, 23–25 Sept. 2015. Turin, Italy, IEEE, 2015, pp. 195–202. DOI: 10.1109/MCSoc.2015.33.
8. Ashouri A. H., Palermo G., Cavazos J., Silvano C. Automatic Tuning of Compilers Using Machine Learning. Cham, Springer, 2018, 78 p. (PoliMI SpringerBriefs in Applied Sciences). DOI: 10.1007/978-3-319-71489-9.
9. Martins L. G. A., Nobre R., Cardoso J. M. P., Delbem A. C. B., Marques E. Clustering-Based Selection for the Exploration of Compiler Optimization Sequences, *ACM Transactions on Architecture and Code Optimization (TACO)*, 2016, Vol. 13, № 1, Art. 14. DOI: 10.1145/2883614.
10. Liou J.-Y., Wang X., Forrest S., Wu C.-J. GEVO: GPU Code Optimization Using Evolutionary Computation, *ACM Trans. on Architecture and Code Optimization*, 2020, Vol. 17, № 4, Art. 33, 28 p. DOI: 10.1145/3418055.

Received 21.04.2025.
Accepted 09.09.2025.

ВИКОРИСТАННЯ ГЕНЕТИЧНИХ АЛГОРИТМІВ В АДАПТИВНИХ КОМПІЛЯТОРАХ ДЛЯ КРОСПЛАТФОРМНОЇ ОПТИМІЗАЦІЇ

Бердник М. Г. – д-р техн. наук, доцент, професор кафедри програмного забезпечення комп’ютерних систем, Національний технічний університет «Дніпровська політехніка», Дніпро, Україна.

Стародубський І. П. – аспірант кафедри програмного забезпечення комп’ютерних систем, Національний технічний університет «Дніпровська політехніка», Дніпро, Україна.

АНОТАЦІЯ

Актуальність. Сучасне програмне забезпечення розробляється в умовах постійно зростаючої складності обчислювальних систем. Сьогодні розробникам доводиться враховувати величезну різноманітність платформ: від мобільних пристроїв із обмеженими ресурсами до серверів із високопродуктивними процесорами та спеціалізованих архітектур, таких як GPU, FPGA та навіть квантові комп’ютери. Ця гетерогенність вимагає від програмного забезпечення здатності ефективно працювати на різних апаратних платформах. Однак переносимість програм залишається однією з найскладніших задач, особливо коли йдеться про досягнення високої продуктивності.

Одним із перспективних напрямів вирішення цієї проблеми є використання адаптивних компіляторів, які здатні автоматично оптимізувати програмний код для різних архітектур. Такий підхід дозволяє розробникам зосередитися на функціональній частині програм, мінімізуючи витрати на оптимізацію та налаштування. Особливе місце серед підходів до створення адаптивних компіляторів займають генетичні алгоритми (ГА). Це потужні еволюційні методи, які дозволяють знаходити оптимальні рішення у складних і багатовимірних параметрів.

Мета роботи. Метою дослідження є застосування генетичних алгоритмів у процесі адаптивної компіляції для забезпечення автоматичної оптимізації програмного забезпечення на різних апаратних платформах.

Метод. Ґрунтується на генетичних алгоритмах для автоматизації процесу компіляції. Основними етапами якого є: ініціалізація популяції – створення початкового набору параметрів компіляції; оцінка функції пристосованості – визначення ефективності кожної комбінації параметрів; еволюційні операції застосування схрещування, мутації та відбору для створення нового покоління параметрів; критерії зупинки – завершення ітераційного процесу при досягненні оптимального результату або стабілізації метрик.

Результати. Розроблений алгоритм був реалізований на Python із використанням бібліотек numpy, multiprocessing і subprocess. Для оцінки ефективності алгоритму використовувалися метрики часу виконання, споживання енергії та обсягу використаної пам’яті.

Висновки. Наукова новизна дослідження полягає в розробки: інноваційного підходу до автоматичної оптимізації параметрів компіляції, заснованому на використанні генетичних алгоритмів; метода динамічного вибору стратегій оптимізації на основі метрик продуктивності для різних архітектур; інтеграції ГА із сучасними компіляторами, такими як LLVM, для автоматичного аналізу проміжного подання та оптимізації кодів; методів застосування адаптивних компіляторів для вирішення задач кросплатформної оптимізації. Практична значимість визначається у використанні генетичних алгоритмів в адаптивних компіляторах, що дозволяє значно підвищити ефективність процесу компіляції, автоматизуючи вибір оптимальних параметрів для різних архітектур. Запропонований підхід може бути успішно застосований у галузях мобільних обчислень, хмарних технологій і високопродуктивних систем.

КЛЮЧОВІ СЛОВА: генетичні алгоритми, адаптивні компілятори, кросплатформна оптимізація, автоматизація.

ЛІТЕРАТУРА

1. FOGA: Flag Optimization with Genetic Algorithm / [B. Tagtekin, B. Höke, M. K. Sezer, M. U. Öztürk] // Proc. IEEE Int. Conf. on Innovations in Intelligent Systems and Applications (INISTA 2021), 29 June–1 July 2021, Kocaeli, Turkey. – IEEE, 2021. – P. 1–6. – DOI: 10.1109/INISTA52462.2021.9456364.
2. Júnior N. L. Q. Combining Machine Learning with a Genetic Algorithm to Find Good Compiler Optimizations Sequences / N. L. Q. Júnior, L. G. A. Rodriguez, A. F. da Silva // Proc. 19th Int. Conf. on Enterprise Information Systems (ICEIS 2017). – Vol. 1. – Porto, Portugal, 26–29 Apr. 2017. – P. 397–404. – DOI: 10.5220/0006270403970404.
3. Optimizing LLVM Pass Sequences with Shackleton: A Linear Genetic Programming Framework / [H. Peeler, S. Li, A. Sloss et al.] // Companion Proc. Genetic and Evolutionary Computation Conf. (GECCO ’22), 9–13 July 2022, Boston, USA. – ACM, 2022. – P. 578–581. – DOI: 10.1145/3520304.3528945.
4. A Survey on Compiler Autotuning Using Machine Learning / [A. H. Ashouri, W. Killian, J. Cavazos et al.] // ACM Computing Surveys. – 2018. – Vol. 51, № 5. – Art. 96. – 42 p. – DOI: 10.1145/3197978.
5. Smart Selection of Optimizations in Dynamic Compilers / [V. M. do Rosario, E. Borin, A. Nacul et al.] // Concurrency and Computation: Practice and Experience. – 2021. – Vol. 33, № 10. – e6089. – DOI: 10.1002/cpe.6089.
6. Cummins C. ProGraML: A Graph-Based Program Representation for Data Flow Analysis and Compiler Optimizations / [C. Cummins, Z. V. Fisches, T. Ben-Nun et al.] // Proc. 38th Int. Conf. on Machine Learning (ICML 2021), 18–24 July 2021, Virtual Event. – PMLR, 2021. – Vol. 139. – P. 2244–2253.
7. Nugteren C. CLTune: A Generic Auto-Tuner for OpenCL Kernels / C. Nugteren, V. Codreanu // Proc. IEEE 9th Int. Symp. on Embedded Multicore/Many-core Systems-on-Chip (MCSoc 2015), 23–25 Sept. 2015, Turin, Italy. – IEEE, 2015. – P. 195–202. – DOI: 10.1109/MCSoc.2015.33.
8. Automatic Tuning of Compilers Using Machine Learning / [A. H. Ashouri, G. Palermo, J. Cavazos, C. Silvano]. – Cham : Springer, 2018. – 78 p. – (PoliMI SpringerBriefs in Applied Sciences). – DOI: 10.1007/978-3-319-71489-9.
9. Clustering-Based Selection for the Exploration of Compiler Optimization Sequences / [L. G. A. Martins, R. Nobre, J. M. P. Cardoso et al.] // ACM Transactions on Architecture and Code Optimization (TACO). – 2016. – Vol. 13, № 1. – Art. 14. – DOI: 10.1145/2883614.
10. GEVO: GPU Code Optimization Using Evolutionary Computation / [J.-Y. Liou, X. Wang, S. Forrest, C.-J. Wu] // ACM Trans. on Architecture and Code Optimization. – 2020. – Vol. 17, № 4. – Art. 33. – 28 p. – DOI: 10.1145/3418055.

ANALYSIS OF PROCEDURES FOR VOICE SIGNAL NORMALIZATION AND SEGMENTATION IN INFORMATION SYSTEMS

Pastushenko M. S. – PhD, Professor, Professor of V. V. Popovskyy Department of Infocommunication Engineering, Kharkiv National University of Radio Electronics, Kharkiv, Ukraine.

Pastushenko O. M. – PhD, Senior Researcher, Serviceman of the Armed Forces of Ukraine, Ukraine.

Faizulaiev T. A. – Director of “TAF-87” LLC, Kharkiv, Ukraine.

ABSTRACT

Context. The current task of evaluating formant data (formant frequencies, their spectral density level, amplitude-frequency spectrum envelope, formant frequency spectrum width) in voice authentication systems is considered. The object of the study is the process of digital preprocessing of the voice signal when extracting formant data.

Objective. Evaluation of the effectiveness of traditional procedures for digital preprocessing of a user voice signal and development of proposals for improving the quality of formant data extraction.

Method. A mathematical model for extracting formant data from an experimental voice signal has been developed to study the influence of normalization and segmentation procedures on the quality of the resulting estimates. By modeling the process of extracting formant data, the results of digital processing of normalized and non-normalized voice signals are compared. The influence of the processed frame duration of the experimental voice signal on the quality of the formant frequencies assessment is estimated. The results are obtained for the experimental phoneme and morpheme.

Results. The obtained results show that when processing a voice signal with a sufficient signal-to-noise ratio, normalization procedures are not mandatory when extracting formant data. Moreover, normalization leads to a less accurate measurement of the spectrum width of formant frequencies. It is also unacceptable to use a processed frame duration of less than 40 ms. These results allow us to modify the traditional method of voice signal preprocessing. The use of the modeling method in the study of the experimental voice signal confirms the reliability of the results obtained.

Conclusions. The scientific novelty of the research results lies in the modification of the voice signal preprocessing methodology in authentication systems. Eliminating normalization procedures at high signal-to-noise ratios of the voice signal, which occurs in user authentication systems, makes it possible to increase the speed of formant data extraction and more accurately estimate the width of the formant frequency spectrum. Selecting a frame duration of at least 40 ms for the processed signal significantly improves the accuracy of formant frequency determination. Otherwise, the estimates of the formant frequencies will be high. Moreover, when processing phonemes, the processed voice signal cannot be divided into frames. Practical application of research results allows to increase the efficiency and accuracy of the formant data generation. Prospects for further research may be studies of the influence of normalization and framing procedures on other elements of a template of the authentication system user.

KEYWORDS: authentication, voice signal, normalization, segmentation, formant data.

NOMENCLATURE

i is an analyzed element of the voice signal;

N_o is a number of elements in a processed voice signal window;

s_i is a non-normalized voice signal element;

s_{ni} is a normalized voice signal element;

$s_{\max}$ is a maximum value of the absolute values of a processed voice signal sample;

abs is a designation of the absolute value of an element being analyzed;

$\max$ is a designation of the maximum value selection function;

T is the duration of a processed section (frame) of a voice signal in seconds;

df is a frequency resolution of spectral analysis in hertz;

f_d is a sampling frequency of an analyzed voice signal in hertz.

institutions of all forms of ownership and purposes is becoming more acute. Information protection and reliable operation of computer systems is one of the most important tasks, which largely depend on the methods of authenticating users and processes. In information systems, authentication procedures are becoming a key component in ensuring security. Authentication systems are widely used in computer systems, mobile devices, online services and play a decisive role in confirming user access rights to systems and their resources.

The known types of authentication have their advantages and disadvantages, and the practice of their application shows their low reliability. In this regard, intensive research is being conducted on the use of biometric user features in authentication systems and various access systems. It is known that among biometric authentication systems, voice systems are preferable by the efficiency/cost criterion [1]. Moreover, modern voice systems do not implement all the achievements of digital signal processing, which are widely and effectively used in radio communications and radiolocation. Currently, the Ukrainian state bank PrivatBank is implementing voice authentication systems, which emphasizes the relevance of the research.

INTRODUCTION

In modern society, the problem of information security of computer systems in various organizations and

The theoretical basis of voice authentication procedures is formed by research conducted in the field of speech recognition, which began in the middle of the last century. Both in the field of speech recognition and in authentication systems, digital processing of the voice signal begins with the preprocessing stage. Then, in authentication systems, specific procedures for forming a user template are performed. The stage of extracting features from the voice signal and placing them in the template is the second stage of voice signal processing. At the third stage, voice systems compare the current user template with the data placed in the template database, and a decision is made on the user access to resources.

User features extracted from a voice signal typically include: pitch frequency, formant data, cepstral and mel-frequency cepstral coefficients, etc. Formant data play a special role, including: formant frequencies; their spectral density level; amplitude-frequency spectrum envelope; spectrum width of formant frequencies.

A formant is a resonant frequency in the human vocal tract that provides a unique timbre and quality of speech sound. Formant frequencies are created by the structure of the vocal tract, the shape and position of the tongue and lips during speech. Formants are decisive in phonetics and speech processing in all practical applications.

Formants also play a decisive role in a new practical task, namely, voice synthesis.

In this case, modern scientific literature focuses on issues of evaluating formant frequencies, for which frequency filters, spectral analysis, wavelet transform, neural networks, etc. are used. In practical applications, preference is given to spectral methods.

The object of study is the process of digital preprocessing of a voice signal during the extraction of formant data.

The subject of the study is the sampling methods and procedures for extracting formant data from the voice signal of the authentication system user. At the preprocessing stage, the processed signal is traditionally normalized and divided into separate sections (frames) to extract user features.

The purpose of the work is to evaluate the effectiveness of traditional procedures for digital preprocessing of a user voice signal and to develop proposals for improving the quality of formant data extraction.

1 PROBLEM STATEMENT

It is known that voice signal normalization is an important stage of preprocessing in voice authentication systems, aimed at bringing the signal to a standardized form to improve recognition accuracy. Even simple normalization, which comes down to dividing each element of the processed voice signal array s_i by the maximum value of the absolute values of the processed data

$$s_{ni} = s_i / s_{\max}, \quad (1)$$

where

$$s_{\max} = \max(\text{abs}(s_i)) \quad (2)$$

for all analyzed elements within the range of change i from 1 to N_o , allows to significantly increase the signal-to-noise ratio. Here N_o is the number of elements in the processed voice signal window. Also, the processed data is converted to the range from -1 to 1 .

However, authentication systems typically process voice signals with a sufficient signal-to-noise ratio (20 dB or more) and the feasibility of normalization procedures for extracting formant data is questionable.

In order to extract user features, authentication systems are further recommended to process in a sliding window with some overlap. In this case, the duration of the processed section (frame) is selected in the range from 20 to 40 ms. Note that in most cases, spectral analysis is used to extract user features, including formant data. It is known that the frequency resolution of spectral analysis is determined by the duration of the processed signal, namely

$$df = 1/T, \quad (3)$$

where T is the duration of the processed section (frame) of the voice signal in seconds. For a discrete signal and digital processing, the specified relationship is transformed to the following form

$$df = f_d / N_o, \quad (4)$$

where f_d is the sampling frequency of the voice signal in hertz.

Setting the window duration to 20 ms opens the question of the formant data extraction quality for authentication purposes.

2 REVIEW OF THE LITERATURE

The theoretical foundations for the acoustic spectral theory of speech were formulated quite a long time ago by the German scientist H. Helmholtz, which are still used today. Here we will also note the works of the Swedish acoustician G. Fant, who developed the theory of distinctive features – a universal acoustic classification of sounds. In the proposed classification, a special place is given to formant data.

The theoretical foundations of digital processing of speech signals are presented in the works of Rabiner L.R. and co-authors, for example, [2]. This is one of the first textbooks in which the concepts of normalization, segmentation (framing), window functions and frame overlapping were introduced.

We would like to draw attention to the modern fundamental work [3]. This work provides detailed research on the issues of digital processing of voice signals, including the issues considered in the article.

Formant data form the basis of modern templates in authentication systems, but they are also used in a number

of other speech-processing applications, including medicine. It is known that the first two formants determine the results of speech recognition, while the third and fourth play a significant role in speaker identification [3].

One of the main procedures of preliminary processing of a voice signal is its normalization. A large number of articles is devoted to the problem of normalization. Let us pay attention to the work [4], which considers various mechanisms of normalization in the recognition of English vowel sounds. It was found that the first and second formants of vowel sounds depend on the size of the vocal tract. It was noted that normalization can partially eliminate this effect. At the same time, it was concluded that for better perception of vowel sounds, preference should be given to simpler normalization procedures.

In [5] the results of studies are presented that indicate the connection between perception and pronunciation. A comparative analysis of the original voice signals perception and their normalized representations is performed. The problem was solved using the example of language learning by people who were not native speakers.

Formant frequencies are increasingly analyzed not only for speech recognition or authentication purposes, but also for other purposes. For example, in [6], the length of the vocal tract is estimated and scale-invariant representations of formant templates are created. Normalization is widely used, both when processing individual sounds and in the interests of combining the estimates obtained for different sounds.

The paper [7] investigates the use of vowel formants for text-independent speaker authentication, including methods for extracting and filtering formants. Normalization procedures and digital processing in the form of individual frames are widely used.

Audio processing has become an integral part of modern applications in areas ranging from healthcare to speech-controlled devices. At the same time, segmentation of the analyzed data plays an important role in audio signal processing. Neural networks and deep learning have been increasingly used to solve this problem. Therefore, the work [8] deserves attention. The analysis presented in this paper confirms and establishes the importance of deep learning methods in audio segmentation.

In [9] a model based on non-negative matrix factorization was developed and investigated. Using procedures for segmenting the analyzed signal, the developed model allows dividing the analyzed data into speech, noise and music.

In the article [10] a compact segmentation model is developed that improves the quality of speech translation by framing and pre-training with punctuation. The advantage of the developed model is the increased efficiency of solving the speech translation problem.

3 MATERIALS AND METHODS

A mathematical model for extracting format data from a voice signal has been developed to study the influence of normalization and segmentation procedures on the quality of the resulting estimates. The experimental voice

signal containing the word “odyn” (“one” in Ukrainian) with a sampling frequency of 64 kHz was analyzed. This signal is presented in Fig. 1.

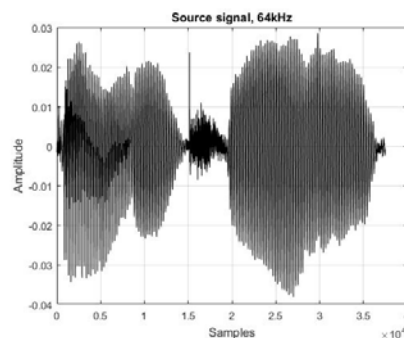


Figure 1 – The analyzed voice signal

We will conduct the study of the proposed signal in three stages. In the first stage, we will analyze the phoneme “o”, which makes up the first part of the analyzed signal. In the second stage, we will analyze the morpheme “dyn”, which makes up the second half of the analyzed signal. Information on the processing of the entire signal will also be obtained.

The main method of evaluating formant data will be spectral analysis of the amplitude-frequency spectrum of the voice signal under study. The spectrum was evaluated using the fast Fourier transform. In this case, we will pay some attention to the voice signal normalization procedures. The normalization procedures will be implemented using the norm method. As a norm, we will use the maximum value of the amplitude of the signal under consideration.

The main studies concern the assessment of the influence of the processed frames duration (sections of the voice signal) on the results of the evaluated formant data. Here, three variants of frame duration are also considered: 20 and 40 ms, as well as the entire analyzed signal.

Currently, when extracting user features from a voice signal, the latter is divided into frames, which are recommended to be selected with a duration of 20 ms. Digital processing is carried out in a “sliding window” that has some overlap.

This approach leads to the fact that spectral analysis must be carried out on a limited amount of initial information. The latter leads to a significant decrease in the resolution of the spectrum, and therefore the accuracy of the formant frequency assessment.

Some attention is paid to the assessment of the influence of the Hamming window function, which is widely used in digital processing of voice signals. The purpose of this function is to eliminate the effect of spectrum “spreading”. This approach will allow to compare traditional approaches to digital voice signal processing and develop practical recommendations for significantly improving the quality of formant data estimates.

4 EXPERIMENTS

The first stage of the research is associated with the analysis of the experimental phoneme “o”, which is presented in Fig. 2. The sampling frequency of the voice signal was 64 kHz.

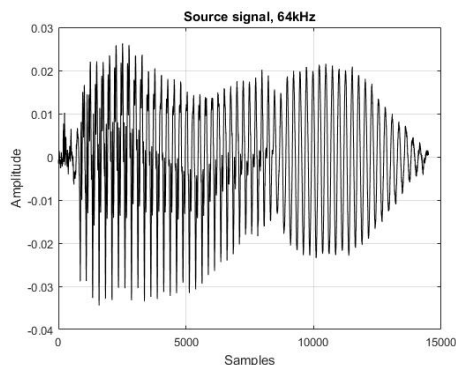


Figure 2 – The analyzed phoneme is “o”

We will begin the analysis with a study of voice signal normalization procedures. Interest in normalization procedures is conditioned by the fact that fundamental and scientific literature has not disclosed in detail the influence of voice signal normalization procedures during its digital processing. The question of the influence of normalization procedures on the quality of generating formant data also remains open.

In addition, the analysis of the Hamming window influence on the formant data of the voice signal is of interest. Before presenting the results of the experimental studies, it is necessary to make a number of comments.

Firstly, the polyharmonic voice signal being processed is recommended to be divided into frames (segments) of 20 ms in known works. This is justified by the fact that in this section the signal being studied will be quasi-stationary. In some works, it is recommended to choose this section of 40 ms.

Secondly, the Hamming window is one of the weighting functions that can be used as one of the characteristics of nonparametric spectral analysis. As stated in a number of works, the purpose of window function processing is to eliminate spectrum spreading.

Thirdly, the real effect during processing of the experimental voice signal is the elimination of anomalous estimates of formant frequencies.

Thus, the procedures of digital processing of the experimental voice signal allowed processing both normalized and non-normalized signals. In addition, to assess the influence of the window function, three duration variants of 20 and 40 ms, as well as a signal of full duration, were studied. Here we will also study the second part of the voice signal under consideration (see Fig. 1), namely the morpheme “dyn”. This will allow us to clarify the features of processing signals including different voice components.

The mathematical model for conducting experimental studies was implemented in the environment of the MatLab computer mathematics system. This approach al-

lowed us to significantly simplify the modeling procedures by using standard MatLab functions. In addition, the use of standard functions allows increasing the reliability of the scientific results obtained.

The mathematical model was based on standard procedures for extracting formant data for different characteristics of the processed voice signal.

5 RESULTS

As the results of experimental studies of the phoneme “o” show, the implementation of voice signal normalization procedures essentially has little effect on formant data. At the same time, the spectral density levels of formant frequencies increase by approximately 30 dB. The spectrum envelope, in this case, moves to the positive region. However, the shape of the normalized amplitude-frequency spectrum envelope does not change.

Normalization does not affect the value of formant frequencies. The values of the estimates of the six analyzed frequencies do not change.

The most significant changes occur when assessing the width of the spectrum of formant frequencies. The changes are associated with an increase in the level of spectral power after normalization procedures. Fig. 3 shows the spectral density peak of the second formant frequency of the phoneme “o”. The width of the formant frequency spectrum was measured at a level of 0.8 (dash-dotted line) from the maximum value. The solid curve shows the peak of the non-normalized formant frequency, and the dashed curve shows the peak of the normalized formant frequency.

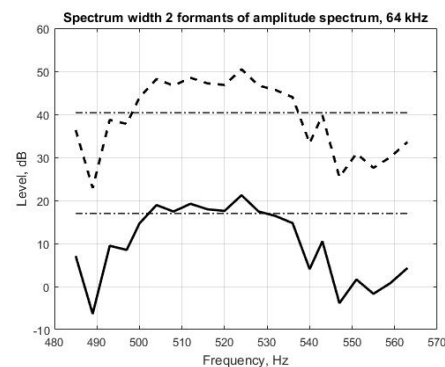


Figure 3 – To the assessment of the formant spectrum width

After the normalization procedures, the spectrum width of the formant frequency increases. The latter is conditioned by the influence of the formant harmonic peak shape and the accuracy of spectrum calculation.

Thus, it is advisable to estimate the spectrum width of formant frequencies for a non-normalized signal; the latter will improve the accuracy of its determination.

Let us proceed to the analysis of the influence of the Hamming window duration on the results of the formant data evaluation. The Hamming window duration corresponded to the duration of the processed section of the voice signal. As indicated above, three variants of the frame duration of the processed voice signal were consid-

ered: 20 ms, 40 ms and the fully analyzed phoneme “o”. Then the spectrum of the analyzed section was calculated using the fast Fourier transform. Formant frequencies were determined based on the amplitude-frequency spectrum.

When using short frames of 20 and 40 ms, the processing was carried out in a sliding window with some overlap. The need for such a processing option will be explained below. For frames of 20 and 40 ms, the mean estimates were used as estimates of formant frequencies. The results of the processing are presented in Fig. 4.

The following notations are used in this Figure: the dashed-dotted line shows the results when processing 20 ms sections, and the dashed line shows 40 ms sections. The solid line shows the processing of the full phoneme. In this case, average estimates were used in short sections (20 and 40 ms). We will consider the estimates obtained when processing the full phoneme to be true.

Let us analyze the presented dependencies.

With a frame duration of 20 ms, after processing by the Hamming window, the mean square deviation is significantly reduced and anomalous measurements of formant frequencies are eliminated. The accuracy of spectrum calculation was approximately 32 Hz. Note that the low accuracy is caused by the insignificant number of samples in the frame. This also leads to insufficient accuracy in determining formant frequencies. As a rule, in this case we obtain high estimates.

Increasing the frame by two times leads to an increase in the accuracy of determining the formant frequencies, which tend to the estimates obtained when processing the phoneme as a single frame. The accuracy of the spectrum estimate for a 40 ms frame was approximately 16 Hz. In this case, the estimates of the standard deviation are also higher. The accuracy of determining the formant frequencies increases significantly.

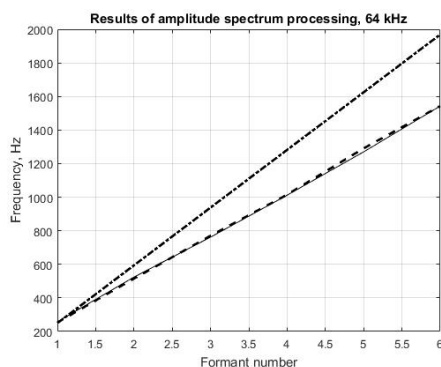


Figure 4 – The estimates of formant frequencies of the phoneme “o”

When processing the entire phoneme “o” as a single frame, the use of the Hamming window function is inappropriate. The use of this function leads to a decrease in the accuracy of the sixth formant estimate. The accuracy

of the spectrum resolution in this case was 4 Hz. Errors in determining formant frequencies when processing 40 ms frames and the full frame of the analyzed phoneme are presented in Table 1. As a rule, the estimates obtained in the 40 ms frame are high. It should be noted that with an increase in the number of formants, the values of the estimates of the standard deviation for these frequencies increase.

Table 1 – Errors in determining the formant frequencies of the phoneme “o”

F1, Hz	F2, Hz	F3, Hz	F4, Hz	F5, Hz	F6, Hz
4.2	8.6	6.4	3	21	1

Now let us move on to the analysis of the morpheme (syllable) “dyn”. As before, we will consider the three variants of digital processing of the voice signal considered above. The results of processing the specified morpheme are presented in Fig. 5.

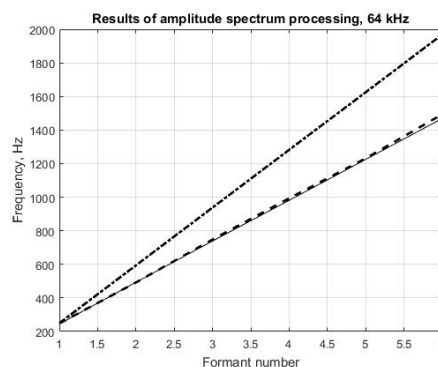


Figure 5 – Formant frequency estimates of the morpheme “dyn”

The analysis of the presented dependencies allows us to draw the following conclusions. As before, the least accurate estimates of formant frequencies are obtained when using a 20 ms frame. They are significantly high. Estimates of the average formant frequencies for a 40 ms frame are also slightly higher than the values obtained when processing the entire morpheme. The errors are presented in Table 2. Estimates of the standard deviation increase with an increase in the formant number.

Table 2 – Errors in determining the formant frequencies of the morpheme “dyn”

F1, Hz	F2, Hz	F3, Hz	F4, Hz	F5, Hz	F6, Hz
3	1	9	12	6	24

Here, attention should be paid to the change in the estimates of the second formant frequency for the 40 ms frame depending on the number of the implementation. This dependence is presented in Fig. 6 and makes it possible to divide the analyzed morpheme into separate sounds.

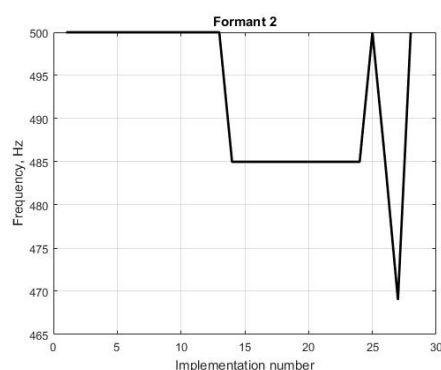


Figure 6 – Estimates of the second formant frequency of the morpheme “dyn”

Table 3 presents estimates of the analyzed formant frequencies for different source information.

When processing the word “odyn”, the estimates of formant frequencies were calculated as underestimated, since the section between syllables was not excluded. This can be seen in Fig. 7, which shows the dependence of the estimate of the first formant frequency on the number of the realization.

Table 3 – Formant frequency estimates

Type of experiment	F1, Hz	F2, Hz	F3, Hz	F4, Hz	F5, Hz	F6, Hz
“o”	258	524	762	1012	1270	1543
“dyn”	245	493	739	981	1225	1469
“odyn”	247	494	776	1015	1268	1515

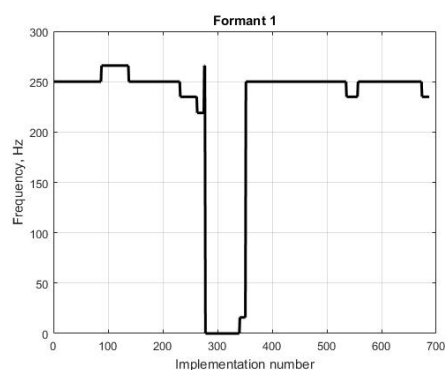


Figure 7 – Estimates of the first formant frequency of the word “odyn”

Fig. 7 shows that some estimates are equal to zero, which corresponds to the section of registration materials between syllables. The latter fact can be used to divide the registration materials into syllables.

6 DISCUSSION

First, let us briefly analyze the influence of normalization procedures on the generation of formant data. Along with the known factors of the influence of normalization procedures on the suppression of the noise component and the increase in the signal-to-noise ratio, in essence, the

indicated procedures do not lead to a change in the formant data.

An exception is the assessment of the spectrum width of formant frequencies, which after normalization is less accurate. Although with more complex digital processing procedures, normalization can lead to a positive effect, primarily due to an increase in the signal-to-noise ratio [11]. Therefore, if it is necessary to assess the spectrum width of formant frequencies, it is advisable to perform these procedures on a non-normalized voice signal.

More interesting results have been obtained when studying the effect of the processed frame duration on formant data. Literature sources on voice signal processing recommend choosing a frame duration of 20 ms to extract user features. However, as the research results show, in this case we get overvalued estimates of formant frequencies. In this case, the spectrum resolution will also be low.

Doubling the frame length results in a significant increase in the accuracy of determining formant frequencies. More accurate estimates can be obtained by processing the full phoneme “o”. Obviously, when digitally processing phonemes, it is advisable not to divide them into frames.

When processing morphemes, it is advisable to choose a frame length of 40 ms. Using such a frame allows dividing the processed morpheme into individual sounds.

CONCLUSIONS

The current scientific problem of studying the influence of normalization procedures and the duration of the processed frame on the assessment of voice signal formant data of the authentication system user is considered. The study was conducted using the developed mathematical model and the experimental voice signal.

The scientific novelty of the obtained results is in the fact that normalization procedures do not change the formant data when evaluating them. Moreover, the estimates of the spectrum width of formant frequencies are less accurate. The duration of the processed frame of 20 ms is unacceptable, since the estimates of formant frequencies will be significantly high. The frame length of 40 ms can be used to divide the processed voice signal into individual sounds. This will improve the accuracy and reliability of the extracted estimates of formant frequencies. More accurate results of formant data are obtained when processing the complete analyzed signal. Thus, the results obtained allow modifying the traditional method for preprocessing voice signals in authentication systems.

The practical significance of the research results is that in some cases it is possible to exclude normalization procedures with a sufficient signal-to-noise ratio. This approach will increase the efficiency of formant data assessment. When choosing the frame length, it is necessary to choose its duration of 40 ms or more. This will allow a more accurate assessment of formant frequencies. When processing phonemes, it is advisable to process the full signal, which will allow a more accurate and efficient assessment of formant data.

The results of the research can be applied in speech recognition, voice authentication systems, forensics, forensic examination and other applications where formant data assessments are used. Clarification of formant data allows for a better formation of a voice authentication system user template, which has a positive effect on reducing errors of the first and second kind.

Further research will focus on assessing the procedures under consideration for other components of the authentication system user template.

REFERENCES

1. Pastushenko M. O., Pastushenko M. S., Petrachenko M. O. Do pytannja ocinky efektyvnosti biometrychnykh system, *Problemy telekomunikacij*, 2023, № 1(32), pp. 37–44. DOI: <https://doi.org/10.30837/pt.2023.1.03>
2. Rabiner L. R., Schafer R. W. Digital Processing of Speech Signals. NJ, Prentice-Hall, Inc., 1978, 512 p. URL: https://ie.u-ryu.ac.jp/~asharif/pukiwiki/attach/Acoustic%20Speech%20Signal%20Processing_Prentice%20Hall%20-%20Digital%20Processing%20of%20Speech%20Signals.pdf
3. Beigi H. Fundamentals of Speaker Recognition. NY, Springer, 2011, 942 p. DOI:10.1007/978-0-387-77592-0
4. Persson A., Barreda S., Jaeger T. F. Comparing normalization against US English listeners' vowel perception, *The Journal of the Acoustical Society of America*, 2025, Vol. 157, № 2, pp. 1458–1482. DOI: <https://doi.org/10.1121/10.0035476>
5. Clopper C. G., Dossey E., Gonzalez R. Raw acoustic vs. normalized phonetic convergence: Imitation of the Northern Cities Shift in the American Midwest, *Laboratory Phonology*, 2024, Vol. 15(1), pp. 1–15. DOI: <https://doi.org/10.16995/labphon.10893>
6. Anikin A., Barreda S., Reby D. A Practical guide to calculating vocal tract length and scale-invariant formant patterns, *Springer Nature Link*, 2023, Vol. 56, pp. 5588–5604. DOI: 10.3758/s13428-023-02288-x
7. Almaadeed N. Aggoun, A., Amira A. Text-Independent Speaker Identification Using Vowel Formants, *Journal of Signal Processing Systems*, 2015, Vol. 82, № 3, pp. 345–356. DOI: <https://doi.org/10.1007/s11265-015-1005-5>
8. Aggarwal S., Vasukidevi G., Selvakanmani S., Pant B., Kaur K., Verma A., Binigde G. N. Audio Segmentation Techniques and Applications Based on Deep Learning, *Journal of Scientific Programming*. Wiley Online Library, 2022, Vol. 2022, pp. 1–9. DOI: <https://doi.org/10.1155/2022/7994191>
9. Lebourdais M., Mariotte T., Almudévar A., Tahon M., Ortega A. Explainable by-design Audio Segmentation through Non-Negative Matrix Factorization and Probing, *arXiv:2406.13385 [eess.AS]*, 2024, pp. 1–5. DOI: <https://doi.org/10.48550/arXiv.2406.13385>
10. Lee J., Kim S., Kim H., Chung J. S. Lightweight Audio Segmentation for Long-form Speech Translation, *arXiv:2406.10549 [eess.AS]*, 2024, pp. 1–5. DOI: <https://doi.org/10.48550/arXiv.2406.10549>
11. Pastushenko M., Krasnozheniuk Ya., Zaika M. Investigation of Informativeness and Stability of Mel-Frequency Cepstral Coefficients Estimates based on Voice Signal Phase Data of Authentication System User, *International Conference Problems of Infocommunications. Science and Technology 6–9 October 2020 (PIC S&T'2020)*. Kharkiv, Ukraine, 2020, pp. 467–472. DOI: 10.1109/PICST51311.2020.9468083

Received 11.08.2025.

Accepted 19.10.2025.

УДК 057.087.1:621.391.26

АНАЛІЗ ПРОЦЕДУР НОРМАЛІЗАЦІЇ ТА СЕГМЕНТАЦІЇ ГОЛОСОВОГО СИГНАЛУ В ІНФОРМАЦІЙНИХ СИСТЕМАХ

Пастушенко М. С. – канд. техн. наук, професор, професор кафедри інфокомунікаційної інженерії ім. В.В. Поповського Харківського національного університету радіоелектроніки, Харків, Україна.

Пастушенко О. М. – канд. техн. наук, старший науковий співробітник, військовослужбовець Збройних Сил України, Україна.

Файзулаєв Т. А. – Директор Товариства з обмеженою відповідальністю «ТАФ-87», Харків, Україна.

АНОТАЦІЯ

Актуальність. Розглядається актуальне завдання оцінки формантних даних (формантних частот, рівня їхньої спектральної щільності, огинаючу амплітудно-частотного спектру, ширини спектрів формантних частот) у системах голосової автентифікації. Об'єктом дослідження був процес цифрової попередньої обробки голосового сигналу під час вилучення формантних даних.

Мета роботи – оцінка ефективності традиційних процедур цифрової попередньої обробки голосового сигналу користувача та розробка пропозицій щодо підвищення якості вилучення формантних даних.

Метод. Розроблено математичну модель вилучення формантних даних з експериментального голосового сигналу для дослідження впливу процедур нормалізації та сегментації на якість одержуваних оцінок. Шляхом моделювання процесу отримання формантних даних порівнюються результати цифрової обробки нормалізованого і ненормалізованого голосового сигналу. Оцінюється вплив тривалості обробленого кадру експериментального голосового сигналу на якість оцінки формантних частот. Результати отримані для експериментальної фонемі та морфемі.

Результати. Отримані результати свідчать, що при обробці голосового сигналу з достатнім співвідношенням сигнал/шум процедури нормалізації не є обов'язковими для отримання формантних даних. Більше того, нормалізація призводить до менш точного виміру ширини спектрів формантних частот. Неприпустимим є використання тривалості оброблюваного кадру менше 40 мс. Зазначені результати дозволяють модифікувати традиційну методику попередньої обробки голосового сигналу. Використання методу моделювання щодо експериментального голосового сигналу підтверджує достовірність отриманих результатів.

Висновки. Проведені експериментальні дослідження показують доцільність виключення процедур нормалізації при високому співвідношенні сигнал/шум голосового сигналу, що має місце в системах автентифікації користувачів. Такий підхід дозволить підвищити оперативність отримання формантних даних і більш точно оцінювати ширину спектрів формантних частот. Результати експериментального дослідження тривалості оброблюваного кадру голосового сигналу свідчать, що його тривалість не має бути менше 40 мс. В іншому випадку оцінки формантних частот будуть завищеними. Більш того, при обробці фонем можна голосовий сигнал, що обробляється, не розбивати на фрейми. Практичне застосування результатів досліджень дозволяє підвищити оперативність та точність формування формантних даних. Перспективами подальших досліджень може бути дослідження впливу процедур нормалізації та фреймінгу на інші елементи шаблону користувача системи автентифікації.

КЛЮЧОВІ СЛОВА: автентифікація, голосовий сигнал, нормалізація, сегментація, формантні дані.

ЛІТЕРАТУРА

1. До питання оцінки ефективності біометричних систем / [М. О. Пастушенко, М. С. Пастушенко, М. О. Петраченко] // Проблеми телекомунікацій. – 2023. – № 1(32). – С. 37–44. DOI: <https://doi.org/10.30837/pt.2023.1.03>
2. Rabiner L. R. Digital Processing of Speech Signals / L. R. Rabiner, R. W. Schafer. – NJ : Prentice-Hall, Inc., 1978. – 512 p. URL: https://ie.u-kyu.ac.jp/~asharif/pukiwiki/attach/Acoustic%20Speech%20Signal%20Processing_Prentice%20Hall%20-%20Digital%20Processing%20of%20Speech%20Signals.pdf
3. Beigi H. Fundamentals of Speaker Recognition / H. Beigi. – NY : Springer, 2011. – 942 p. DOI:10.1007/978-0-387-77592-0
4. Persson A. Comparing normalization against US English listeners'vowel perception / A. Persson, S. Barreda, T. F. Jaeger // The Journal of the Acoustical Society of America. – 2025. – Vol. 157, № 2. – P. 1458–1482. DOI: <https://doi.org/10.1121/10.0035476>
5. Clopper C. G. Raw acoustic vs. normalized phonetic convergence: Imitation of the Northern Cities Shift in the American Midwest / C. G. Clopper, E. Dossey, R. Gonzalez // Laboratory Phonology. – 2024. – Vol. 15(1). – P. 1–15. DOI: <https://doi.org/10.16995/labphon.10893>
6. Anikin A. A Practical guide to calculating vocal tract length and scale-invariant formant patterns / A. Anikin, S. Barreda, D. Reby // Springer Nature Link. – 2023. – Vol. 56. – P. 5588–5604. DOI: 10.3758/s13428-023-02288-x
7. Almaadeed N. Text-Independent Speaker Identification Using Vowel Formants / N. Almaadeed, A. Aggoun, A. Amira // Journal of Signal Processing Systems. – 2015. – Vol. 82, № 3. – P. 345 – 356. DOI: <https://doi.org/10.1007/s11265-015-1005-5>
8. Audio Segmentation Techniques and Applications Based on Deep Learning / [S. Aggarwal, G. Vasukidevi, S. Selvakamani et al.] // Journal of Scientific Programming. Wiley Online Library. – 2022. – Vol. 2022. – P. 1–9. DOI: <https://doi.org/10.1155/2022/7994191>
9. Explainable by-design Audio Segmentation through Non-Negative Matrix Factorization and Probing / [M. Lebourdais, T. Mariotte, A. Almudévar, M. et al.] // arXiv:2406.13385 [eess.AS]. – 2024. P. 1–5. DOI: <https://doi.org/10.48550/arXiv.2406.13385>
10. Lightweight Audio Segmentation for Long-form Speech Translation / [J. Lee, S. Kim, H. Kim, J. S. Chung] // arXiv:2406.10549 [eess.AS]. – 2024. – P. 1–5. DOI: <https://doi.org/10.48550/arXiv.2406.10549>
11. Pastushenko M. Investigation of Informativeness and Stability of Mel-Frequency Cepstral Coefficients Estimates based on Voice Signal Phase Data of Authentication System User / M. Pastushenko, Ya. Krasnozheniuk, M. Zaika // International Conference Problems of Infocommunications. Science and Technology 6–9 October 2020 (PIC S&T'2020). – Kharkiv, Ukraine, 2020. – P. 467–472. DOI: 10.1109/PICST51311.2020.9468083

AN ANALYTICAL APPROACH TO MULTI-CRITERIA CHOOSING TECHNOLOGICAL SCHEME FOR INFORMATION PROCESSING

Popov M. – Dr. Sc., Professor, Corresponding Member of the NAS of Ukraine, Director, State Institution “Scientific Centre for Aerospace Research of the Earth of the Institute of Geological Sciences of the National Academy of Sciences of Ukraine”, Kyiv, Ukraine.

Zaitsev O. – Dr. Sc., Associate Professor, Head of Department, Yevhenii Bereznyak Military Academy, Kyiv, Ukraine.

Stefantsev S. – PhD, Senior Lecturer of Department, Yevhenii Bereznyak Military Academy, Kyiv, Ukraine.

ABSTRACT

Context. Today, effective information processing is critically important for making strategic, tactical, and operational management decisions. The increasing volumes of information and the need for its rapid analysis necessitate the development and implementation of new methods for multi-criteria choosing technological schemes for information processing. The application of analytical approaches, such as the Ordered Weighted Averaging (OWA) method, allows for the improvement of the quality of final information products, which is relevant for analysis and research in various fields.

Objective. The aim of the research is to develop an analytical approach to multi-criteria choosing an information processing technological scheme using the OWA operator.

Method. The paper uses an analytical approach based on the multi-criteria decision-making method. Specifically, the Ordered Weighted Averaging (OWA) operator is applied, which allows taking into account the weight coefficients of the criteria and their ranking significance to determine the optimal information processing technological scheme.

Results. The research results show that the application of the OWA operator effectively aggregates the evaluation of alternatives and selects the technological scheme that best meets the specified criteria for the quality of the end information product. The conducted experiments confirmed the effectiveness of the proposed approach in evaluating alternative information processing schemes.

Conclusions. The proposed approach to multi-criteria selection of a technological scheme for processing intelligence data allows for improved quality of the end information product and considers the importance of various criteria. Further research could be focused on the development of automated decision support systems taking into account the metadata of intelligence data.

KEYWORDS: information processing, Technological scheme, End information product, Multi-criteria analysis, OWA operator.

ABBREVIATIONS

OWA – Ordered Weighted Aggregation;

EIP – End Information Product;

TS – Technological Scheme.

NOMENCLATURE

$A = \{a_1, \dots, a_n, \dots, a_N\}$ is a set of numbers to be aggregated;

b_j is a j -th element of a set A after ordering its constituent elements $\{a_1, \dots, a_n, \dots, a_N\}$ according to the principle of non-increasing (decreasing) their values;

$b_{m,j}$ is a j -th element of an ordered sequence levels of the criteria satisfaction for the TS_m scheme case;

$C = (c_1, \dots, c_n, \dots, c_N)$ is a set of criteria for assessing the End Information Product quality;

c_n is n -th partial criterion of End Information Product quality;

D_m is a generalized criterial level of the End Information Product quality achieved using TS_m scheme;

$F(A)$ is a resultant of the aggregation of the numbers $\{a_1, \dots, a_n, \dots, a_N\}$ by means of the OWA operator;

$l_{m,n}$ is satisfaction level of the n -th partial criterion by TS_m scheme;

M is a total number of alternative technological schemes;

N is a total number of quality assessment criteria for the EIP;

TS_m is m -th technological scheme for information processing;

$u_{m,j}$ is an importance of the criterion associated with $b_{m,j}$;

V_j is an importance of the j -th criterion;

$W = (w_1, \dots, w_n, \dots, w_N)$ is a set of weight coefficients (weights).

INTRODUCTION

As is known, one of the most important factors determining the successful functioning of virtually any organizational system is the level of its information support [1]. At the same time, taking into account the high dynamism of the modern world, special importance is paid to updating previously obtained data and replenishing it. However, the permanent increase of information volumes complicates carrying out in-depth and comprehensive analysis its and requires the use of adequate technological methods and schemes.

For conducting comprehensive, including quantitative, analysis of information, there are various TS that differ in their data analysis methods, software and instrumental

support, level of automation, quality of created EIP, and so on [2].

Content and presentation form of the information analysis results are usually determined by the customer. In particular, as such a result can be:

- information decision;
- subjective evaluate of the object or the situation of interest state;
- in-brief report;
- thematic map, etc.

Quality of any EIP may be assessed using such criteria as relevance, reliability, completeness, timeliness, but not only [3]. However, it is namely the listed criteria are usually used when choosing a TS.

In our work, the choosing the relevant scheme TS for information processing is considered as a multi-criteria task, and to solve it an analytical approach based on the application of the OWA operator, introduced by R. Yager in [4], is proposed.

The unique properties of the OWA operator include the ability to combine (aggregate) criterion assessments taking into account their position in the rank statistics, as well as, if necessary, to help find a compromise between the assessments of alternatives across different criteria [5, 6].

The object of the study is a procedure for choosing technological scheme for information processing.

The subject of the study is a methodological approach to the multi-criteria choosing the technological scheme TS leveraging the OWA operator.

The aim of the study is to develop a methodological approach to choosing the relevant scheme TS for information processing based on the OWA operator.

1 PROBLEM STATEMENT

We consider a situation where an analyst has obtained specific pieces of information $\{Info\}$ on the basis of which he must create the next EIP. It is assumed that to process the obtained information and to form the EIP, the analyst can use one of the M available TS $\{TS_m | m = 1, 2, \dots, M\}$ with different capabilities.

The task for an analyst is to choose, among the available $\{TS\}$, the scheme that allows to form the EIP whose quality best meets the totality of specified criteria.

There are two or more such criteria, and they may differ in importance.

Formally, the task can be written as follows: to choose from $TS | TS \in \{TS_m | m = 1, 2, \dots, M\}$ one scheme TS^* which will provide the transformation $\{Info\}$ in the product $KIII^*$ according to the rule (1):

$$\{Info\} \xrightarrow{TS^*} KIII^*, \quad (1)$$

at the same time the product $KIII^*$ will satisfy the follow criterion of quality:

$$D^* = \max\{D_m | m = 1, 2, \dots, M\}. \quad (2)$$

An analytical approach proposed and described below allows to solve this task.

2 REVIEW OF LITERATURE

Issues of applying multi-criteria decision analysis which assists decision-making in various fields, are actively studied (see, e.g., [8–14]).

In [8], a review of approaches to multi-criteria decision analysis in the field of sustainable energy is presented. The authors examine in detail the specific features characteristic of the various stages of the decision-making process, namely: the stages of criteria selection, criteria weighting, evaluation, and final aggregation.

The experience of practical application of multi-criteria selection and decision-making methods in ecology has been analyzed by researchers from the Massachusetts Institute of Technology (USA) in [9]. The presented examples demonstrate the high effectiveness of such methods in solving various environmental problems.

Reference [10] discusses the use of a multi-criteria approach to support strategic decision-making and provides a number of proposals for adapting the approach based on the specific application area.

Reference [11] presents the results of research on the comparative evaluation of the effectiveness of modern multi-criteria methods (TOPSIS, VIKOR, COPRAS, MULTIMOORA, and PROMETHEE GAIA) in the task of selecting renewable energy sources (hydropower plants, solar photovoltaic systems) for the energy sector of Latvia. It is concluded that the choice of the most effective multi-criteria method depends on the type of energy source to which it is applied.

Based on a comprehensive content analysis of approximately 3,000 scientific articles, the authors of [12] identified a growing trend in the application of multi-criteria methods (AHP/ANP, TOPSIS, MAUT/MAVT, etc.) in the field of the environment (including water, air, energy, natural resources, and waste management), including in geospatial analysis tasks.

Researchers from the Swiss Federal Institute of Aquatic Science and Technology [13] analyzed over 300 literature sources (have been published between 2000 and 2015) that explore the potential for combining multi-criteria decision analysis methods. The study examined how extensively problem-structuring methods and multi-criteria decision analysis are used together, how they are integrated, and what advantages and challenges such an approach presents.

The study by researchers from the O. Beketov Kharkiv National University of Urban Economy [14] is devoted to solving a multi-criteria assignment problem under a categorical efficiency criterion.

At the same time, the issue of choosing and applying the most effective methods and schemes for multi-criteria analysis of information remains open.

3 MATERIALS AND METHODS

As mentioned above, the proposed approach is based on the use of the OWA operator. Its essence is as follows. Let there be a set of N numbers $\{a_1, \dots, a_n, \dots, a_N\}$, and the task is to combine (aggregate) them into a point-valued number $F(A)$. Using the OWA operator, this number is calculated according to the following formula [4]:

$$F(A) = \sum_{j=1}^N w_j \cdot b_j = w_1 b_1 + \dots + w_n b_n + \dots + w_N b_N. \quad (3)$$

Let us now turn to the essence of the proposed approach. It is assumed that there is a set of criteria C that the EIP must meet, with each criterion having its importance V_j ($j=1, 2, \dots, N$). It is further assumed that there are M different data analysis schemes $\{TS_1, \dots, TS_m, \dots, TS_M\}$, where each the m -th scheme ($m=1, \dots, M$) provides its level $\{l_{m,1}, \dots, l_{m,n}, \dots, l_{m,N}\}$ of satisfaction for the corresponding criterion c_n during data processing.

It is assumed that the levels of satisfaction of criteria c by any specific data analysis scheme TS_m are known. These levels can be obtained based on the application history of the given TS, through expert evaluations, or by statistical modeling methods.

Each element c_n from the C set can be considered a partial criterion ($n=1, 2, \dots, N$). The level of satisfaction of partial criteria is measured in relative units. For example, the partial criterion c_n is defined as the maximum time resource allocated for the production of the EIP, which should not exceed two hours.

Let us assume that when applying the m -th data analysis scheme TS , this criterion is fully satisfied at 100%, which means $l_{m,n} = 1.0$. At the same time, the $(m+1)$ -th TS requires 2.5 hours to complete the same task, exceeding the allocated resource by 20%; this means that $l_{m,n} = 0.8$.

According to the proposed approach, all partial criteria are combined (aggregated) into a generalized criterion. The task then is to analytically and justifiably select, among the M available $\{TS\}$ the scheme that will ensure the formation of the EIP with the biggest value of the generalized criterion.

Let us denote by D_m the generalized criterion level corresponding to the EIP formed using the m -th TS during the processing of the given information. Applying the OWA operator, the generalized criterion level is calculated using formula (4):

$$D_m = \sum_{j=1}^N w_{m,j} \cdot b_{m,j} = w_{m,1} b_{m,1} + \dots + w_{m,n} b_{m,n} + \dots + w_{m,N} b_{m,N}, \quad (4)$$

$$(m=1, 2, \dots, M).$$

To perform calculations according to formula (4), certain logical and computational procedures need to be carried out.

First, from the known elements of the sequence L_m , the sequence (5) is constructed:

$$b_{m,j} = \{b_{m,1}, \dots, b_{m,n}, \dots, b_{m,N}\}, \quad (5)$$

whose elements are related to each other as $b_{m,1} \geq b_{m,2} \geq \dots \geq b_{m,N}$, at what:

$$b_{m,1} = \max\{l_{m,1}, \dots, l_{m,n}, \dots, l_{m,N}\};$$

$$b_{m,N} = \min\{l_{m,1}, \dots, l_{m,n}, \dots, l_{m,N}\}.$$

Second, a correspondence table is compiled:

$w_{m,j}$	$b_{m,j}$	$u_{m,j}$
$w_{m,1}$	$b_{m,1}$	$u_{m,1}$
$\vdots$	$\vdots$	$\vdots$
$w_{m,n}$	$b_{m,n}$	$u_{m,n}$
$\vdots$	$\vdots$	$\vdots$
$w_{m,N}$	$b_{m,N}$	$u_{m,N}$

which reflects the fact that the weight values $w_{m,j}$ are functions of two arguments, namely $b_{m,j}$ and $u_{m,j}$. Note that the values of the argument $b_{m,j}$ are defined by sequence (3), and thus are known.

Regarding the elements of column $u_{m,j}$: according to the construction rule of sequence $b_{m,j}$, each of its elements $b_{m,n}$ is associated with a corresponding partial criterion from the set $\{c_1, \dots, c_n, \dots, c_N\}$. The column $u_{m,j}$ consists of the importance values V_j assigned to the partial criteria associated with the respective elements of sequence $b_{m,j}$.

Third, the construction of sequence $u_{m,j} = \{u_{m,1}, \dots, u_{m,1}, \dots, u_{m,1}\}$ allows proceeding to the calculation of weights $w_{m,j} = \{w_{m,1}, \dots, w_{m,1}, \dots, w_{m,1}\}$, for which formula (6) is proposed in [7]:

$$w_{m,j} = \left(\frac{\sum_{k=1}^j u_{m,k}}{\sum_{k=1}^N u_{m,k}} \right)^2 - \left(\frac{\sum_{k=1}^{j-1} u_{m,k}}{\sum_{k=1}^N u_{m,k}} \right)^2, \quad (6)$$

($m = 1, 2, \dots, M$).

Having the results of the logical and computational procedures, the criterion level ($D_m | m = 1, 2, \dots, M$) that can be achieved by the EIP formed using each available TS is determined according to formula (4).

Finally, comparing the calculated criterion levels from D_1 to D_M provides the analyst with the opportunity to select the relevant and most effective scheme TS for information processing according to criterion c_n .

4 EXPERIMENTS

Assume the EIP quality is evaluated by three criteria: reliability (c_1), completeness (c_2) and timeliness (c_3). The importance of these criteria is given respectively as: $V_1 = 1.0$; $V_2 = 0.60$; $V_3 = 0.80$.

When planning the data analysis process, the analyst has two alternative technological schemes – TS_1 and TS_2 . Using one of the methods mentioned above, it was established that the TS_1 scheme can generate the EIP that satisfies the accepted criteria to the following extent: criterion (c_1) – 100%, criterion (c_2) – 70%, and criterion (c_3) – 80%. The TS_2 scheme allows for the formation of an EIP that satisfies the accepted criteria as follows: criterion (c_1) – 90%, criterion (c_2) – 80%, and criterion (c_3) – 100%.

The question arises: which TS should be preferred? According to the proposed approach, preference is given to the one of the two that can generate the EIP with a higher value of the generalized criterion.

Let us perform the necessary calculations for the schemes TS_1 and TS_2 , and then compare their results.

5 RESULTS

For the scheme TS_1 , the correspondence table will be as follows:

c	$w_{1,j}$	$b_{1,j}$	$u_{1,j}$
c_1	$w_{1,1}$	1.0	1.0
c_2	$w_{1,2}$	0.8	0.8
c_3	$w_{1,3}$	0.7	0.6

$$w_{1,1} = \left(\frac{u_{1,1}}{\sum_{k=1}^3 u_{1,k}} \right)^2 = \left(\frac{1.0}{1.0 + 0.8 + 0.6} \right)^2 = 0.174;$$

$$w_{1,2} = \left(\frac{\sum_{k=1}^2 u_{1,k}}{\sum_{k=1}^3 u_{1,k}} \right)^2 - \left(\frac{u_{1,1}}{\sum_{k=1}^3 u_{1,k}} \right)^2 = \left(\frac{1.0 + 0.8}{1.0 + 0.8 + 0.6} \right)^2 - 0.174 = 0.388;$$

$$w_{1,3} = \left(\frac{\sum_{k=1}^3 u_{1,k}}{\sum_{k=1}^3 u_{1,k}} \right)^2 - \left(\frac{\sum_{k=1}^2 u_{1,k}}{\sum_{k=1}^3 u_{1,k}} \right)^2 = 1 - \left(\frac{1.0 + 0.8}{1.0 + 0.8 + 0.6} \right)^2 - 0.5625 = 0.438.$$

Substituting the calculated weight values, as well as the data from the correspondence table into formula (4), we obtain:

$$D_1 = \sum_{j=1}^3 w_{1,j} \cdot b_{1,j} = 0.174 \cdot 1.0 + 0.388 \cdot 0.8 + 0.438 \cdot 0.7 = 0.791.$$

For the TS_2 scheme, the correspondence table will be as follows:

c	$w_{2,j}$	$b_{2,j}$	$u_{2,j}$
c_3	$w_{2,1}$	1.0	0.8
c_1	$w_{2,2}$	0.9	1.0
c_2	$w_{2,3}$	0.8	0.6

The weights are calculated using formula (6):

$$w_{2,1} = \left(\frac{u_{2,1}}{\sum_{k=1}^3 u_{2,k}} \right)^2 = \left(\frac{0.8}{1.0 + 0.8 + 0.6} \right)^2 = 0.111;$$

$$w_{2,2} = \left(\frac{\sum_{k=1}^2 u_{2,k}}{\sum_{k=1}^3 u_{2,k}} \right)^2 - \left(\frac{u_{2,1}}{\sum_{k=1}^3 u_{2,k}} \right)^2 = \left(\frac{0.8 + 1.0}{1.0 + 0.8 + 0.6} \right)^2 - \left(\frac{0.8}{1.0 + 0.8 + 0.6} \right)^2 = 0.451;$$

$$w_{2,3} = \frac{\left(\sum_{k=1}^3 u_{2,k} \right)^2}{\sum_{k=1}^3 u_{2,k}} - \frac{\left(\sum_{k=1}^2 u_{2,k} \right)^2}{\sum_{k=1}^2 u_{2,k}} =$$

$$= 1 - \left(\frac{0.8 + 1.0}{1.0 + 0.8 + 0.6} \right)^2 = 0.438.$$

Substituting the calculated weight values and the data from the correspondence table into formula (4), we obtain:

$$D_2 = \sum_{j=1}^3 w_{2,j} \cdot b_{2,j} =$$

$$= 0.111 \cdot 1.0 + 0.451 \cdot 0.9 + 0.438 \cdot 0.8 = 0.867.$$

A comparison of the criterion values D_1 and D_2 shows that, under the given conditions, the TS_2 scheme is capable of providing higher the EIP quality.

6 DISCUSSION

The proposed approach to choose relevant TS for information processing based on the OWA operator demonstrates advantages compared to traditional methods, particularly due to its ability to account for the ranked importance of partial criteria. This enables a more flexible and adaptive decision-making process, especially in cases where different criteria have varying weights and importance.

The results of the numerical example indicate that even minor differences in the values of partial criteria can significantly influence the choice of TS when using the OWA operator.

A comparison with similar studies [8–14] shows that most existing multi-criteria decision analysis methods rely on rigid ranking rules and do not always consider the relative importance of criteria. In this context, the proposed approach offers greater flexibility, especially when integrating expert assessments or fuzzy data.

The application of the OWA operator requires the prior determination of weighting coefficients, which may necessitate involving experienced experts or additional normalization procedures. This aspect defines the dependence of result accuracy on the quality of input data and the justification of the weights.

Further improvement of the approach may involve the use of machine learning methods for the automatic determination of weights based on historical data or predictive models.

Practical significance of the obtained results: the analytical approach can be effectively used in decision support systems under conditions of limited time or incomplete information.

The proposed approach is not only theoretically justified but also demonstrates high effectiveness in the numerical example. This opens prospects for further re-

search in the direction of automating TS evaluation processes, integrating with intelligent agents, and developing dynamic models for adapting criteria to changing environments.

CONCLUSIONS

The paper proposes the analytical approach to multi-criteria choosing TS for information processing based on the OWA operator leveraging. The content of the approach is described in detail, and the numerical example of its application is provided.

The scientific novelty of the study is that the proposed approach to multi-criteria choosing TS for information processing makes it possible to adapt the weighting coefficients of criteria and their ranking structure considering the properties of available TS, thereby helping to find a compromise between the evaluations of alternatives across different criteria.

The practical significance lies in the possibility of applying the proposed approach in decision support systems for information processing, which will improve the quality of the EIP. The methodology can be used in information-analytical structures engaged in the evaluation of strategic information, as well as in military and intelligence units.

For future research, it is important to expand the proposed methodology by integrating machine learning and neural network approaches to automate the process of choosing the relevant TS. Additionally, a promising direction is the development of algorithms for predicting the quality of the EIP based on the analysis of metadata of the input information.

REFERENCES

1. Valacich J. Information Systems Today: Managing in the Digital World. 9th ed. UK, Pearson Education Ltd, 2021, 564 p.
2. Waltz E. Quantitative Intelligence Analysis: Applied Analytic Models, Simulations, and Games. Lanham: Rowman & Littlefield Publishers, 2014, 282 p.
3. Prunckun H. Handbook of Scientific Methods of Inquiry for Intelligence Analysis. Lanham, Scarecrow Press, 2010, 248 p.
4. Yager R. R. On Ordered Weighted Averaging Aggregation Operators in Multicriteria Decision Making, *IEEE Transactions on Systems, Man and Cybernetics*, 1988, Vol. 18, pp. 183–190. DOI: 10.1109/21.87068.
5. Yager R. R. Families of OWA Operators, *Fuzzy Sets and Systems*, 1993, Vol. 59, Issue 2, pp. 125–148. DOI: 10.1016/0165-0114(93)90194-M.
6. Grabisch M., Marichal J. L., Mesiar R. et al. Aggregation Functions: Means, *Information Science*, 2011, Vol. 181, Issue 1, pp. 1–22. DOI: 10.1016/j.ins.2010.08.043.
7. Yager R. R. Quantifier Guided Aggregation Using OWA Operators, *International Journal of Intelligent Systems*, 1996, Vol. 11, Issue 1, pp. 49–73. DOI: 10.1002/(SICI)1098-111X(199601)11:1<49::AID-INT3>3.0.CO;2-Z.

8. Wang J. J., Jing Y. Y., Zhang C. F. et al. Review on Multi-Criteria Decision Analysis Aid in Sustainable Energy Decision-Making / [// *Renewable and Sustainable Energy Reviews*, 2009, Vol. 13, Issue 9, pp. 2263–2278. DOI: 10.1016/j.rser.2009.06.021.
9. Huang I. B., Keisler J., Linkov I. Multi-Criteria Decision Analysis in Environmental Sciences: Ten Years of Applications and Trends, *Science of The Total Environment*, 2011, Vol. 409, Issue 19, pp. 3578–3594. DOI: 10.1016/j.scitotenv.2011.06.022.
10. Montibeller G., Franco A. eds. C. Zopounidis, P. Pardalos Multi-Criteria Decision Analysis for Strategic Decision Making, *Handbook of Multicriteria Analysis*. Heidelberg, Springer, 2010, (Applied Optimization; Vol. 103), pp. 25–48. DOI: 10.1007/978-3-540-92828-7_2.
11. Zlaugotne B., Zihare L., Balode L. et al. Multi-Criteria Decision Analysis Methods Comparison, *Environmental and Climate Technologies*, 2020, Vol. 24, Issue 1, pp. 454–471. DOI: 10.2478/rtuct-2020-0028.
12. Cegan J., Filion A., Keisler J. et al. Trends and Applications of Multi-Criteria Decision Analysis in Environmental Sciences: Literature Review, *Environmental Systems and Decision*, 2017, Vol. 37, pp. 123–133. DOI: 10.1007/s10669-017-9642-9.
13. Marttunen M., Lienert J., Belton V. Structuring Problems for Multi-Criteria Decision Analysis in Practice: A Literature Review of Method Combinations, *European Journal of Operational Research*, 2017, Vol. 263, Issue 1, pp. 1–17. DOI: 10.1016/j.ejor.2017.04.041.
14. Novozhylova M. V., Karpenko M. Y. Solution of a Multicriteria Assignment Problem Using a Categorical Efficiency Criterion, *Radio Electronics, Computer Science, Control*. – 2024, No. 4, pp. 75–84. DOI: 10.15588/1607-3274-2024-4-7.

Received 10.06.2025.

Accepted 18.09.2025.

УДК 519.816

АНАЛІТИЧНИЙ ПІДХІД ДО БАГАТОКРИТЕРІАЛЬНОГО ВИБОРУ ТЕХНОЛОГІЧНОЇ СХЕМИ ОПРАЦЮВАННЯ ІНФОРМАЦІЇ

Попов М. О. – д-р техн. наук, професор, член-кореспондент НАН України, директор, Державна установа «Науковий центр аерокосмічних досліджень Землі Інституту геологічних наук Національної академії наук України», Київ, Україна.

Зайцев О. В. – д-р техн. наук, доцент, начальник кафедри, Воєнна академія імені Євгенія Березняка, Київ, Україна.

Стефанцев С. С. – канд. техн. наук, старший викладач кафедри, Воєнна академія імені Євгенія Березняка, Київ, Україна.

АНОТАЦІЯ

Актуальність. Сьогодні ефективне опрацювання інформації є критично важливим для прийняття стратегічних, тактичних та оперативних управлінських рішень. Зростаючий обсяг інформації та необхідність її швидкого аналізу вимагають розробки та впровадження нових методів багатокритеріального вибору технологічних схем обробки інформації. Застосування аналітичних підходів, таких, як метод упорядкованого зваженого агрегування (OWA), дозволяє покращити якість кінцевих інформаційних продуктів, що є актуальним для аналізу та досліджень у різних сферах людської діяльності.

Мета. Метою дослідження є розробка аналітичного підходу до багатокритеріального вибору технологічної схеми опрацювання інформації із застосуванням оператора упорядкованого зваженого агрегування (OWA-оператора).

Метод. В роботі використано аналітичний підхід, заснований на методі багатокритеріального прийняття рішень. Зокрема, застосовується OWA-оператор, що дозволяє враховувати вагові коефіцієнти критеріїв та їх рангову значущість для визначення найбільш ефективної технологічної схеми опрацювання інформації.

Результати. Результати досліджень показують, що застосування OWA-оператора дозволяє ефективно агрегувати оцінки альтернатив та обрати технологічну схему, яка найбільше серед тих, що є в наявності, відповідає заданим критеріям якості кінцевого інформаційного продукту. Проведені експерименти підтвердили ефективність запропонованого підходу при оцінюванні альтернативних схем опрацювання інформації.

Висновки. Запропонований підхід до багатокритеріального вибору технологічної схеми опрацювання розв'язувальних даних дозволяє підвищити якість кінцевого інформаційного продукту та враховувати при цьому важливість розглядуваних критеріїв. Подальші дослідження плануються спрямувати на розробку автоматизованої системи підтримки прийняття рішень з урахуванням результатів аналізу метаданих вхідних даних.

КЛЮЧОВІ СЛОВА: опрацювання інформації, технологічна схема, кінцевий інформаційний продукт, багатокритеріальний аналіз, OWA-оператор.

ЛІТЕРАТУРА

1. Valacich J. Information Systems Today: Managing in the Digital World. 9th ed. / J. Valacich. – UK : Pearson Education Ltd, 2021. – 564 p.
2. Waltz E. Quantitative Intelligence Analysis: Applied Analytic Models, Simulations, and Games / E. Waltz. – Lanham : Rowman & Littlefield Publishers, 2014. – 282 p.
3. Prunckun H. Handbook of Scientific Methods of Inquiry for Intelligence Analysis / H. Prunckun. – Lanham : Scarecrow Press, 2010. – 248 p.
4. Yager R. R. On Ordered Weighted Averaging Aggregation Operators in Multicriteria Decision Making / R. R. Yager // IEEE Transactions on Systems, Man and Cybernetics. – 1988. – Vol. 18. – P. 183–190. DOI: 10.1109/21.87068.
5. Yager R. R. Families of OWA Operators / R. R. Yager // Fuzzy Sets and Systems. – 1993. – Vol. 59, Issue 2. – P. 125–148. DOI: 10.1016/0165-0114(93)90194-M.
6. Aggregation Functions: Means / [M. Grabisch, J. L. Marichal, R. Mesiar et al.] // Information Science. – 2011. – Vol. 181, Issue 1. – P. 1–22. DOI: 10.1016/j.ins.2010.08.043.
7. Yager R. R. Quantifier Guided Aggregation Using OWA Operators / R. R. Yager // International Journal of Intelligent Systems. – 1996. – Vol. 11, Issue 1. – P. 49–73. DOI: 10.1002/(SICI)1098-111X(199601)11:1<49::AID-INT3>3.0.CO;2-Z.
8. Review on Multi-Criteria Decision Analysis Aid in Sustainable Energy Decision-Making / [J. J. Wang, Y. Y. Jing, C. F. Zhang et al.] // Renewable and Sustainable Energy Reviews. – 2009. – Vol. 13, Issue 9. – P. 2263–2278. DOI: 10.1016/j.rser.2009.06.021.
9. Huang I. B. Multi-Criteria Decision Analysis in Environmental Sciences: Ten Years of Applications and Trends / I. B. Huang, J. Keisler, I. Linkov // Science of The Total Environment. – 2011. – Vol. 409, Issue 19. – P. 3578–3594. DOI: 10.1016/j.scitotenv.2011.06.022.
10. Montibeller G., Franco A. Multi-Criteria Decision Analysis for Strategic Decision Making / G. Montibeller, A. Franco // Handbook of Multicriteria Analysis / eds. C. Zopounidis, P. Pardalos. – Heidelberg: Springer, 2010. – (Applied Optimization; Vol. 103). – P. 25–48. DOI: 10.1007/978-3-540-92828-7_2.
11. Multi-Criteria Decision Analysis Methods Comparison / [B. Zlaugotne, L. Zihare, L. Balode et al.] // Environmental and Climate Technologies. – 2020. – Vol. 24, Issue 1. – P. 454–471. DOI: 10.2478/rtuct-2020-0028.
12. Trends and Applications of Multi-Criteria Decision Analysis in Environmental Sciences: Literature Review / [J. Cegan, A. Fillion, J. Keisler et al.] // Environmental Systems and Decision. – 2017. – Vol. 37. – P. 123–133. DOI: 10.1007/s10669-017-9642-9.
13. Marttunen M., Lienert J., Belton V. Structuring Problems for Multi-Criteria Decision Analysis in Practice: A Literature Review of Method Combinations / M. Marttunen, J. Lienert, V. Belton // European Journal of Operational Research. – 2017. – Vol. 263, Issue 1. – P. 1–17. DOI: 10.1016/j.ejor.2017.04.041.
14. Novozhylova M. V., Karpenko M. Y. Solution of a Multicriteria Assignment Problem Using a Categorical Efficiency Criterion / M. V. Novozhylova, M. Y. Karpenko // Radio Electronics, Computer Science, Control. – 2024. – Vol. 4. – P. 75–84. DOI: 10.15588/1607-3274-2024-4-7.

MODIFY SHADERS AND RENDER TEXTURES ON CURVED SURFACES

Suhoniak I. – Associate Professor, PhD in Technical Sciences, Associate Professor of the Department of Computer Science, State University “Zhytomyr Polytechnic”, Zhytomyr, Ukraine.

Marchuk G. – Senior Lecturer of the Department of Computer Science, State University “Zhytomyr Polytechnic”, Zhytomyr, Ukraine.

Oleksiuk O. – Applicant of the Department of Computer Science, State University “Zhytomyr Polytechnic”, Zhytomyr, Ukraine.

ABSTRACT

Context. The display of curvilinear surfaces on flat screens is a complex task. The development of an interface for such surfaces is a relevant task that requires the solution of numerous issues. This paper presents an approach to UI development for curvilinear surfaces and shader modifications for the creation of realistic landscape elements. The object of the research is the development of an interface system based on a custom raycaster to ensure interactivity and create an immersive effect within the game environment.

Objective. The purpose of the paper. The primary objective of this research is to create, optimise and adapt shaders on curved surfaces to achieve more efficient rendering with high-quality visualisation.

Method. The development of user interfaces (UI) for curvilinear surfaces requires consideration of geometric parameters. To resolve this issue, a custom component based on BaseRaycaster was developed, enabling the computation of the intersection between the camera ray and the physical surface. To provide correct and efficient interaction with the canvas a custom component based on BaseRaycaster was created. The developed component solves the problem by identifying the ray intersection point from the camera with the canvas surface. The implementation of this component involves an algorithm for detecting the camera's ray intersections with colliders, using a mathematical model to process the detected elements, according for their depth to ensure proper interaction.

Results. This approach facilitates the creation of interfaces on arbitrary static curved surfaces that are applicable in various gaming and interactive scenarios.

Conclusions. The use of splines and modified shaders ensures the placement of text on curvilinear surfaces and the natural arrangement of roads and other landscape elements according to the terrain contours. This approach is important for developing open-world games or games with complex geometry, where the UI on curvilinear surfaces appears natural and integrated into the environment.

KEYWORDS: game development, shaders, spline, rendering, custom raycaster, curvilinear surfaces, user interface, visualization.

ABBREVIATIONS

UI – User Interface.

NOMENCLATURE

C_p is a position of the camera in 3D space;

C_d is a direction of the beam from the camera;

C_s is a parameters of the collider;

P_c is a collider parameters;

P_n is a normal parameters;

S is a set of surfaces in 3D space, where each surface $s \in S$ has a collider C_s ;

T_w is a width of the canvas texture;

T_h is a height of the canvas texture;

UI is a set of UI elements on the texture, where each element $ui \in UI$;

$T_{c,min}$ is a minimum coordinates on the texture;

$T_{c,max}$ is a maximum coordinates on the texture;

T_{px} is a coordinates of the intersection point on the texture of each UI element along the x-axis;

T_{py} is a coordinates of the intersection point on the texture of each UI element along the y-axis;

$z-depth$ is a depth of the element;

P_i is a control points for creating splines;

I_p is a point of intersection of the ray with the surface;

T_c is a coordinates of the intersection point on the texture;

ui is a active UI element;

R_c is a ray from the camera;

I_{px} is a coordinates of the intersection point on the surface along the x-axis;

I_{py} is a coordinates of the intersection point on the surface along the y-axis;

$Lerp$ is a interpolation of textures based on the alpha channel to create smooth transitions between pre-horn segments.

INTRODUCTION

In the modern world of 3D games, user interaction with the virtual environment plays a vital role in ensuring the quality of the user experience. Interfaces are often displayed on the main screen's surfaces, which does not allow maximum immersion in the game. Still, in cases where the interface is located on the curved surface of an object, additional implementation difficulties arise. Texture rendering allows you to create such interfaces, but ensuring accurate and intuitive interaction on curved surfaces is more complex, as standard interaction processing methods are designed for flat surfaces.

The object of research. The study focuses on creating an interface system using a specially developed ray tracing method (raycasting) to increase the player's interactivity and immersion in the game world.

The placement of the interface on a curved surface is an essential aspect of modern game design, especially in terms of maximum immersion in the game world, which allows for a more organic and realistic interaction between the player and the game world. The use of rendered textures and modified shaders opens up new opportunities for integrating UI elements into complex three-dimensional environments, increasing the level of immersion and aesthetic integrity of the game.

The subject of study. Modern computer graphics methods allow you to create virtual worlds with a high degree of realism. However, modeling curved surfaces often found in nature and technology remains challenging. Traditional methods based on polygonal approximation can lead to loss of detail and shape distortion, so there is a need to develop effective methods for modifying shaders and rendering textures that would allow accurate and realistic display of curved surfaces. The subject of study is adapting existing tools and developing new ones to reproduce realistic 3D images on varying complex surfaces.

The study aims to develop and adapt computer graphics methods for creating realistic 3D images on various curved surfaces.

1 PROBLEM STATEMENT

One key approach is to map the interface canvas to a texture, which allows it to be displayed on any surface,

including curves or uneven planes. It is achieved by using a separate camera that displays only the necessary layer with the canvas. It renders it not on the screen but into a texture, which is subsequently used to overlay the object material. A shader then processes this rendered texture to realize the retro style using *ShaderGraph*.

For example, let's consider creating text on the screen of an old tube TV. Since the kinescope in such TVs has a convex shape, several visual effects must be applied to make the image realistic, such as additional geometry curvature (Fig. 1(1)); pixelation effect (Fig. 1(2)); noise and interference (Fig. 1(3)); stripes (Fig. 1(4)), which can be vertical or horizontal; displacement and freeze effect (Fig. 1(5)). To create the desired result, you must develop each effect separately and combine them into a single whole (Fig. 1(6)).

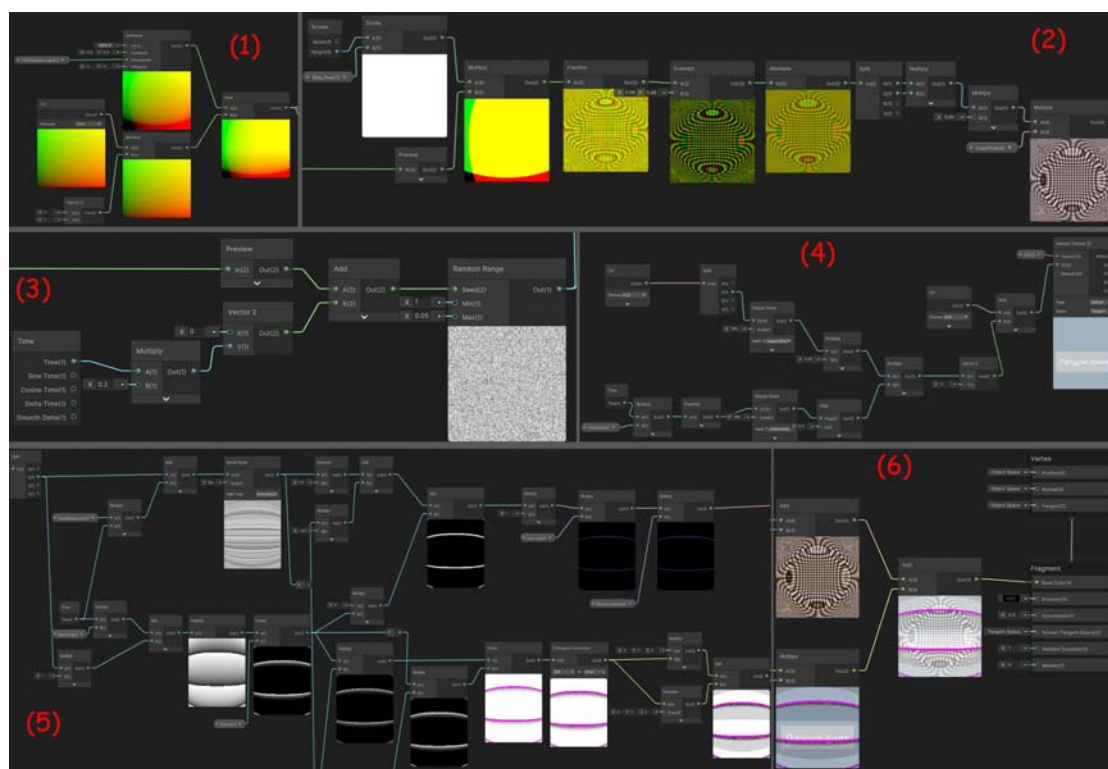


Figure 1 – Step-by-step parameters of the shader: screen curvature (1); pixelation (2); noise (3); stripes on the screen (4); image displacement and freezing (5); holistic picture (6)

Shaders are essential in creating a visual atmosphere and immersing the player in the game. However, the quality of shaders can vary, and they do not always meet high standards. Poor-quality shaders can ruin the atmosphere of the game and reduce the level of immersion. Poorly optimized shaders can lead to a decrease in game performance and freezes. Poor-quality shaders can make game graphics less attractive and detailed. It is also important to note that the textures generated by the shaders also mediate the player's interaction with the game world. Due to the use of a low-quality shader, the text on the

screen of the old monitor is displayed with defects, as seen in Figure 2.



Figure 2 – The result of a low-quality shader

Shader development affects visual perception, as it determines how the player will perceive the game world's elements and how realistically they will be able to immerse themselves in it. However, it is essential to provide visual effects and ensure that the texture on which the shader is applied remains interactive. This means that the player should be able to interact with the objects covered by the texture with the active shader, for example, by clicking on interface elements or performing other actions.

Thus, the shader not only adds an aesthetic component but also considers functionality and interactivity, ensuring the usability of game elements. Developing shaders that combine effects to improve graphics and the possibility of interactive interaction is a complex but essential task to achieve complete immersion in the virtual world.

The algorithm for implementing a custom Raycaster component using splines can be represented as a sequence of stages and key elements that ensure the exact intersection of the beam with the surface, projection onto the texture, and construction of splines for modeling curves (Fig. 3).

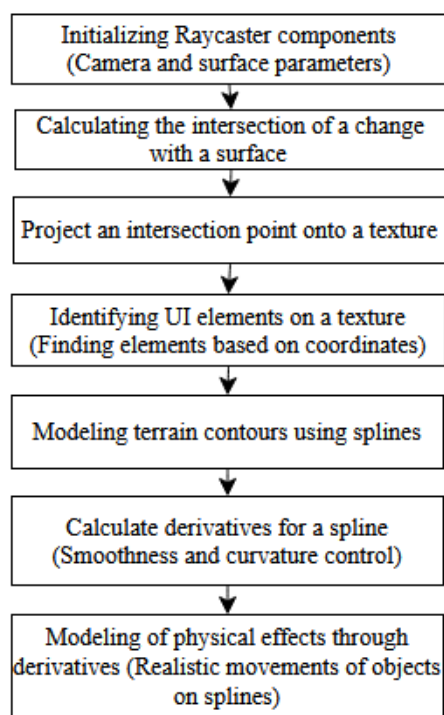


Figure 3 – Flowchart with the definition of the main stages and mathematical components

To initialize the Raycaster components, let the input data be given:

1. Camera parameters (position, beam direction).
2. Physical surface parameters (collider parameters, normal parameters).

The task of creating a Raycaster component for interacting with 3D objects is as follows:

– calculating the point of intersection of the ray with the surface: $I_p = \text{raycast}(C_p, C_d, P_c)$;

- identifying UI elements on the texture: $UIlist = \text{findUI}(T_c)$;
- processing interaction with UI elements: $I = \text{handleInteraction}(UIactive)$;
- determining the intersection of the ray with the surface: $T_c = \text{project}(I_p)$;
- applying depth to interact with UI elements: $I' = \text{apply z-depth}(i)$ – where $i=0, 1, \dots, n$.

2 REVIEW OF THE LITERATURE

In today's video games, shaders are critical to creating visual effects that make games more realistic and immersive. The book Building Quality Shaders for Unity® [1] will help beginners understand what shaders are and how they work. Game engines use different methods to process shaders and achieve the desired result.

Creating game worlds is a complex process requiring developers' considerable effort and resources. Shaders are tiny programs that run on the GPU and are responsible for image processing. The quality of shaders can vary and lead to rendering failures, which can cause objects to disappear, blink, or display incorrectly. They can significantly reduce performance, leading to low frame rates and freezes. The study's authors [2] analyzed 12 shader compilers from Intel, AMD, Nvidia, ARM, and Apple and found more than 16 thousand input data that cause errors affecting the rendering quality. These errors were identified and described, and the developers confirmed their typicality.

Article [3] explores in detail the technical aspects of using textures and shaders in various game engines to achieve various goals. The authors also provide recommendations on how to use these tools effectively.

Article [4] represents a research about combining mesh and shader levels of detail (LOD) using two special algorithms for leads to smoother visual transitions. The mesh LOD effectively reduces the quantity of polygons for distant objects, which hard load the GPU. Shader LOD, on the other hand, allows to reproduce or save small image details (such as texture nuances or complex normals) even on models with a reduced polygon count, thus improving visual quality without significantly increasing the complexity of the geometry

In article [5], the authors propose several improvements (Improved shader and texture detail level using ray cones) to the ray cone algorithm that improve image quality, increase performance, and simplify integration with game engines. The results show that a GPU-based tracer can reduce frame processing time by about 10%. The authors also provide an open-source implementation of this algorithm.

The paper [6] focuses on low-power optimization, which targets texture memory, geometry engine, pixel, and rasterization, as these components contribute significantly to a typical GPU's power consumption. The proposed methodology integrates the dynamic voltage frequency scaling (DVFS) technique, adjusting voltage and frequency based on analyzing the frame rate workload for 3D game scenes.

The authors of [7] present a tool for researching shader design using an interactive evolutionary algorithm integrated with the Unity editor. The framework uses an essential graph-based representation of the latest shader editors and interactive evolution to allow designers to explore multiple visual options starting from an existing shader. The framework encodes a graphical representation of the current shader as a chromosome used to spawn the evolution of a population of shaders. It applies graph-based recombination and mutation with a set of heuristics to generate possible raiders.

In [8], the authors presented a new particle-based fluid surface reconstruction method that includes mesh shaders for the first time. This approach eliminates storing triangular meshes in GPU global memory, significantly reducing the required memory. In addition, our method uses a bidirectional two-level uniform mesh, which speeds up the computationally expensive stage of surface cell detection and effectively solves the problem of vertex overflow among mesh shaders.

The authors of [9] propose using tetrahedra as primitives for volumetric rendering and a pipeline for their rasterization. The work relies on recently introduced mesh shaders to encode each tetrahedron so that the rasterizer calculates the depth of the front and back faces simultaneously when interpolating vertex attributes. However, the fragment shader receives two depths and can calculate its shading in the interval. The presented method is easy to implement and efficient, opening up new possibilities for the pasteurization pipeline.

Paper [10] presents a deep learning-based framework for predicting the shader simplification space, where shader variants can be immediately embedded into a metric space for efficient quality assessment. The new structure allows for a single embedding of the space rather than a single instance. In addition, simplification errors can be interpreted by the mutual attention between shader fragments, presenting an informative focus-aware simplification structure that can help experts optimize codes. The results show that the new framework achieves significant speedups over existing search approaches. The focus-aware simplification framework opens up new possibilities for interpreting shaders for different programs.

The article [11] discusses various ways to use a geometric shader for particle visualization. Geometric shaders are practical and flexible ways to visualize particles. They allow you to create a variety of visual effects with significant performance. This article is helpful for developers interested in visualizing particles using the GPU.

A study [12] used it to fit flash images to a model that can be efficiently evaluated in real-time using a custom shader in the Unity game engine. The custom shader accurately represents the object's mirror properties more accurately than the standard Unity default shaders, with little to no performance impact.

Due to the many options for using modern GPUs, it is usually difficult for engineers to manually check the many shader codes arising from these programs. To this end, the authors of [13] developed a framework that converts

shader codes into graphs and uses sophisticated graph analysis and machine learning techniques in several applications to simplify the analysis of shader graphs efficiently and understandably, aimed at speeding up the entire debugging process and improving the overall performance of the hardware. The authors studied the evolution of shader codes by analyzing time graphs and structure analysis with frequent subgraphs, using them as the primary tools, and performing scene detection and selecting representative frames. The scene (program) was grouped to identify typical scenes and predict inefficient shaders of a new program.

3 MATERIALS AND METHODS

Developing a UI for curved surfaces is a complex task that requires considering the peculiarities of geometry and user interaction.

To provide interaction with the canvas, you should add a custom component based on *BaseRaycaster* (since the standard *GraphicRaycaster* cannot interact with objects in this way for several reasons: it is tightly tied to the main camera and not to the interactive object and does not take information from it), which determines the intersection of the beam from the camera with the physical surface. Given the intersection coordinates, it is necessary to determine the position on the canvas and a list of controls for further processing. The implementation includes searching for beam intersections with colliders using a mathematical formula and further processing the found elements, considering their depth for proper interaction. This approach allows you to create an interface on any curved static surface, which can be used for various gaming and interactive scenarios.

Implementing the custom Raycaster allows you to determine the camera beam's intersection with the object's physical surface. The Raycaster displays the intersection coordinates and is used to further search for UI elements (Fig. 4).

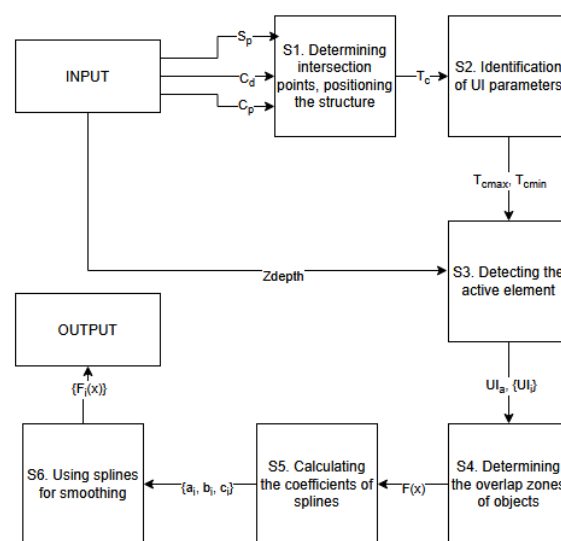


Figure 4 – A formalized spline interpolation model for smoothing planes

List of input variables:

- Camera position in 3D space;
- Camera ray direction;
- Set of surfaces in 3D space, where each surface $s \in S$ has a collider;
- Width and height of the canvas texture.

The UI is a set of UI elements on the texture, where each element $ui \in UI$ has a collider:

- Minimum and maximum coordinates on the texture;
- The depth of the element;
- Set of control points for the construction of splines.

List of output variables:

- Point of intersection of the ray with the surface;
- Coordinates of the intersection point on the texture;
- Active UI element.

Criteria:

- Maximum accuracy of determining the intersection point;
- Accurate projection of T_c onto the texture;
- Correct identification of the active UI element ui , considering z -depth;
- The smoothness and continuity splines;
- The realism of physical effects modeled using derived splines.

Limitations:

- Restrictions on computing resources;
- Restrictions on the time of operations;
- Restrictions on the accuracy of calculations;
- Restrictions on the complexity of splines.

The area where the beam intersects the physical surface is described by equation (1):

$$I_p = R_c \bullet C_s. \quad (1)$$

The point of intersection is projected onto the fabric texture with the following coordinates:

$$T_c = \left(\frac{I_{px}}{T_w}, \frac{I_{py}}{T_h} \right). \quad (2)$$

The resulting coordinates T_c define the position of the texture

After determining the coordinates on the texture, you can find the controls in this area and process them, considering the depth (for proper interaction with objects at different levels). Suppose rectangles or other geometric shapes represent the elements in the UI system. In that case, we can compare the resulting point T_c with the coordinates and dimensions of these elements on the texture.

Let's formalize the parameters of UI elements: $T_{c,min}$ and $T_{c,max}$ are the minimum and maximum coordinates of the UI element on the texture, defining its rectangular area.

If the point $T_c = (T_{cx}, T_{cy})$ is within the UI element, then this element is active:

$$T_{c,min} \leq T_c \leq T_{c,max}. \quad (3)$$

If several UI elements may overlap in the projection, you should add additional criteria for selecting an element, such as depth or layer priority. One of the most common ways to solve this problem is to use z -depth. Z -depth is a value that determines which element is higher than others on the screen. An element with a higher z -depth value appears on top of elements with a lower value:

$$z - depth = element\ depth\ UI. \quad (4)$$

For correct interaction with UI elements in a 3D environment, it is necessary not only to find the elements but also to determine exactly which element the user interacts with, especially when they intersect, using the depth of the element. Taking into account the accuracy of coordinates and perspective ensures correct interaction, and a formalized interface allows you to standardize the processing of input events such as clicks, hovering, and pressing. The formalized definition of the interface for interacting with the UI is determined by formula (2). Identification of a UI element by formula (3), where

$$\begin{aligned} T_{c,min} &= (T_{cx,min}, T_{cy,min}), \\ T_{c,max} &= (T_{cx,max}, T_{cy,max}). \end{aligned} \quad (5)$$

When a user interacts with the interface, it is essential to know how the elements are relative to each other, that is, their depth. The depth of a UI element on a texture determines its distance from the user or other elements on the screen. This distance is denoted by the value z , where $z=0$ means that the element is closest, and larger z values mean that the element is farther away.

A custom Raycaster determines which UI elements should respond to user actions. To accurately determine which UI elements are closest to the user and should be activated, Raycaster calculates the depth (z -depth) of the point where the ray intersects the surface. This allows you to prioritize UI elements correctly, providing an intuitive interaction.

When the ray crosses the surface, you need to determine which UI is at this point. Since UI elements can overlap, depth (z -depth) determines priority, allowing you to select the element closest to the user.

Identifying UI elements using z -depth consists of three steps.

Step 1. Collecting data on the depth of UI elements. Each UI element on the texture has a z -depth, i , determining its distance from the camera in 3D space. The smaller the depth, the closer the element is.

Let's say for n UI elements on a texture, we have sets of coordinates and depths:

$$\{T_{cx}, T_{cy}, z - depth\}, i = 1, 2, \dots, n. \quad (6)$$

I is an element of the texture, which varies from 1 to n .

Step 2: Determining the active UI element. The task is to determine which UI element on the texture will be active for interaction, considering the *z-depth* of the elements. For each intersection point on the texture T_c , the active UI element is determined by the following formula:

$$\text{active UI element} = \arg \min(z - \text{depth}, i). \quad (7)$$

$\arg \min(z - \text{depth}, i)$ is the index of the element for which the depth is *z-depth* and is the minimum among all elements that overlap the intersection point.

Step 3. Check for overlapping UI elements. Since several UI elements on the texture can overlap the same T_c point, selecting the element closer to the user is important. If several elements have the same T_c coordinate, the element is selected using depth minimization.

Each element, i , is checked to see whether its area intersects the point T_c , and then its *depth*, i , is determined for each element.

Of particular interest is the use of modified shaders to create roads that smoothly follow the contours of the terrain. This approach uses splines for road placement and ensures its smooth overlay on the landscape, creating more realistic and detailed scenes. The cubic polynomial is the basis for building splines, which are widely used in computer graphics and modeling to create smooth curves and contours. Splines defined by cubic polynomials provide a smooth transition between points, making them ideal for modeling surfaces and paths in 3D environments. The basic form of a cubic polynomial used to create splines is as follows:

$$F(x) = ax^3 + bx^2cx + d. \quad (8)$$

The coefficients a , b , c , and d determine the shape of the curve.

A cubic spline consists of a set of polynomials (8), where each polynomial corresponds to a separate segment between adjacent points. It is essential to ensure smoothness at the junction points, i.e., that the first and second derivatives coincide. This ensures the smoothness of the curve, which is critical for implementing tasks such as camera trajectories, building roads, or object contours in a game. For example, for a cubic spline defined by a four-point formula, it looks like this:

$$P(x) = (1-x)^3P_0 + 3(1-x)^2xP_1 + 3(1-x)x^2P_2 + x^3P_3. \quad (9)$$

To ensure smoothness and control the curvature of the spline, you need to calculate the first and second derivatives that describe the slope and curvature of the curve. We use the product and chain rules to calculate the derivative. The first derivative, $P'(t)$, is calculated by the formula (5).

$$P'(x) = -3(1-x)^2xP_0 + [3(1-x)^2 - 6(1-x)x]P_1 + [6(1-x)x - 3(1-x)x^2]P_2 + 3x^2P_3. \quad (10)$$

The second derivative, $P''(t)$, is calculated using the following formula:

$$P''(x) = 6(1-x)P_0 + [-6(1-x) + 6x]P_1 + [6(1-x) - 12x]P_2 + 6xP_3. \quad (11)$$

Both derivatives play a key role in the formation of splines. The smoothness of a spline is determined by the first derivative, $P'(x)$, which is responsible for the slope (or rate of change) of the spline. For a spline to look smooth, the first derivatives on neighboring segments must be identical. This means that the spline shouldn't have sudden slope changes, ensuring a smooth transition between points.

The first derivatives also reflect the rate of change of the spline's position over time, which can be useful in animation or motion modeling.

The second derivative, $P''(x)$, is responsible for the curvature of the spline. It describes how the slope of the spline changes over time. To ensure a smooth transition, the second derivatives must also coincide at the boundaries of neighboring segments. This ensures that the curvature of the spline does not change suddenly.

The second derivative also helps control the shape of the spline, allowing you to create smooth curves and avoid sharp turns. This is especially important in graphics and modeling, where aesthetics are critical.

In physical simulations, the second derivative can calculate the acceleration of objects moving along the spline, allowing you to model realistic movements.

Thus, the first and second derivatives in splines provide smoothness, curvature control, and aesthetic appeal of curves, making them essential tools in computer graphics, animation, and modeling, especially in landscape construction.

4 EXPERIMENTS

To conduct the experiments, a prototype game system was developed for the post-apocalyptic game *Broken World Wandere Last Horizon*, where players can explore, interact, and develop in a post-apocalyptic environment.

However, during the development of the menu, a problem arose with dropdown lists. The standard *Dropdown List* used in Unity uses *GrapiRaycaster* by default, which does not allow interactivity using a custom *Raycaster* system. This happens because *GraphicRaycaster* is tied to the main camera and does not consider the interaction through the new *Raycaster*. To solve this problem, we created a custom component called *CustomDropdown* based on *TMP_Dropdown*. It rewrote the interaction in such a way as to disable *GraphicRaycaster* on the *Dropdown List* element and replace it with a custom *MashColliderRaycaster* with the transfer of parameters to the new system. This approach ensures correct interaction with the drop-down list within the created system, allowing you to conveniently use the menu in the game, considering all the effects of shaders and other interactivity on any surface (Fig. 5).

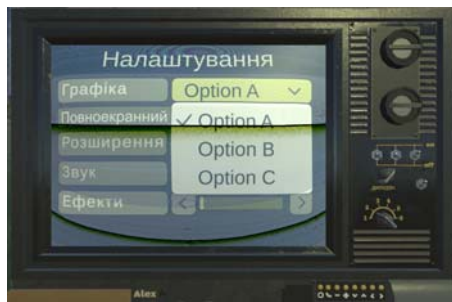


Figure 5 – The drop-down list

This approach is critical for creating more interactive and realistic interfaces, especially in games with open worlds or complex geometry. For example, in post-apocalyptic or sci-fi game worlds where the player can interact with large three-dimensional maps or objects, UI on a curved surface does not look artificial or disconnected from the environment. Instead, it seamlessly integrates into the world, following the shape of the surface it is placed on, making interaction more natural and attractive. With transparent textures and alpha channels, shaders allow you to layer different textures that can adapt to the surface while maintaining transparency and image structure, allowing you to create more sophisticated gaming interfaces.

It is worth noting the use of modified shaders for the realization of roads and other elements that should naturally follow the shape of the landscape. Using splines to create roads allows them to be placed smoothly and organically in accordance with the contour of the cityscape, ensuring realism and smooth transitions between different segments. To do this, the redefined shader and the code component responsible for width and smoothing dynamically consider and display the spline zones. The modified shader based on the spline ensures the correct repetition of the shape of the road surface by passing the *height* parameter, allowing it to naturally “wrap” the terrain (Fig. 6) while maintaining high-quality textures and details. This is especially important for games where landscape details play a significant role in the overall game experience, such as racing, shooters, or adventure games, or it makes it possible to use a spline to change the shape of the landscape dynamically during the game.

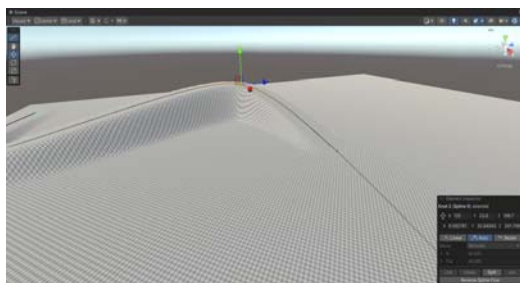


Figure 6 – Passing a spline height parameter for a road

The modified shader overlays the height of the road, taking into account the surface normals and its contours:

$$\text{Shader}_{\text{road}} = \text{Lerp}(\text{Texture}_{\text{base}}, \text{Texture}_{\text{detail}}, \text{Alpha Chan-nel}). \quad (12)$$

Lerp is an alpha-channel-based texture interpolation to create smooth transitions between road segments.

Using this approach allows you to place roads and other landscape elements by the shape of the terrain, ensuring realism and smooth transitions between road segments, i.e., giving smooth bends at their edges for greater realism (Fig. 7).

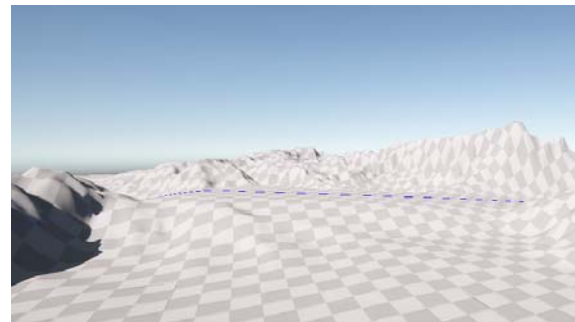


Figure 7 – Road on the landscape

In addition, modified shaders allow you to control texture updates in real-time, which can significantly improve game performance. For example, the interface texture can be updated on demand only when necessary, reducing the load on the system and increasing optimization. This approach allows the implementation of an interface that can adapt to any surface, such as a curved monitor panel in a futuristic vehicle or an interface superimposed on a virtual reality object.

In addition to improving interactivity, modified shaders significantly impact the visual style of the game. They allow you to achieve a new level of depth and detail in the design of game environments, making them more believable and rich. Well-chosen shaders can change the perception of the game, making it unique and inimitable. For example, in post-apocalyptic games like *Metro Exodus*, the integration of complex shaders can emphasize the abandonment and destruction of the environment by adding detail to broken or warped objects.

The image in Figure 8 shows how cubic splines behave during the experiment.

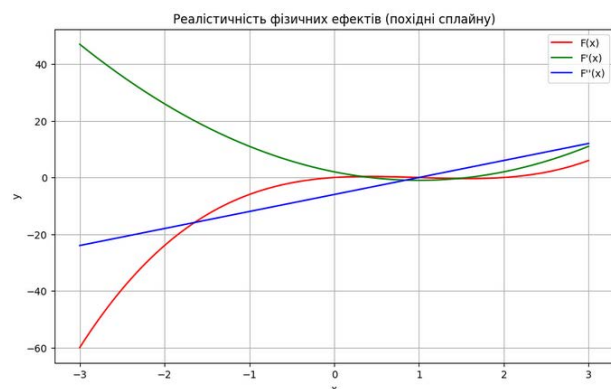


Figure 8 – Realistic physical effects

The curve shown in red ($F(x)$) shows smoothness, indicating that the spline provides continuous and smooth transitions between specified points. This is critical for modeling physical phenomena where abrupt changes produce unrealistic results.

The green line ($F'(x)$) shows how the slope of the red line changes. When the green line crosses the horizontal axis, the red line becomes flat and changes direction. And all this happens very smoothly, without jerks.

The blue line ($F''(x)$) shows the linear change in curvature, i.e., how the red line's curvature changes. This change occurs evenly in cubic splines, as if in a straight line. This demonstrates their ability to accurately model curvature where precise control of the curve shape is required.

5 RESULTS

The use of standard GraphicRaycaster is unacceptable because it depends on the main camera and cannot receive information directly from the interactive object. Based on the obtained intersection coordinates, the position on the canvas is determined, and a list of controls is generated for further processing.

A custom component based on BaseRaycaster was developed to ensure correct and efficient interaction with the canvas. The proposed component solves this problem by determining the point of intersection of the beam coming out of the camera with the physical surface of the canvas. The implementation includes an algorithm for searching for beam intersections with colliders using a mathematical formula and further processing the found elements, considering their depth to ensure correct interaction.

This approach allows for creating an interface on an arbitrary curved static surface, opening up vast possibilities for use in various games and interactive scenarios.

6 DISCUSSION

The problem of UI interactivity on uneven surfaces is relevant in modern game development, especially in open-world games, where complex geometric shapes are often used.

The presented research focuses on developing and optimizing graphical user interfaces in a 3D gaming environment, particularly on curved surfaces. Custom rails and modified shaders were used to solve the problem of UI interactivity on uneven surfaces. The proposed method, using mathematical formulas to determine the intersection of the ray with a physical surface, provides accurate localization of UI elements. This is consistent with studies [9] emphasizing the importance of accurate geometry and optimized rasterization pipelines. In both the study and article [9], accurate geometry is key to achieving high-quality results. In the study, accurate geometry is used to calculate ray intersections, and in [9], it is used to generate shading intervals.

The study used modified shaders to create roads that smoothly follow the contours of the landscape. This demonstrates the effectiveness of using splines and cubic

polynomials. The review notes that shaders are critical in creating realistic visual effects [1, 3].

The study's results confirm the effectiveness of using shaders to create realistic roads that smoothly follow the contours of the landscape. The use of splines and cubic polynomials allows for a high level of detail and naturalness of display.

The use of shaders allows for significant optimization and increased performance, which is important in creating immersive gaming environments.

The proposed approach achieves a high level of detail and smooth transitions, essential for open-world games. This is consistent with research [5], which proposes improvements to ray tracing algorithms to improve image quality and performance.

CONCLUSIONS

In summary, the use of rendered textures and modified shaders allows you to create more realistic game scenes with a high level of immersion and adds visual harmony between the interface and the game world. Spline technology and advanced shaders significantly impact interactivity, performance, and game style, providing smoothness, realism, and visual appeal.

The use of rendered textures and modified shaders significantly improves the integration of the interface into the game environment, providing a more organic interaction between the player and UI elements. This opens new possibilities for creating complex game worlds with interactive UI on non-standard surfaces, such as curved panels, monitors, or books. Another important aspect is the improvement of performance and the ability to adapt textures during the game.

This paper presents a comprehensive approach to designing UI for curved surfaces and using modified shaders to create realistic roads and other landscape elements.

The scientific novelty of the research is developing and successfully applying a ray-cast component using splines to provide efficient and intuitive interaction with the user interface on complex, curved 3D surfaces. This innovative approach solves the problem of the limitations of standard rake tools, which cannot adequately handle interaction with objects on uneven surfaces. Achieve high accuracy in determining the points of intersection of the beam with the surface and their projections by using mathematical spline models for detailed modeling of curves. Ensure reliable identification of active interface elements, considering their depth and possible overlap, which is key to creating an intuitive user interface.

The obtained results have practical significance because the developed and modified shaders can adapt textures during the game, which is an essential aspect for achieving high performance.

Further research is needed to develop methods for automatically adapting UI elements to the geometry of curved surfaces, which will greatly simplify the process of creating UI for complex objects.

The study results can be used to create more realistic and interactive virtual worlds.

REFERENCES

1. Ilett D. Building Quality Shaders for Unity®. Using Shader Graphs and HLSL Shaders. Berkeley, Apress, 2022, 734 p. ISBN 978-1-4842-8651-7. Mode of access: <https://doi.org/10.1007/978-1-4842-8652-4>.
2. Xiao D., Liu Z., Wang S. Metamorphic Shader Fusion for Testing Graphics Shader Compilers, *IEEE/ACM 45th International Conference on Software Engineering (ICSE)*, Melbourne, Australia, 14–20 May. 2023. Melbourne, Australia, IEEE, 2023, pp. 2400–2412. Mode of access: <https://doi.org/10.1109/ICSE48619.2023.00201>.
3. Souza D. A., Mota R. R. Materials for games – An overview on creating materials for games materials for games, *Extended Proceedings of the XX Brazilian Symposium on Games and Digital Entertainment, Porto Alegre, Brazil, 18–21 October. 2021*. Porto Alegre, Brazilian Computer Society, 2021, pp. 136–142. URL: https://doi.org/10.5753/sbgames_estendido.2021.19633.
4. Liang Y. et al. Automatic Mesh and Shader Level of Detail, *EEE Transactions on Visualization and Computer Graphics (TVCG)*, 2023, vol. 29, No. 10, pp. 4284–4295. Mode of access: <https://doi.org/10.1109/TVCG.2022.3188775>.
5. Akenine-Möller T. et al. Improved shader and texture level of detail using ray cones, *Journal of Computer Graphics Techniques (JCGT)*, 2021, Vol. 10, No 1, pp. 1–24. ISSN 2331-7418. Mode of access: <http://jcgt.org/published/0010/01/01/>.
6. Konurmah G., Chickerur S. GPU Shader Analysis and Power Optimization Model, *Engineering, Technology & Applied Science Research*, 2024, vol. 14, no 1, pp. 12925–12930. Mode of access: <https://doi.org/10.48084/etasr.6695>.
7. Sasso E., Loiacono D., Lanzi P. L. A Tool for the Procedural Generation of Shaders Using Interactive Evolutionary Algorithms, *2024 IEEE Gaming, Entertainment, and Media Conference (GEM)*, Turin, Italy, 5–7 June. 2024. Turin, Italy, 2024, pp. 1–4. Mode of access: <https://doi.org/10.1109/GEM61861.2024.10585418>.
8. Nishidate Y., Fujishiro I. Efficient Particle-Based Fluid Surface Reconstruction Using Mesh Shaders and Bidirectional Two-Level Grids, *Proceedings of the ACM on Computer Graphics and Interactive Techniques*, 2024, Vol. 7, No 1, pp. 1–14. Mode of access: <https://doi.org/10.1145/3651285>.
9. Tricard T. Interval Shading: using Mesh Shaders to generate shading intervals for volume rendering, *Proceedings of the ACM on Computer Graphics and Interactive Techniques*, 2024, Vol. 7, No 3, pp. 1–11. Mode of access: <https://doi.org/10.1145/3675380>.
10. Huo Y. et al. ed.: Nandigjav M., Mitra N. J., Hertzmann A. ShaderTransformer: Predicting Shader Quality via One-shot Embedding for Fast Simplification, *SIGGRAPH '22: ACM SIGGRAPH 2022 Conference Proceedings, Vancouver BC, Canada, 7–11 July 2022*. New York, NY, United States, 2022, pp. 1–9. ISBN 978-1-4503-9337-9. Mode of access: <https://doi.org/10.1145/3528233.3530722>.
11. Wiryadi F. I., Kosala R. R. Particle rendering using geometry shader, *1st International Conference on Game, Game Art, and Gamification (ICGGAG)*, Jakarta, Indonesia, 19–21 December 2016. Jakarta, Institute of Electrical and Electronics Engineers (IEEE), pp. 1–9. Mode of access: <https://doi.org/10.1109/ICGGAG.2016.8052661>.
12. Bissell B. et al. A cultural heritage game for entertainment, *Archiving Conference 2021, Granada, Spain, 1–6 July 2021*. Kilworth Lane, Springfield, VA 22151 USA, Society for Imaging Science and Technology, 2021, pp. 1–6. Mode of access: <https://doi.org/10.2352/issn.2168-3204.2021.1.0.2>.
13. Zhao L. et al. Graph Mining and Machine Learning for Shader Codes Analysis to Accelerate GPU Tuning, *Complex Networks and Their Applications XI, 4 January 2023*. Cham, Springer, 2023, pp. 426–439. Mode of access: https://doi.org/10.1007/978-3-031-21127-0_35.

Received 20.05.2025.
Accepted 18.09.2025.

УДК 004.925.8 + 519.6

МОДИФІКАЦІЯ ШЕЙДЕРІВ І РЕНДЕР-ТЕКСТУР НА КРИВОЛІНІЙНИХ ПОВЕРХНЯХ

Сугоняк І. І. – канд. техн. наук, доцент, доцент кафедри комп'ютерних наук, Державний університет «Житомирська політехніка», м. Житомир, Україна.

Марчук Г. В. – старший викладач кафедри комп'ютерних наук, Державний університет «Житомирська політехніка», м. Житомир, Україна.

Олексюк О. С. – здобувач кафедри комп'ютерних наук, Державний університет «Житомирська політехніка», м. Житомир, Україна.

АНОТАЦІЯ

Актуальність. Криволінійні поверхні є складними для відображення на плоскому екрані. Розробка інтерфейсу для таких поверхонь є актуальною задачею, яка потребує вирішення багатьох проблем. У даній роботі представлено підхід до розробки UI для криволінійних поверхонь, а також модифікація шейдерів для створення реалістичних елементів ландшафту. Об'єктом дослідження є розробка інтерфейсної системи на основі кастомного райкаста для забезпечення інтерактивності та занурення у ігровий світ.

Мета роботи. Основна мета даної роботи полягає у створенні, вдосконаленні та адаптації шейдерів на криволінійних поверхнях для досягнення більш ефективного рендерингу, зберігаючи при цьому високу якість візуалізації.

Метод. Розробка UI для криволінійних поверхонь потребує врахування особливостей геометрії. Для вирішення цієї проблеми було розроблено кастомний компонент на базі BaseRaycaster, який дозволяє визначати перетин променя від камери з фізичною поверхнею. З метою забезпечення коректної та ефективної взаємодії з полотном, розроблено кастомний компонент на основі BaseRaycaster. Запропонований компонент вирішує проблему шляхом визначення точки перетину променя, що виходить з камери, з фізичною поверхнею полотна. Реалізація включає в себе алгоритм пошуку перетинів променя з колайдерами за допомогою математичної формули та подальшу обробку знайдених елементів з урахуванням їх глибини для забезпечення коректної взаємодії. Завдяки цьому підходу, з'являється можливість створювати інтерфейс на довільній кривій статичній поверхні, що відкриває широкі можливості для використання в різноманітних ігрових та інтерактивних сценаріях.

Результати. Представлений підхід до розробки UI для криволінійних поверхонь та використання модифікованих шейдерів дозволяє створювати більш інтерактивні та реалістичні інтерфейси та ігрові світи.

Висновки. Використання сплайнів та модифікованих шейдерів забезпечує розташування тексту на криволінійних поверхнях та органічне розташування доріг та інших елементів ландшафту відповідно до контурів місцевості. Цей підхід є важливим для створення ігор з відкритим світом або складною геометрією, де UI на криволінійній поверхні виглядає природно та інтегровано в середовище.

КЛЮЧОВІ СЛОВА: розробка ігор, шейдери, сплайн, рендеринг, кастомний райкастер, криволінійні поверхні, користувацький інтерфейс, візуалізація.

ЛІТЕРАТУРА

1. Ilett D. Building Quality Shaders for Unity®. Using Shader Graphs and HLSL Shaders / D. Ilett. – Berkeley : Apress, 2022. – 734 p. – ISBN 978-1-4842-8651-7. – Mode of access: <https://doi.org/10.1007/978-1-4842-8652-4>.
2. Xiao D. Metamorphic Shader Fusion for Testing Graphics Shader Compilers / D. Xiao, Z. Liu, S. Wang // IEEE/ACM 45th International Conference on Software Engineering (ICSE), Melbourne, Australia, 14–20 May. 2023. – Melbourne, Australia : IEEE, 2023. – P. 2400–2412. – Mode of access: <https://doi.org/10.1109/ICSE48619.2023.00201>.
3. Souza D. A. Materials for games – An overview on creating materials for games materials for games / D. A. Souza, R. R. Mota // Extended Proceedings of the XX Brazilian Symposium on Games and Digital Entertainment, Porto Alegre, Brazil, 18–21 October. 2021. – Porto Alegre : Brazilian Computer Society, 2021. – P. 136–142. – URL: https://doi.org/10.5753/sbgames_estendido.2021.19633.
4. Automatic Mesh and Shader Level of Detail / Y. Liang et al. // IEEE Transactions on Visualization and Computer Graphics (TVCG). – 2023. – Vol. 29, No 10. – P. 4284–4295. – Mode of access: <https://doi.org/10.1109/TVCG.2022.3188775>.
5. Improved shader and texture level of detail using ray cones / T. Akenine-Möller et al. // Journal of Computer Graphics Techniques (JCGT). – 2021. – Vol. 10, No 1. – P. 1–24. – ISSN 2331-7418. – Mode of access: <http://jcgt.org/published/0010/01/01/>.
6. Konnurmuth, G. GPU Shader Analysis and Power Optimization Model / G. Konnurmuth, S. Chickerur // Engineering, Technology & Applied Science Research. – 2024. – Vol. 14, No 1. – P. 12925–12930. – Mode of access: <https://doi.org/10.48084/etasr.6695>.
7. Sasso E. A Tool for the Procedural Generation of Shaders Using Interactive Evolutionary Algorithms / E. Sasso, D. Loiacono, P. L. Lanzi // 2024 IEEE Gaming, Entertainment, and Media Conference (GEM), Turin, Italy, 5–7 June. 2024. – Turin : Italy, 2024. – P. 1–4. – Mode of access: <https://doi.org/10.1109/GEM61861.2024.10585418>.
8. Nishidate Y. Efficient Particle-Based Fluid Surface Reconstruction Using Mesh Shaders and Bidirectional Two-Level Grids / Y. Nishidate, I. Fujishiro // Proceedings of the ACM on Computer Graphics and Interactive Techniques. – 2024. – Vol. 7, No 1. – P. 1–14. – Mode of access: <https://doi.org/10.1145/3651285>.
9. Tricard T. Interval Shading: using Mesh Shaders to generate shading intervals for volume rendering / T. Tricard // Proceedings of the ACM on Computer Graphics and Interactive Techniques. – 2024. – Vol. 7, No 3. – P. 1–11. – Mode of access: <https://doi.org/10.1145/3675380>.
10. ShaderTransformer: Predicting Shader Quality via One-shot Embedding for Fast Simplification / Y. Huo et al. // SIGGRAPH '22: ACM SIGGRAPH 2022 Conference Proceedings, Vancouver BC, Canada, 7–11 July 2022 / ed.: M. Nandigjav, N. J. Mitra, A. Hertzmann. – New York, NY : United States, 2022. – pp. 1–9. – ISBN 978-1-4503-9337-9. – Mode of access: <https://doi.org/10.1145/3528233.3530722>.
11. Wiryadi F. I. Particle rendering using geometry shader / F. I. Wiryadi, R. R. Kosala // 1st International Conference on Game, Game Art, and Gamification (ICGAG), Jakarta, Indonesia, 19–21 December 2016. – Jakarta : Institute of Electrical and Electronics Engineers (IEEE). – P. 1–9. – Mode of access: <https://doi.org/10.1109/ICGAG.2016.8052661>.
12. A cultural heritage game for entertainment / B. Bissell et al. // Archiving Conference 2021, Granada, Spain, 1–6 July 2021. – Kilworth Lane, Springfield, VA 22151 USA : Society for Imaging Science and Technology, 2021. – P. 1–6. – Mode of access: <https://doi.org/10.2352/issn.2168-3204.2021.1.0.2>.
13. Graph Mining and Machine Learning for Shader Codes Analysis to Accelerate GPU Tuning / L. Zhao et al. // Complex Networks and Their Applications XI, 4 January 2023. – Cham : Springer, 2023. – P. 426–439. – Mode of access: https://doi.org/10.1007/978-3-031-21127-0_35.

МЕХАНІЗМИ СХЕМОТЕХНІЧНОГО МОДЕЛЮВАННЯ НА ОСНОВІ ВЕКТОРНОЇ ЛОГІКИ

Хаханов В. І. – д-р техн. наук, професор кафедри автоматизації проектування обчислювальної техніки, Харківський національний університет радіоелектроніки, Україна.

Обрізан В. І. – канд. техн. наук, докторант кафедри автоматизації проектування обчислювальної техніки, Харківський національний університет радіоелектроніки, Україна.

Шевченко О. Ю. – канд. техн. наук, старший викладач кафедри автоматизації проектування обчислювальної техніки, Харківський національний університет радіоелектроніки, Україна.

Хаханова Г. В. – д-р техн. наук, професор кафедри автоматизації проектування обчислювальної техніки, Харківський національний університет радіоелектроніки, Україна.

Хаханов І. В. – канд. техн. наук, асистент кафедри автоматизації проектування обчислювальної техніки, Харківський національний університет радіоелектроніки, Україна.

Демченко О. І. – аспірант кафедри автоматизації проектування обчислювальної техніки, Харківський національний університет радіоелектроніки, Україна.

Воронов А. О. – аспірант кафедри автоматизації проектування обчислювальної техніки, Харківський національний університет радіоелектроніки, Україна.

АНОТАЦІЯ

Актуальність. Розглядаються актуальні для EDA-ринку питання зменшення вартості та часу тестування й верифікації цифрових проєктів шляхом синтезу логічного вектора цифрової схеми, що дозволяє суттєво спростити алгоритми good-value simulation, а синтез карти тестування звести до трьох матричних операцій.

Мета. Мета дослідження – зменшення вартості та часу тестування і верифікації цифрових проєктів завдяки синтезу логічного вектора цифрової схеми, що дозволяє спростити алгоритм побудови карти тестування до трьох матричних операцій.

Метод. Пропонується синтез логічного вектора комбінаційної схеми для good-value та fault як адресів моделювання в архітектурі промпт-інтелектуального in-меморі комп'ютигу. Логічний вектор є найбільш технологічним, компактним та вичерпним представленням схеми для економного вирішення всіх задач дизайну та тестування. Пропонується декартова логіка, яка завдяки експоненційній надлишковості 2^{n+m} є ефективним інтелектуальним механізмом для вирішення комбінаційних задач (моделювання, simulation, тестування, діагностування) алгоритмами лінійної обчислювальної складності. Пропонуються механізми побудови логічного вектора процесу або явища, функції або структури на основі матричних структур декартової логіки для вирішення задач Modeling for Simulation. Пропонується інженерний метод прямого паралельного good-value моделювання схеми на основі використання логічних векторів елементів та таблиць істинності. Декартова логіка – це логічний вектор (матриця), як результат моделювання декартових логічних відношень між бітами логічних векторів або адресами таблиці істинності. Декартова логіка вирішує задачі: 1. Good-value modeling circuit logic vector без алгоритму моделювання справної поведінки. 2. Fault modeling testing map of logic без алгоритму моделювання несправностей.

Результати. Запропоновано математичний апарат декартової логіки, представлений у вигляді логічного вектора (матриці), що є результатом моделювання декартових логічних відношень між бітами логічних векторів або адресами таблиці істинності. На основі теорії векторної логіки розроблено механізми та програмне застосування для економного розв'язання задач design and test в архітектурі in-memory комп'ютигу.

Висновки. Практична значущість дослідження орієнтована на вирішення всіх задач design and test на основі використання простих моделей векторної логіки, орієнтованої на економічний in-меморі комп'ютигу на основі read-write транзакцій, вільний від інструкцій процесора. Пропонується використовувати логічний вектор специфікації в архітектурі in-memory комп'ютигу для тестування, верифікації, діагностування та експлуатації. In-memory vector logic computing – це економічне вирішення багатьох обчислювальних проблем з точки зору енергетичних, часових і вартісних витрат. Новизна дослідження полягає у використанні векторної логіки в архітектурі in-memory комп'ютигу на основі read-write транзакцій, що дозволяє зменшити споживання ресурсів у вигляді часу та енергії. Імплементація векторної логіки в in-memory комп'ютигу робить його масовим і енергоефективним, вільним від складних алгоритмів аналізу великих даних.

КЛЮЧОВІ СЛОВА: векторна логіка, логічний вектор, декартова логіка, інтелектуальний комп'ютигу, карта тестування, моделювання несправностей, good-value моделювання, in-memory моделювання для симуляції, prompt-інжиніринг.

АБРЕВІАТУРИ

EDA – Electronic Design Automation (Автоматизація проектування електроніки);

GUI – Graphical User Interface (Графічний інтерфейс користувача);

HDL – Hardware Description Language (Мова опису апаратури);

ДНФ (DNF) – диз'юнктивна нормальна форма;

КНФ (CNF) – кон'юнктивна нормальна форма;

AI – Artificial Intelligence (Штучний інтелект);

NI – Number of Inputs (Кількість входів);

MOSI – Modeling for Simulation.

НОМЕНКЛАТУРА

X – вхідні логічні вектори / черговий елемент схеми;

X_i – біт вхідного ефекту / координата вхідного набору;

Y – вихідні логічні вектори / результат декартового добутку;

L – логічний вектор (результат моделювання, матриця активності, специфікація);

L_{ij} – координата логічного вектора;

M – модель (пам'ять, матриця, структура даних);

M_i – біт вектора моделювання / координата пам'яті;

T – таблиця істинності;

A – алгоритм, зокрема перекодування / складність моделі;

F – помилки, несправності;

D – матриця дедуктивних векторів;

H – матриця перекодування;

G – утворювальний логічний вектор (функція декартової операції);

n – кількість бітів (змінних / входів);

m – середня кількість входів у логічному елементі;

k – довжина тесту;

$2^n, 2^m, 2^k$ – розмірність логічного вектора / кількість бітів;

$\otimes$ – декартова логічна операція (комбінаторна суперпозиція);

$\vee, \wedge, \neg$ – диз'юнкція (OR), кон'юнкція (AND), заперечення (NOT).

ВСТУП

Розглядаються актуальні для EDA-ринку питання зменшення вартості та часу тестування й верифікації цифрових проєктів шляхом синтезу логічного вектора цифрової схеми, що дозволяє суттєво спростити алгоритми good-value simulation, а синтез карти тестування звести до трьох матричних операцій. Структури даних для промпт-комп'ютингу (моделювання) логічного вектора схеми визначаються як: 1) два вхідних логічних вектори, довжина яких кратна ступеню двійки; 2) номери для ідентифікації вхідних змінних кожного вектора, які беруться з логічної схеми; 3) результуючий вектор схеми та вектор ідентифікаторів вхідних змінних схеми. Декартовий добуток двох векторів різної або однакової довжини, що відповідають фрагменту схеми з двох елементів, виконується на основі бінарної логічної операції (and, xor, or). Ця операція над векторами дозволяє отримати нові, раніше невідомі властивості в результуючому логічному векторі (матриці) для обробки даних.

Об'єкт дослідження – комп'ютинг на основі векторної і декартової логіки для оцифрованих процесів та явищ.

Предмет дослідження – механізми в моделюванні цифрових проєктів.

Мета дослідження – зменшення вартості та часу тестування і верифікації цифрових проєктів завдяки синтезу логічного вектору цифрової схеми, що дозволяє спростити алгоритм побудови карти тестування до трьох матричних операцій.

1 ПОСТАНОВКА ЗАДАЧІ

Розглядається задача моделювання логічного вектора для цифрової структури. Нехай задано логічні вектори X і Y . Практичне завдання полягає у побудові

матриці відношень за допомогою логічної декартової операції ($\otimes$ -вектора) $L = X \otimes Y$ над бітами двох логічних векторів X і Y з метою вирішення задач modeling for simulation: 1) моделювання L -матриці активності вихідних станів логічної функції Y ; 2) моделювання H -матриці перекодування; 3) побудова матриці дедуктивних векторів $D = L_H$ для моделювання несправностей функціональності або схеми як адрес таблиці істинності; 4) використання декартової AND-логіки для синтезу матриці, яка визначає властивість similarity (схожості) двох векторів у двовимірному просторі; 5) побудови карти тестування схеми за вектором.

Задачі дослідження: 1) аналіз існуючих підходів до побудови логічного вектора процесу або явища, функції чи структури; 2) визначення метрики алгоритмічної складності розв'язання задач моделювання схем на основі елементів або логічного вектора; 3) визначення матричних надлишкових структур декартової логіки для ефективного вирішення задач vector logic modeling and simulation [24–27]; 4) розробка методу прямого паралельного моделювання схеми на основі використання декартової логіки з метою отримання логічного вектора; 5) кодування та верифікація запропонованого методу на різних структурах логічних схем; 6) використання логічного вектора для prompt-інжинірингу інтелектуального комп'ютингу з метою побудови карти тестування.

2 ОГЛЯД ЛІТЕРАТУРИ

Розгалуження в логічній схемі [1–4] породжують однойменні вхідні змінні (координати) на одному або кількох логічних елементах. Така особливість схеми призводить до мінімізації результативного вектора (координат матриці) фрагмента структури за рахунок однойменних логічних змінних – шляхом залишення координат, які відповідають однаковим значенням вхідних сигналів (адресних бітів) цих змінних, та виокреслення координат з різними вихідними сигналами. У результуючому векторі L завжди залишається кількість координат, кратна ступеню двійки (рис. 1).

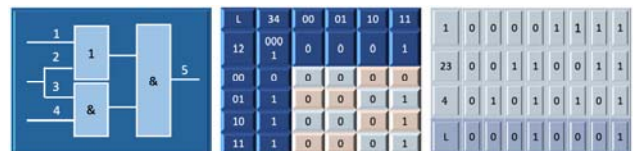


Рисунок 1 – Синтез логічного вектора для фрагмента схеми

Побудова результуючого логічного вектора двох взаємодіючих елементів на основі третього бінарного логічного вектора від двох змінних. Це універсальна процедура, вільна від інструкцій процесора, яка використовує лише read-write транзакції, орієнтовані на інтеграцію комп'ютинг. Тому при побудові карти тестування будь-якої логічної функціональності можна використовувати логічні вектори та read-write транза-

кції. Більше того, всі задачі design and test (як і будь-яке інше завдання) можна вирішувати на основі read-write транзакцій над логічними векторами без участі інструкцій процесора. Моделювання результуючого логічного вектора (у більше ніж двох вимірах) на основі третього логічного вектора, від більш ніж двох змінних, застосовується для формування відношень між більш ніж двома векторами. Це паралельна процедура для математичного гурмана. Слід мати на увазі, що логічний вектор для будь-якої схеми може бути побудований шляхом рекурсивної суперпозиції завжди двох векторів: поточного актуального вектора схеми та вектора наступного елемента. Побудову логічного вектора схеми можна розглядати як паралельне good-value пряме моделювання вичерпного тесту або вхідних наборів – це обов'язкова процедура при верифікації проєкту.

Якщо розглядати схему та логічний вектор схеми, то дуже важко знайти пояснення, чому використовується саме схема замість простого логічного вектора для задання функціональності. Переваги схеми раніше визначалися як високошвидкісний механізм отримання значень на виходах при мінімальному обсязі пам'яті, яка тоді була дуже дорогою.

Сьогодні read-write транзакції в пам'яті мають ту саму затримку (latency – 0,6 нс), що й логічні елементи. При цьому кожен користувач практично не обмежений у розмірах пам'яті, яку можна використати для розв'язання задачі. Висновок простий: векторно-логічний комп'ютинг у пам'яті на основі read-write транзакцій – це майбутнє масових обчислень. Він має мінімальні: вартість створення моделі (яка не потребує синтезу), енергоспоживання, затримку під час експлуатації моделі. До речі, всі компанії, що займаються штучним інтелектом, сьогодні переходять на прості двійкові моделі, розміщені в пам'яті, представлені у вигляді логічних векторів або матриць (а це одне й те саме) [5–10].

Схемотехніка вже 80 років потребує механізмів синтезу, які є гальмом на шляху створення інтелектуального in-memory комп'ютингу [11–13]. З іншого боку, існує альтернатива – будь-який процес або явище можна легко представити у вигляді векторів (патернів), двійково й унітарно закодованих на універсумі примітивних символів. Синтез не потрібен. Потрібне лише моделювання – процес розміщення векторної або матричної моделі в пам'яті. Потім цю двійкову модель можна використати для розв'язання всіх задач штучного інтелекту методами логічного двійкового good-value моделювання на основі read-write транзакцій. Чи не час уже замислитися над масовим економічним in-memory комп'ютингом на основі таблиць істинності або логічних векторів як інженерної основи інтелектуального комп'ютингу?

Час декогеренції [14–15] – це період існування стійкого динамічного процесу або явища в мікро- або макросвіті, підтримуваного ззовні. У мікросвіті квантового комп'ютингу таку підтримку забезпечує тер-

мостат або додаткове охолодження апаратури. У макросоціальному світі – це зовнішні гроші або зброя. Для класичного або квантового комп'ютингу – це енергетична підтримка обчислень потужним процесором або апаратурою охолодження. В усіх випадках стабільність динамічного процесу зникає, коли припиняється зовнішня допомога. Тому in-memory комп'ютинг на основі read-write транзакцій виступає як майбутній масовий інтелектуальний комп'ютинг для обслуговування людства. Інтелектуальний комп'ютинг – це інтеграція класичних, квантових і штучно-інтелектуальних in-memory архітектур для формування гармонійних відносин між метою та засобами її досягнення [16]. Для кожної задачі потрібно обирати економічний механізм отримання ефективного розв'язання з точки зору енерговитрат і часу (рис. 2).

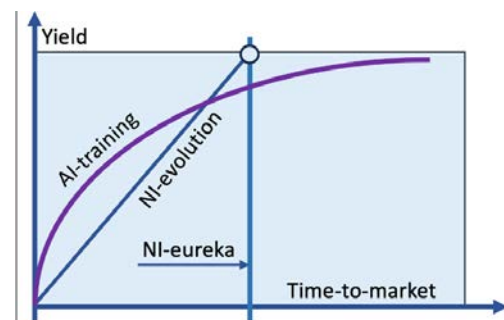


Рисунок 2 – Відносини між штучним та природним інтелектом

Індукція [17] – це шлях переходу від окремого до загального – мати штучного інтелекту (Artificial Intelligence), яка проходить довгий імовірнісний шлях навчання, що веде крізь море помилок. Дедукція [18] – це шлях від загального до конкретного – мати природного інтелекту (Natural Intelligence), – це множина великих і малих помилок, що призводять до великих і малих відкриттів. «Ймовірність (випадковість) – це фіговий листок на голому тілі нашого невігластва», – Альберт Ейнштейн. Імовірнісний процес із часом вивчення стає детермінованим. І навпаки – коли багато детермінованих явищ перетворюються на випадкові процеси. Штучний інтелект (Artificial Intelligence) гарантує рішення задачі з визначеною точністю за заданий час, але не може одразу відповісти на просте запитання: скільки буде 2×2 ? Для цього потрібні декілька ітерацій навчання AI-механізму. Природний інтелект (Natural Intelligence) може одразу дати точну відповідь на просте завдання або знайти розв'язок складної задачі за невизначений час (годину, день, місяць, рік) – або взагалі не знайти. Завжди потрібно обирати прийнятні – детерміновані або ймовірнісні – механізми розв'язання задач.

3 МАТЕРІАЛИ І МЕТОДИ

Векторна та декартова логіка. Логічний вектор – це представлення процесу або явища, функції чи

структури у вигляді послідовності нулів і одиниць довжиною 2^n , де n – кількість бітів двійкових змінних для формування адрес бітів вектора. Це визначення робить логічний вектор незалежним від таблиці істинності. Водночас обов'язковим неявним атрибутом вектора є стандартні адреси, які легко згенерувати для будь-якого вектора. Декартовий добуток одного вектора з самим собою формує рефлексивне (рис. 3) відношення $L = Y \otimes Y$, яке генерує декартову матрицю – у цьому випадку матрицю активності або логічний вектор L розмірності 2^{n+1} .

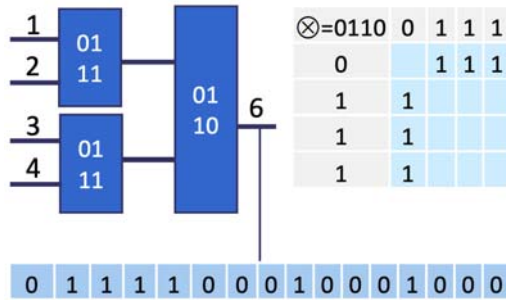


Рисунок 3 – Рефлексивне хог-відношення вектора 0111

Декартовий добуток двох різних векторів формує або синтезує новий вектор розмірності $2^{n+m} = 2^n \times 2^m$. Схематичне позначення такої операції може бути представлено трьома елементами (рис. 4). Кожне розширення простору логічного вектора супроводжується появою нових цікавих властивостей, орієнтованих на практичне розв'язання задач аналізу. Наприклад, матриця-вектор, отриманий у результаті декартової операції, вказує ті координати, які відрізняють два вектори (дві функціональності) – 1000 і 11011011 – між собою на вичерпному тесті. Наведена схема є хорошим прикладом, що демонструє моделювання логічного вектора на основі векторів логічних елементів, які входять до складу схеми. Моделювання логічного вектора – це задача, актуальна для EDA-ринку, оскільки логічний вектор схеми – це найекономніша модель для розв'язання всіх задач design and test (і не тільки).

Наведена схема є архітектурою (структурою) верифікації будь-якого цифрового проекту на основі специфікації та актуальної моделі цифрового пристрою. Результат верифікації представлений логічним вектором, де одиничні координати показують невідповідність результатів тестування у просторі верифікації проекту.

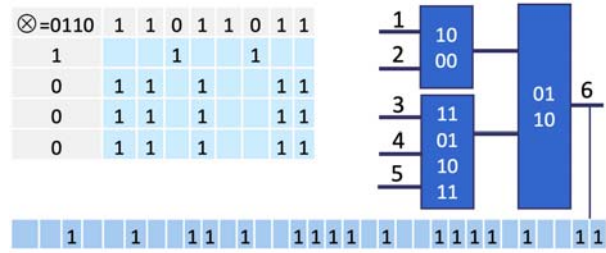


Рисунок 4 – Схематика декартового добутку

Розклад або декомпозиція логічного вектора – це практично орієнтоване завдання: як один елемент або вектор представити сукупністю векторів чи елементів із тими ж властивостями. Частково таке завдання вирішується шляхом синтезу схеми на основі логічного вектора. Проте це шлях назад до комп'ютингу вчорашнього дня. Фактично, теорія векторно-логічного комп'ютингу є дизраптором не лише для синтезу в традиційну систему елементів кремнієвого кристалу, а й для класичного комп'ютингу, що базується на архітектурі фон Неймана з центральним процесором. Крім того, логічний вектор – це ідеальна модель для представлення даних у кубітах квантового комп'ютингу. Логічним вектором і його похідними можна описати будь-які процеси, що існують сьогодні в кіберпросторі, фізичному просторі та соціальному просторі. Більше того, за допомогою логічного вектора можна керувати будь-якими об'єктами в цих трьох просторах на основі загального моніторингу й аналізу, які також можуть бути реалізовані за допомогою логічного вектора. Масштабування логічного вектора на будь-якому рівні ієрархії та розмірності залишає його вектором. Єдиним недоліком логічного вектора є експоненційний обсяг пам'яті, необхідної для його зберігання. Проте пам'ять у сучасному кіберпросторі вже не є дефіцитною. Оригінальність дослідження полягає у використанні векторної логіки в архітектурі in-мемогу комп'ютингу на основі read-write транзакцій, що дозволяє зменшити споживання ресурсів у вигляді часу та енергії. Імплементация векторної логіки в in-мемогу комп'ютинг робить його масовим та енергоощадним, вільним від складних алгоритмів аналізу структур даних.

Векторна логіка – це представлення логічних операцій векторами станів таблиці істинності. Таблиця істинності має стандартизовані адреси бітів логічного вектора вихідних станів, що дає змогу використовувати логічний вектор як самостійну модель, розміщену в пам'яті. У разі потреби для цієї векторної моделі завжди можна автоматично згенерувати адреси кожного біта або побудувати таблицю істинності. Генерація логічного вектора передбачає його розміщення в пам'яті, де доступ до кожного біта забезпечується архітектурою адресних дешифраторів.

Декартова логіка – це матриця відношень, побудована за допомогою логічної декартової операції ($\otimes$ -вектора) $L = X \otimes Y$ над бітами двох логічних векторів X і Y . Усі три вектори X , Y , $\otimes$ є логічними. Вектор $\otimes$

(G) виступає як вектор, що формує відношення між X та Y . І навпаки: будь-який з них (X , Y , G) може виступати як вектор, що встановлює відношення між двома іншими. Для формування декартового відношення необхідно дотримуватись таких умов для форматів векторів: $X = 2^n$, $Y = 2^m$, $G = 2^k$ (де n , m – будь-які, а $k = 2$). Декартова логіка – це розширення або надлишковість векторної логіки, яка дозволяє технологічно вирішувати ті задачі, які створюють проблеми для самої векторної логіки. Із теорії та практики відомо, що будь-яка надлишковість структур даних знижує обчислювальну складність алгоритму аналізу таких моделей. Оскільки модель і алгоритм утворюють єдиний механізм комп'ютигу, збільшення одного з компонентів призводить до зменшення іншого, і навпаки. Якщо необхідно розв'язати комбінаторну задачу лінійним алгоритмом, потрібно збільшити надлишковість моделі до експоненційної складності. Це сьогодні – тренд №1 у світі, коли інтелектуальні надлишкові моделі використовуються для вирішення всіх задач комп'ютигу лінійними або нульовими алгоритмами.

Кілька прикладів практичного застосування декартової логіки для задач modeling for simulation (рис. 5):

1. Моделювання L -матриці активності вихідних станів логічної функції Y : $L = Y \oplus Y$. Одиничні координати матриці визначають активність виходів при подачі вхідного набору, що задається адресами бітів логічного вектора.

2. Моделювання H -матриці перекодування $H = A \oplus A$, яка являє собою десяткові коди відстаней між адресами таблиці істинності будь-якої логічної функції. Це логічна константа для генерації всіх комбінацій несправностей.

3. Побудова матриці дедуктивних векторів $D = L_H$ для моделювання несправностей функціональності або схеми як адрес таблиці істинності.

4. Декартова AND-логіка для синтезу матриці, яка визначає властивість similarity (схожості) двох векторів у двовимірному просторі.

Побудова карти тестування схеми за вектором [24–26] – це один із потужних практичних аргументів на користь генерації логічного вектора, який звільняє інженера від необхідності писати складні алгоритми синтезу тестів і моделювання несправностей (рис. 6).

Механізм побудови карти тестування за логічним вектором 10010100 передбачає всього чотири матричні операції. Отриманий результат дозволяє вирішувати всі завдання: моделювання несправностей, тестування, верифікація, пошук і діагностування дефектів, оцінка якості тесту, вибір мінімального тесту. Константна складність матричних операцій на передбачуваному обсязі структур даних – це все, що потрібно для отримання карти тестування функціональності, заданої логічним вектором.

Можна говорити про дві метрики побудови карти тестування схеми:

1. Механізм моделювання несправностей на дедуктивних векторах елементів схеми:

$$C = k \times n \times 2^m \times n^2,$$

де k – довжина тесту, n – кількість логічних елементів у схемі, m – середня кількість входів в одному елементі, $n \times 2^m$ – складність генерації дедуктивних векторів, n^2 – складність моделювання несправностей лінії схеми на одному векторі.

1	L-matrix: $L=Y \oplus Y$	2	H-matrix: $H=A \oplus A$	3	D-matrix: $D=L_H$	4	L-matrix: $L=X \wedge Y$
Y	0 0 0 1 1 0 1 1	Y	000 001 010 011 100 101 110 111	Y	Y \ T \ 000 001 010 011 100 101 110 111	Y	0 0 0 1 1 0 1 1
0		000	0 1 2 3 4 5 6 7	0 000		0 000	
0		001	1 0 3 2 5 4 7 6	0 001		0 001	
0		010	2 3 0 1 6 7 4 5	0 010		0 010	
1	1 1 1 1	011	3 2 1 0 7 6 5 4	1 011		1 011	
1	1 1 1 1	100	4 5 6 7 0 1 2 3	1 100		1 100	
0		101	5 4 7 6 1 0 3 2	0 101		0 101	
1	1 1 1 1	110	6 7 4 5 2 3 0 1	1 110		1 110	
1	1 1 1 1	111	7 6 5 4 3 2 1 0	1 111		1 111	

Рисунок 5 – Практичне використання розширеної декартової логіки

	L-matrix	H-matrix	D-matrix	C-matrix (testing map)
	1 0 0 1 0 1 0 0		000 001 010 011 100 101 110 111	000 001 010 011 100 101 110 111
1	1 1 1 1 1 1	0 1 2 3 4 5 6 7	000 1 1 1 1 1 1	000 .1 .1. 1.. 11. 111
0 1	1 1	1 0 3 2 5 4 7 6	001 1 1 1	001 ..0 .1. 1..
0 1	1 1	2 3 0 1 6 7 4 5	010 1 1	010 ..1 .0. 101
1 1 1 1 1 1	3 2 1 0 7 6 5 4	011 1 1 1 1 1 1	011 1 1 1 1 1 1	011 ..0 .0. 1.. 1.0 100
0 1	1 1	4 5 6 7 0 1 2 3	100 1	100 ..1 0.. 011
1 1 1 1 1 1	5 4 7 6 1 0 3 2	101 1 1 1 1 1	101 1 1 1 1 1	101 ..0 .1. .10 0.. 010
0 1	1 1	6 7 4 5 2 3 0 1	110 1 1 1 1	110 ..01 0.1 00.
0 1	1 1	7 6 5 4 3 2 1 0	111 1 1 1 1	111 ..0. 0.. 000

Рисунок 6 – Побудова карти тестування схеми за її логічним вектором

2. Механізм моделювання карти тестування за логічним вектором схеми: $C = 3 \times 2^m$, де 3×2^m – довжина логічного вектора схеми, m – кількість зовнішніх змінних схеми, 3 – кількість матричних операцій для моделювання карти тестування.

3. Слід також враховувати складність алгоритмів і моделей, необхідних для програмування коду моделювання несправностей. У другому випадку код становить лише 100 операторів мови Python, у першому – близько 1000.

4. Варто враховувати алгоритмічну складність побудови логічного вектора схеми: $C = n \times 2^{2m}$, де n – кількість елементів у схемі, m – середня кількість входів у кожному елементі.

5. Інтегральна метрика складності побудови карти тестування для схеми через моделювання логічного вектора має вигляд: $C = n \times 2^{2m} + 3 \times 2^m$, що співмірно зі складністю алгоритму дедуктивного моделювання несправностей у схемі.

Практичний висновок – якщо специфікація функціональності задана логічним вектором, то немає потреби використовувати його для синтезу комбінаційної схеми, щоб потім моделювати цю схему задля отримання логічного вектора для подальшої верифікації функціональності. Слід безпосередньо використовувати логічний вектор специфікації в архітектурі in-методу комп'ютиру для тестування, верифікації, діагностування та експлуатації. In-memory vector logic комп'ютинг – це економічне вирішення багатьох обчислювальних проблем з погляду енергетичних, часових і вартісних витрат.

Карту тестування також можна отримати за одну декартову операцію над вектором логічної функції. Для цього необхідно виконати хог-операцію над бітами логічного вектора. Одиначні значення координат отриманої матриці формують активність тестових наборів, на яких перевіряється комбінація несправностей, визначених одиначними координатами адрес таблиці істинності логічних несправностей. При цьому ознаки несправностей завжди будуть інверсними щодо бітів тестового набору. Синтез логічного вектора схеми передбачає на початковому етапі розбиття або декомпозицію багатовходових елементів на двовходові. Це дозволяє застосувати декартову логіку для отримання логічного вектора схеми або її фрагмента. Для регулярних функцій (and, or, xor) і їхніх похідних завжди можна побудувати таке розбиття. Для довільної векторної логіки таке розбиття також завжди існує (рис. 7), оскільки будь-який логічний

вектор може бути представлений у вигляді диз'юнктивної нормальної форми (ДНФ) або кон'юнктивної нормальної форми (КНФ), у яких використовуються регулярні логічні операції and, or, not.

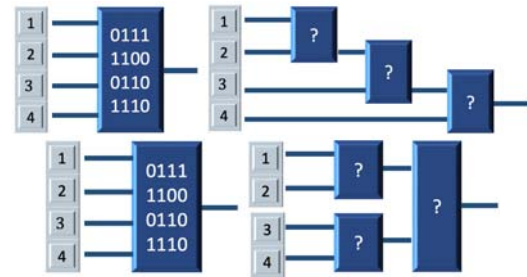


Рисунок 7 – Розбиття складних елементів на двовходові

У загальному випадку слід використовувати такий шаблон побудови логічного вектора за двома іншими векторами будь-якої розмірності, кратної 2^n , тобто ступеню двійки, та логічною функцією від двох змінних, яка є утворювальною (рис. 8).

Можна стверджувати, що завжди формується логічний вектор з кількістю бітів 2^{n+m} , де n і m – довжини двох векторів, які утворюють матричне декартове відношення. Моделювання логічного вектора для цифрової структури. Логічний вектор процесу або явища дозволяє розв'язувати всі задачі комп'ютиру в економічному режимі за лінійний час виконання алгоритму. Тому надзвичайно важливо також будувати логічний вектор для цифрової схеми. Для цього використовуємо рекурсивний алгоритм на основі декартової логіки. Це метод отримання матриці логічного вектора L схеми шляхом декартового добутку бітів двох векторів $X \otimes Y$ за правилом логічної функції G ($\otimes$). При цьому вектор Y – це сукупний вектор уже оброблених елементів схеми, а вектор X – це вектор чергового елемента схеми, який підлягає обробці. У результаті фактично отримується логічний вектор схеми на повному вичерпному тесті, який можна інтерпретувати як суперпаралельне good-value моделювання змінної структури.

Декомпозиція логічного вектора (рис. 9) здійснюється на основі рівняння його синтезу $L = X \otimes Y$. Необхідно знати два компоненти рівняння: L та $\otimes$. Потрібно визначити вектори X і Y , які перебувають між собою у відношенні $\otimes$, де X і Y повинні відповідати певним співвідношенням довжин, що важливо для формування формату матриці вектора.

$\otimes$	1	0	0	1	0	1	1	0
0								
1								
1								
0								

$\otimes=0111$	1	0	0	1	0	1	1	0
0	1			1	1	1		
1	1	1	1	1	1	1	1	1
1	1	1	1	1	1	1	1	1
0	1			1	1	1		

$\otimes=0110$	1	0	0	1	0	1
0	1			1		
1			1	1		
1			1	1		
0	1			1		

Рисунок 8 – Побудова логічного вектора-матриці фрагмента схеми на основі декартової логіки

Для визначення векторів X та Y працює правило: $X \otimes L_{ij} = Y$. У цьому процесі також може бути використана модель штучного інтелекту.

$\otimes=0001$	1	0	0	1	0	1	1	0
0								
1	1			1		1	1	
1	1			1		1	1	
0								

$L=X \otimes Y$

Рисунок 9 – Декомпозиція логічного вектора

Для декомпозиції вектора немонотонної функції необхідно представити її у вигляді матриці на основі відомого утворювального G -вектора (рис. 10).

Initial form for decomposition								
$\otimes=0001$								
00=0								
01=0	1			1		1	1	
10=0	1			1		1	1	
11=1								

$L=X \otimes Y \rightarrow 1=1 \otimes 1$

Resulting decomposition								
$\otimes=0001$	1	0	0	1	0	1	1	0
0								
1	1			1		1	1	
1	1			1		1	1	
0								

$L=X \otimes Y$

Initial form for decomposition								
$\otimes=0111$								
00=0	1			1		1	1	
01=1	1	1	1	1	1	1	1	1
10=1	1	1	1	1	1	1	1	1
11=1	1			1		1	1	

$L=X \otimes Y \rightarrow 0=0 \otimes 0$

Resulting decomposition								
$\otimes=0111$	1	0	0	1	0	1	1	0
0	1			1		1	1	
1	1	1	1	1	1	1	1	1
1	1	1	1	1	1	1	1	1
1	1			1		1	1	

$L=X \otimes Y$

Рисунок 10 – Декомпозиції логічних векторів

Мета цієї процедури – знайти два вектори, які формують матрицю шляхом конкатенації бітів двох векторів, які потрібно визначити. Конкатенація двох одиниць утворює адресу 11, яка формує всі одиничні координати в матриці логічного вектора. В іншому випадку – два нулі в адресі функціональності формують усі нульові координати матриці логічного вектора. У загальному випадку завдання формується так: знайти два логічні вектори, які на основі декартової логіки бітів генерують інтегральний вектор у формі матриці. Існує два умови для розв'язання цього завдання: 1) Утворювальна функція відома. Алгоритмічна складність розв'язання задачі – $A = 2^2 \times 2^n$.

2) Якщо утворювальна функція невідома (рис. 11), то алгоритмічна складність розв'язання задачі становить: $A = 2^2 \times 2^2 \times 2^{n-2} = 2^4 \times 2^{n-2}$.

Це завдання вирішується повним перебором усіх можливих підстановок усіх 16 функцій від двох змінних, за винятком двох тривіальних функцій: 0000 (завжди 0) і 1111 (завжди 1).

Initial form for decomposition								
$\otimes=2^4 \times 2^{n-2}$								
00=0...1		1			1		1	1
01=0...1		1	1	1	1	1	1	1
10=0...1		1	1	1	1	1	1	1
11=1...0		1			1		1	1

$L=X \otimes Y$

Рисунок 11 – Декомпозиція логічного вектора на основі перебору

Може існувати двійковий вектор, який неможливо отримати шляхом декартової суперпозиції двох векторів. У такому випадку залишається лише один шлях декомпозиції логічного вектора – представлення логічного вектора у вигляді ДНФ за одиничними координатами вектора. Наприклад, для вектора 10010100 розклад у вигляді ДНФ матиме вигляд: $L = x_1 x_2 x_3 \vee x_1 x_1 x_2 \vee x_1 x_2 x_2$ (рис. 12).

Кожен елемент схеми тут представлений логічним вектором монотонної логічної функції (and, or) з базису Поста.

Тому немає труднощів у побудові логічного вектора цього фрагмента схеми, навіть якщо в ньому буде 2^{n-1} одиничних значень бітів. Монотонна функція визначається логічним вектором, який має лише один перехід з одиниці в нуль або з нуля в одиницю в таблиці істинності.

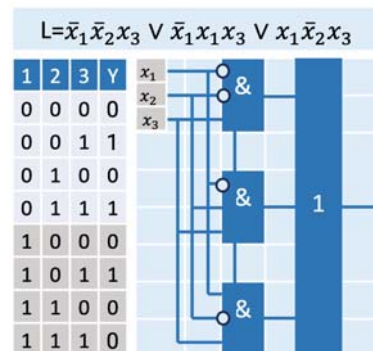


Рисунок 12 – Декомпозиції логічного вектора шляхом синтезу ДНФ

Розділити логічний вектор на два можна за допомогою схеми комутації. Мета цієї процедури – отримати схемну структуру для немонотонного логічного вектора, що складається з двовходових елементів. Це цілком передбачуваний алгоритм, який працює для будь-якого нерегулярного вектора (рис. 13). Недолік такої процедури полягає в тому, що в результаті виникає певна надлишковість у вигляді схеми комутації, де елементи мають більше ніж два входи. Ця схема обслуговує поділ вектора 11010101 на дві рівновеликі частини.

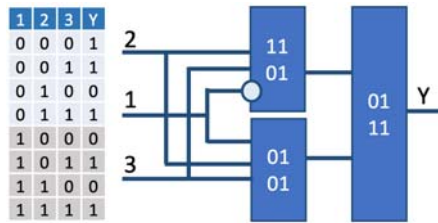


Рисунок 13 – Поділ логічного вектора на два з використанням комутатора

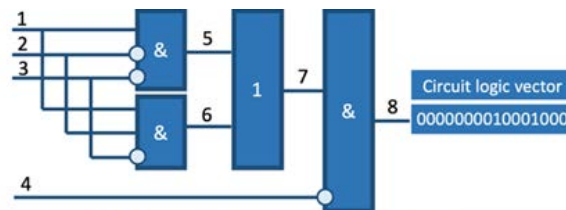
Моделювання логічного вектора схеми шляхом good-value simulation. Моделювання всіх 2^n вхідних наборів. Процедура є ефективною та технологічно дуже простою. Для обробки елементів схеми, представлених логічними векторами L , розміщеними в пам'яті, використовується єдина формула: $M_i = L[M(X_i)]$.

Для аналізу наведених структур даних схеми застосовуються read-write транзакції in-меморію computing над бітами логічних векторів. Щоб визначити вихід оброблюваного елемента, необхідно сформувати вхідний двійковий вектор $M(X)$

відповідно до станів координат вектора моделювання M , номери яких вказані в полі Inputs (рис. 14).

Поле NI визначає кількість входів у кожному елементі. Після цього вхідний двійковий вектор $M(X)$ розглядається як адреса бітової комірки відповідного логічного вектора з поля L -vectors, яка формує вихідний стан поточного елемента. Це моделювання дозволяє отримати всі стани ліній схеми на вичерпному тесті, включаючи лінію з виходом «вісім», на якій формується логічний вектор усієї схеми.

Моделювання логічного вектора схеми шляхом good-value simulation дозволяє отримати карту тестування всіх комбінацій вхідних несправностей (рис. 15) усього за три матричні операції [26]. Моделювання несправностей на основі схемної структури – на порядок складніше, ніж good-value simulation. Саме цим пояснюється той факт, що під час верифікації цифрових проєктів зазвичай використовується лише good-value simulation схеми.



The data structure of the fault-free simulation																								
M	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	Inputs: 1,2,3,4 outputs 8							
Simulation matrix on the complete test																L -vectors				Inputs				N
1	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1								
2	0	0	0	0	1	1	1	1	0	0	0	0	1	1	1	1								
3	0	0	1	1	0	0	1	1	0	0	1	1	0	0	1	1								
4	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1								
5	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	00001000	1	2	3		3		
6	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	00000010	1	2	3		3		
7	0	0	0	0	0	0	0	0	1	1	0	0	1	1	0	0	0111		5	6		2		
8	0	0	0	0	0	0	0	0	1	0	0	0	1	0	0	0	0010		7	4		2		

Рисунок 14 – Структура даних безпомилкового моделювання (fault-free simulation) схеми

L-matrix															Testing map																	
0	0	0	0	0	0	0	0	1	0	0	0	1	0	0	0		0000	0001	0010	0011	0100	0101	0110	0111	1000	1001	1010	1011	1100	1101	1110	1111
0								1								0000									1...				11..			
0								1								0001										1..0				11.0		
0								1								0010											1.0.				110.	
0								1								0011												1.00				1100
0								1								0100									1...				10..			
0								1								0101											1..0				10.0	
0								1								0110												1.0.				100.
0								1								0111													1.00			1000
1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1000	...	1	..1	..11		..11	..11	..11	0...	0..1	0..1	0..11	01..	01..1	011.	0111
0								1								1001	...	0					..10									
0								1								1010			..0.				..10.									
0								1								1011				..00				..100								
1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1100	...	1	..1	..11		..0.1	..01.	..011	0...	0..1	0..1	0..11	00..	00..1	001.	0011
0								1								1101	...	0					..0.0									
0								1								1110			..0.				..00.									
0								1								1111				..00				..000								

Рисунок 15 – Моделювання карти тестування за логічним вектором 0000000010001000 для схеми

Однак дедуктивне векторно-логічне моделювання несправностей [27] практично не поступається за обчислювальною складністю алгоритму good-value simulation. Тут представлено всі комбінації логічних несправностей входних змінних комбінаційної схеми, що перевіряються на вичерпному тесті. Це яскравий приклад prompt-комп'ютинг, коли вся діагностична інформація для верифікації проекту отримується на основі введення логічного вектора схеми. Складність алгоритмів моделювання є лінійною завдяки експоненційній надлишковості структур даних, представлених логічними векторами, таблицями істинності та матрицями декартової логіки.

Метод декартової суперпозиції логічних векторів. Метод моделювання справної роботи для отримання логічного вектора має один суттєвий недолік – це необхідність моделювання повного (вичерпного) тесту довжиною 2^n , де n – кількість входних змінних, для всіх елементів схеми. Більш оптимальним і мінімальним є метод, заснований на суперпозиції векторів за допомогою декартової логіки.

Рисунок 17 – Отримання логічного вектора схеми на основі декартової логіки

2. Виконуємо побудову логічного вектора кожного двовходового елемента за допомогою декартової логіки. 3. Моделюємо матрицю двох векторів за допомогою утворювального вектора довжиною чотири біти. Кожна таблиця для обробки елемента представлена: двома логічними векторами (вхідними), адресами бітів, утворювальним логічним вектором, які використовуються для виконання декартового добутку. 4. Видаляємо координати матриці, що суперечать вхідним наборам. Для елемента з виходом 7 залишаються лише координати головної діагоналі; решта вважаються суперечливими за адресами вхідних бітів. 5. Перевіряємо отриманий логічний вектор схеми за допомогою good-value simulation. 6. У результаті отримуємо логічний вектор схеми: 0000000010001000.

4 ЕКСПЕРИМЕНТИ

Алгоритм forward modeling (прямого моделювання): кожен логічний елемент розглядається як вузол, у якому виконується декартова логіка (декартовий добуток) завжди двох векторів, що подаються на входи цього елемента. Декартова логіка кожного елемента – це його логічний вектор більшої розмірності 2^{n+1} .

За один прохід по всіх елементах схеми формується логічний вектор, який надалі можна використати для побудови карти тестування. Розглянемо основні етапи алгоритму побудови логічного вектора схеми на основі декартової логіки на прикладі комбінаційної схеми (рис. 16).

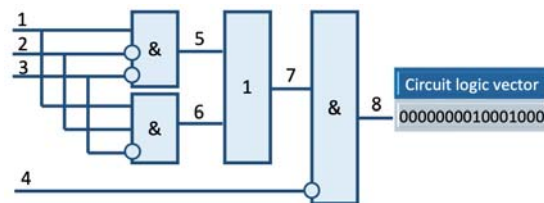


Рисунок 16 – Логічна схема зі зсодженими розгалуженнями

Моделювання логічного вектора схеми методом декартової суперпозиції векторів (рис. 17): 1. Умовно розбиваємо всі елементи на двовходові. Це спрощує процедуру моделювання та дозволяє уніфікувати обробку елементів.

Під час виконання декартової логіки враховуються особливості структури логічної схеми:

- 1) на елемент подані два зовнішніх входи – виконується декартова операція: $01 \text{ G } 01$;
- 2) на елемент подані три зовнішніх входи – послідовно виконуються дві декартові операції $(01 \text{ G } 01) \text{ G } 01$. На елемент подано понад три зовнішні входи, виконується рекурсія: $(01 \text{ G } 01) \text{ G } 01) \dots \text{ G } 01$;
- 3) якщо на елемент подано один зовнішній вхід і вихід іншого елемента – виконується операція: $01 \text{ G } Y$ або $X \text{ G } 01$;
- 4) якщо на елемент подано два виходи з інших елементів – виконується: $X \text{ G } Y$. Якщо на елемент подано понад два виходи з інших елементів – виконується рекурсія: $(X \text{ G } Y) \text{ G } Y) \dots \text{ G } Y$ (рис. 18);
- 5) обмеження на пряме моделювання векторів схеми пов'язані з нерегулярним характером деяких логічних векторів, які містять більше ніж дві змінні. У такому разі потрібно виконати розклад або декомпозицію початкового вектора на кілька векторів регулярних логічних функцій від двох змінних, після чого застосовується рекурсивна процедура прямого моделювання багатовходового елемента на основі декартових операцій двовходових елементів.

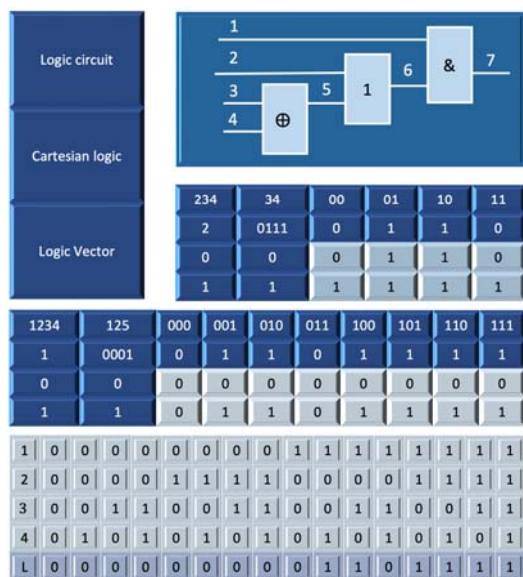


Рисунок 18 – Моделювання логічного вектора для ітеративної цифрової структури

Зворотнє моделювання (на відміну від forward modeling) передбачає обробку схеми від виходу до входів. Порядок обробки елементів: на кожному кроці вибирається той елемент, вихід якого з'єднаний із входом уже обробленого елемента. Обидва методи – forward і зворотнє моделювання – мають однакову обчислювальну складність. Вихідний результат – це таблиця істинності T за зовнішніми входами схеми разом із логічним вектором L і цифровими ідентифікаторами логічних входних змінних.

Хмарні сервіси modeling for simulation (MOSI) представлені такими компонентами: 1) синтез карти тестування за логічним вектором функціональності; 2) синтез логічного вектора цифрової комбінаційної схеми, заданої логічними векторами елементів; 3) моделювання справної поведінки логічної схеми на основі векторного представлення логічних елементів; 4) моделювання несправностей як адресів у логічній схемі на основі дедуктивних векторів логічних елементів; 5) пошук несправностей у логічній функціональності за її картою тестування; 6) графічний інтерфейс (GUI) для введення логічної схеми, де елементи представлені логічними векторами.

Інтерфейс застосунку GenLV. Вхідні дані: два логічні вектори X і Y однакової або різної довжини. Декартова логіка – це спосіб отримання матриці логічного вектора L шляхом декартового добутку бітів двох векторів за правилом логічної функції G : $L = X \otimes Y$, де $2^{n+m} = 2^n \times 2^m$. Утворювальний вектор G повинен мати довжину $G = 2^{n-2} = 4$.

1. Цифрові ідентифікатори логічних змінних: кількість визначається як двійковий логарифм довжини кожного з векторів: $\log_2 X$, $\log_2 Y$, $\log_2 G$, $\log_2 L$.

2. Інтерфейс вихідних даних декартової логіки: матриця M , що формує результат синтезу логічного вектора $L = X \otimes Y$, і містить усі три вектори (X , Y , L), ідентифікатори змінних і координати синтезованого вектора L у вигляді матриці, де $\otimes$ – будь-яка логіка, представлена вектором G .

3. Вихідний результат – це таблиця істинності T разом із логічним вектором L і цифровими ідентифікаторами логічних змінних (рис. 19).

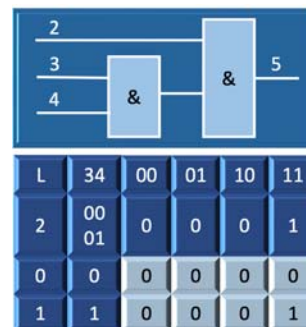


Рисунок 19 – Логічний вектор для фрагмента схеми без озгалужень

5 РЕЗУЛЬТАТИ

Запропоновано механізми побудови логічного вектора для цифрової комбінаційної схеми з метою подальшого синтезу карти тестування функціональності за допомогою лінійного алгоритму обробки векторно-матричних даних для розв'язання комбінаторних задач моделювання одиночних і кратних логічних несправностей [24–27]. Запропоновано апарат декартової логіки для розв'язання таких задач за допомогою алгоритмів лінійної обчислювальної складності. Декартова логіка, завдяки експоненційній надлишковості 2^{n+m} , є ефективним інструментом для задач моделювання, симуляції, тестування та діагностування. Наведено приклади практичного застосування векторної та декартової логіки для верифікації цифрових схем і функціональності. Представлено хмарні сервіси, що дозволяють вивчати та застосовувати ці механізми в моделюванні цифрових проєктів шляхом побудови тестів на основі моделювання несправностей. За простотою та оригінальністю ці механізми не мають аналогів.

На основі теорії векторної логіки розроблено механізми та програмне забезпечення для економного розв'язання задач design and test в архітектурі інтегрованого комп'ютерингу (рис. 20).

Хмарні сервіси cloud modeling for simulation (MOSI) представлені такими компонентами:

1. Синтез карти тестування за логічним вектором функціональності.

2. Синтез логічного вектора цифрової комбінаційної схеми, заданої логічними векторами її елементів.

MOSI – vector logic in-memory MOdeling for Simulation test and fault as address			
Design and test service	Complexity	Limitation	Novelty
GUI logic circuit	n^2 (line number)	Screen Area	Logic vector
Good value circuit simulation	n – linear	Screen Area	Logic vector
Circuit fault simulation	2^n (exponential)	Screen Area	fault as address
Modeling circuit logic vector	n^2 (line number)	Screen Area	Cartesian logic
Modeling logic testing map	n – linear	$2^{n=9}$ (input<10)	fault as address
Logic fault detection on test	n – linear	$2^{n=9}$ (input<10)	fault as address
The services are focused on processor-free in-memory computing via read-write transactions			
Service economics – low latency, low energy, low code, low time-to-market, high yield			

Рисунок 20 – Шість сервісів хмарної функціональності на основі векторної логіки

3. Моделювання справної поведінки логічної схеми на основі векторного представлення логічних елементів.

4. Моделювання несправностей як адрес у логічній схемі на основі дедуктивних векторів логічних елементів.

5. Діагностика функціональності з використанням її карти тестування.

6. Графічний інтерфейс користувача (GUI) для введення логічної схеми, в якій елементи представлені логічними векторами.

6 ОБГОВОРЕННЯ

Представлені сервіси, що показані на рисунку 20, мають одну важливу спільну рису – це логічний вектор, який використовується як модель для побудови механізмів Modeling for Simulation. Сервіси реалізовані мовами програмування Python і JavaScript з орієнтацією на архітектури майбутнього масового in-memory (quantum) комп'ютингу, що базується на read-write транзакціях без інструкцій центрального процесора. Така реалізація робить механізми Modeling for Simulation перспективними для EDA-ринку з точки зору часових та енергетичних витрат. Представлені сервіси орієнтовані на інженерне розуміння розв'язання задач design and test за технологією prompt-комп'ютингу, що є надзвичайно важливим для навчання студентів комп'ютерній інженерії у технічних вишах світу.

Особливості запропонованого методу декартової суперпозиції логічних векторів полягають у наступному: 1) необхідно використовувати адреси таблиці істинності або біти логічного вектора для мінімізації кількості координат, які формують вектор чергового оброблюваного елемента; 2) якщо логічна схема не має сходження розгалужень, то ця процедура мінімізації не застосовується; 3) недоліком методу є необхідність виконання процедури розбиття елемента на

двовходові для їх рекурсивної обробки з метою побудови логічного вектора схеми; 4) метод є технологічно зручним для написання програмного коду, якщо всі елементи схеми – двовходові; 5) якщо не розбивати елемент на двовходові підементи, то потрібно використовувати утворювальні логічні вектори довжиною більше ніж 4 біти.

Якщо специфікація функціональності задана логічним вектором – немає потреби синтезувати комбінаційну схему, а потім моделювати її, щоб знову отримати логічний вектор для верифікації. Замість цього логічний вектор специфікації слід безпосередньо використовувати в архітектурі in-memory комп'ютингу для тестування, верифікації, діагностування та експлуатації. In-memory vector logic комп'ютинг – це економічне рішення багатьох обчислювальних задач з погляду енерговитрат, часу та вартості.

Запропоновано математичний апарат декартової логіки, представлений у вигляді логічного вектора (матриці), що є результатом моделювання декартових логічних відношень між бітами логічних векторів або адресами таблиці істинності. Декартова логіка вирішує задачі: 1) Good-value modeling circuit logic vector – без алгоритму моделювання справної поведінки. 2) Fault modeling testing map of logic – без алгоритму моделювання несправностей.

Хмарний сервіс з GUI-інтерфейсом для введення схеми на основі логічних елементів у вигляді векторів дозволяє в 3–5 разів скоротити час створення моделі у порівнянні з описом схеми мовою HDL.

Моделювання логічного вектора схеми дає змогу вдвічі зменшити час моделювання несправностей у порівнянні з моделюванням схеми, поданої у вигляді логічних елементів. Саме моделювання займає лише 25% часу, необхідного для моделювання несправності схеми на основі логічного вектора.

Описано механізм good-value simulation цифрової схеми для синтезу логічного вектора з подальшим його використанням для побудови карти тестування несправностей на повному тесті лише за три матричні операції. Цей in-memory механізм є простим у реалізації, використовує лише read-write транзакції над бітами логічних векторів і не потребує інструкцій процесора.

ВИСНОВОК

Практичне значення дослідження – впровадження в EDA-ринку інженерного підходу до розв’язання задач design and test на основі використання простих моделей векторної логіки, орієнтованої на економічний in-memory комп’ютинг [19–20], що базується на read-write транзакціях і вільний від інструкцій процесора.

«Векторна логіка» представлена переважно у вигляді матрично-векторного формалізму для логічних функцій і значень істинності/хибності [21–23]. Проте це майже нічого не означає з практичної точки зору детермінованої комп’ютерної інженерії.

Новизна дослідження полягає у використанні векторної логіки [24–27] в архітектурі in-memory комп’ютингу на основі read-write транзакцій, що дозволяє зменшити споживання ресурсів у вигляді часу та енергії. Імплементація векторної логіки в in-memory комп’ютинг робить його масовим і енергоефективним, вільним від складних алгоритмів аналізу великих даних.

ЛІТЕРАТУРА

1. Integrated Circuit (IC) Open Library Architecture (OLA) : IEEE Std 1481-2009. – [Effective from 2010-03-11]. – New York : IEEE, 2010. – 658 p.
2. Esmaili E. Fanout-Based Reliability Model for SER Estimation in Combinational Circuits [Text] / E. Esmaili, Y. Sedaghat, A. Peiravi // IEEE Transactions on Circuits and Systems I: Regular Papers. – 2025. – Vol. 72, No. 1. – P. 228–240. – DOI: 10.1109/TCSI.2024.3458864.
3. Hwang M.-E. Slope Interconnect Effort: Gate-Interconnect Interdependent Delay Modeling for Early CMOS Circuit Simulation [Text] / M.-E. Hwang, S.-O. Jung, K. Roy // IEEE Transactions on Circuits and Systems I: Regular Papers. – 2009. – Vol. 56, No. 7. – P. 1428–1441. – DOI: 10.1109/TCSI.2008.2006217.
4. A Hybrid Model-Based Diagnosis Approach for Open-Switch Faults in PMSM Drives [Text] / [W. Huang, J. Du, W. Hua et al.] // IEEE Transactions on Power Electronics. – 2022. – Vol. 37, No. 4. – P. 3728–3732. – DOI: 10.1109/TPEL.2021.3123144.
5. Efficient Cartesian Genetic Programming-Based Automatic Synthesis Framework for Reversible Quantum-Flux-Parametron Logic Circuits [Text] / [R. Fu, R. Wille, N. Yoshikawa et al.] // IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems. – DOI: 10.1109/TCAD.2025.3546884.
6. Accuracy and Size Trade-off of a Cartesian Genetic Programming Flow for Logic Optimization [Text] / [A. Berndt et al.] // 2021 34th SBC/SBMicro/IEEE/ACM Symposium on Integrated Circuits and Systems Design (SBCCI), Campinas, Brazil, 2021 : proceedings. – 2021. – P. 1–6. – DOI: 10.1109/SBCCI53441.2021.9529968.
7. Utilizing the fuzzy logic algorithm for cartesian robot control system in conditions of mechanical damage transmission [Text] / [O. Dunaeva, S. Autso, K. Kudelina et al.] // IECON 2024 – 50th Annual Conference of the IEEE Industrial Electronics Society, Chicago, IL, USA, 2024 : proceedings. – 2024. – P. 1–5. – DOI: 10.1109/IECON55916.2024.10905445.
8. Khan G. M. Very Short-Term Load Forecasting Using Cartesian Genetic Programming Evolved Recurrent Neural Networks (CGPRNN) [Text] / G. M. Khan, F. Zafari, S. A. Mahmud // 2013 12th International Conference on Machine Learning and Applications, Miami, FL, USA, 2013 : proceedings. – 2013. – P. 152–155. – DOI: 10.1109/ICMLA.2013.181.
9. Inglese P. Side Channel and Fault Analyses on Memristor-Based Logic In-Memory [Text] / P. Inglese, E.-I. Vatajelu, G. Di Natale // IEEE Design & Test. – 2024. – Vol. 41, No. 3. – P. 29–35. – DOI: 10.1109/MDAT.2023.3324522.
10. A Novel Parallel In-Memory Logic Array Based on Programmable Diodes [Text] / [J. Ye, J. Zhu, J. Cao et al.] // IEEE Journal of the Electron Devices Society. – 2024. – Vol. 12. – P. 738–744. – DOI: 10.1109/JEDS.2024.3457021.
11. Chen X., Song T., Han Y. RRAM-based Analog In-Memory Computing: Invited Paper [Text] / X. Chen, T. Song, Y. Han // 2021 IEEE/ACM International Symposium on Nanoscale Architectures (NANOARCH), AB, Canada, 2021 : proceedings. – 2021. – P. 1–6. – DOI: 10.1109/NANOARCH53687.2021.9642235.
12. XNOR-BSNN: In-Memory Computing Model for Deep Binarized Spiking Neural Network [Text] / [V.-T. Nguyen, Q.-K. Trinh, R. Zhang et al.] // 2021 International Conference on High Performance Big Data and Intelligent Systems (HPBD&IS), Macau, China, 2021 : proceedings. – 2021. – P. 17–21. – DOI: 10.1109/HPBDIS53214.2021.9658467.
13. Truth Table Based Intelligent Computing [Text] / [V. Hahanov, E. Litvinova, Z. Davitadze et al.] // 2024 31st International Conference on Mixed Design of Integrated Circuits and System (MIXDES), Gdansk, Poland, 2024 : proceedings. – 2024. – P. 199–204. – DOI: 10.23919/MIXDES62605.2024.10614035.
14. Decoherence of Dynamically Manipulated Qubits [Text] / V. Privman, D. Solenov // 2006 Sixth IEEE Conference on Nanotechnology, Cincinnati, OH, USA, 2006 : proceedings. – 2006. – P. 842–845. – DOI: 10.1109/NANO.2006.247790.
15. Moiré Localization Induced Enhancement in the Decoherence Time of Interlayer Excitons in WSe₂-MoSe₂ Heterobilayers [Text] / [M. A. Durmuş, K. Demiralay, M. M. Khan et al.] // 2024 Conference on Lasers and Electro-Optics (CLEO), Charlotte, NC, USA, 2024 : proceedings. – 2024. – P. 1–2.
16. Intelligent Computing: The Latest Advances, Challenges, and Future [Text] / [Shiqiang Zhu, Ting Yu, Tao Xu et al.] // Intelligent Computing. – DOI: 10.34133/icomputing.0006.
17. Tarek A. A Proof Theoretic Exploration of Mathematical Induction in Computational Paradigm [Text] / A. Tarek, A. Alveed, A. Farhan // 2023 Congress in Computer Science, Computer Engineering, & Applied Computing (CSCE), Las Vegas, NV, USA, 2023 : proceedings. – 2023. – P. 963–972. – DOI: 10.1109/CSCE60160.2023.00162.
18. Kamide N. Natural Deduction with Explosion and Excluded Middle [Text] / N. Kamide // 2023 IEEE 53rd International Symposium on Multiple-Valued Logic (ISMVL), Matsue, Japan, 2023 : proceedings. – 2023. – P. 24–29. – DOI: 10.1109/ISMVL57333.2023.00016.
19. Behavioral Level Simulation Framework to Support Error-Aware CNN Training with In-Memory Computing [Text] / S.-H. Chang, C.-N. J. Liu, A. Küster // 2022 18th International Conference on Synthesis, Modeling, Analysis and Simulation Methods and Applications to Circuit Design (SMACD), Villasimius, Italy, 2022 : proceedings. – 2022. – P. 1–4. – DOI: 10.1109/SMACD55068.2022.9816307.

20. Framework Independent Modeling for SRAM-Based In-Memory Computing [Text] / [S. Li, N. Zhang, W. Zhang et al.] // 2024 2nd International Symposium of Electronic Design Automation (ISED), Xi'an, China, 2024 : proceedings. – 2024. – P. 777. – DOI: 10.1109/ISED62518.2024.10617799.
21. Westphal J. Logic as a Vector System [Text] / J. Westphal, J. Hardy // Journal of Logic and Computation. – 2005. – Vol. 15, No. 5. – P. 751–765. – DOI: 10.1093/logcom/exi040.
22. Mizraji E. Vector Logic: A Natural Algebraic Representation of the Fundamental Logical Gates [Text] / E. Mizraji // Journal of Logic and Computation. – 2008. – Vol. 18, No. 1. – P. 97–121. – DOI: 10.1093/logcom/exm057.
23. Mizraji E. Vector logic allows counterfactual virtualization by the square root of NOT [Text] / E. Mizraji // Logic Journal of the IGPL. – 2020. – Vol. 29, No. 5. – P. 859–870. – DOI: 10.1093/jigpal/jzaa026.
24. Vector Synthesis of Fault Testing Map For Logic [Text] / [V. Hahanov, W. Gharibi, S. Chumachenko et al.] // IAES International Journal of Robotics and Automation (IJRA). – 2024. – Vol. 13, No. 3. – P. 293–306. – DOI: 10.11591/ijra.v13i3.pp293-306.
25. Vector-Logical In-Memory Simulation of Faults as Truth Table Addresses [Text] / [V. Hahanov, E. Litvinova, H. Hahanova et al.] // 2024 IEEE East-West Design & Test Symposium (EWDTS), Yerevan, Armenia, 2024 : proceedings. – 2024. – P. 1–6. – DOI: 10.1109/EWDTS63723.2024.10873615.
26. Prompt-Testing of Logic [Text] / [V. Hahanov, D. Devadze, I. Hahanov et al.] // 2024 IEEE East-West Design & Test Symposium (EWDTS), Yerevan, Armenia, 2024 : proceedings. – 2024. – P. 1–5. – DOI: 10.1109/EWDTS63723.2024.10873774.
27. Faults-as-address simulation [Text] / [V. Hahanov, S. Chumachenko, E. Litvinova et al.] // IAES International Journal of Robotics and Automation. – 2024. – Vol. 13, No. 4. – P. 452–468. – DOI: 10.11591/ijra.v13i4.pp452-468.
- Received 14.07.2025.
Accepted 14.10.2025.

UDC 681.326

MECHANISMS OF SCHEMATIC MODELING BASED ON VECTOR LOGIC

Hahanov V. I. – Dr. Sc., Professor of the Design Automation Department, Kharkiv National University of Radio Electronics, Kharkiv, Ukraine.

Obrizan V. I. – PhD, Post-Doctoral Student of the Design Automation Department, Kharkiv National University of Radio Electronics, Kharkiv, Ukraine.

Rakhlis D. Y. – PhD, Associate Professor of the Design Automation Department, Kharkiv National University of Radio Electronics, Kharkiv, Ukraine.

Khakhanova H. V. – Dr. Sc., Professor of the Design Automation Department, Kharkiv National University of Radio Electronics, Kharkiv, Ukraine.

Hahanov I. V. – PhD, Assistant of the Design Automation Department, Kharkiv National University of Radio Electronics, Kharkiv, Ukraine.

Demchenko O. I. – Post-graduate student of the Design Automation Department, Kharkiv National University of Radio Electronics, Kharkiv, Ukraine.

Voronov A. O. – Post-graduate student of the Design Automation Department, Kharkiv National University of Radio Electronics, Kharkiv, Ukraine.

ABSTRACT

Relevance. This paper addresses issues relevant to the EDA market – reducing the cost and time of testing and verification of digital projects by synthesizing the logic vector of a digital circuit, which significantly simplifies the algorithms of good-value simulation and reduces the synthesis of the test map to three matrix operations.

Purpose. The aim of the study is to reduce the cost and time of testing and verification of digital projects by synthesizing the logic vector of a digital circuit, which allows for simplifying the algorithm for constructing the test map to three matrix operations.

Method. The synthesis of a logic vector for a combinational circuit is proposed for good-value and fault as modeling addresses within the architecture of prompt-intelligent in-memory computing. The logic vector is the most technological, compact, and exhaustive representation of the circuit for efficiently solving all design and testing tasks. Cartesian logic is proposed as an effective intelligent mechanism for solving combinatorial problems (modeling, simulation, testing, diagnostics) using algorithms of linear computational complexity, thanks to its exponential

redundancy 2^{n+m} . Mechanisms are proposed for constructing a logic vector of a process or phenomenon, function or structure, based on matrix structures of Cartesian logic for solving Modeling for Simulation tasks. An engineering method of direct parallel good-value modeling of the circuit is proposed, based on the use of logic vectors of elements and truth tables. Cartesian logic is a logic vector (matrix) as the result of modeling Cartesian logical relations between the bits of logic vectors or truth table addresses. Cartesian logic solves the following tasks: 1. Good-value modeling of a circuit logic vector without the need for a functional behavior modeling algorithm. 2. Fault modeling testing map of logic without a fault simulation algorithm.

Results. A mathematical apparatus of Cartesian logic is proposed, represented as a logic vector (matrix), which is the result of modeling Cartesian logical relations between the bits of logic vectors or addresses of the truth table. Based on the theory of vector logic, mechanisms and software have been developed for efficiently solving design and test tasks within the in-memory computing architecture.

Conclusions. The practical value of the study lies in addressing all design and test tasks using simple models of vector logic, oriented toward economical in-memory computing based on read-write transactions and free from processor instructions. It is proposed to use the specification logic vector within the in-memory computing architecture for testing, verification, diagnostics, and operation. In-memory vector logic computing is an economical solution to many computational problems in terms of energy, time, and cost efficiency. The novelty of the study lies in the use of vector logic within the in-memory computing architecture based on read-write transactions, which reduces resource consumption in terms of time and energy. Implementing vector logic in in-memory computing makes it scalable and energy-efficient, and frees it from complex big data analysis algorithms.

KEYWORDS: vector logic, logic vector, Cartesian logic, intelligent computing, test map, fault modeling, good-value modeling, in-memory modeling for simulation, prompt engineering.

REFERENCES

1. Integrated Circuit (IC) Open Library Architecture (OLA) : IEEE Std 1481-2009. [Effective from 2010-03-11]. New York, IEEE, 2010, 658 p.
2. Esmaili E., Sedaghat Y., Peiravi A. Fanout-Based Reliability Model for SER Estimation in Combinational Circuits [Text], *IEEE Transactions on Circuits and Systems I: Regular Papers*, 2025, Vol. 72, No. 1, pp. 228–240. DOI: 10.1109/TCSI.2024.3458864.
3. Hwang M.-E., Jung S.-O., Roy K. Slope Interconnect Effort: Gate-Interconnect Interdependent Delay Modeling for Early CMOS Circuit Simulation [Text], *IEEE Transactions on Circuits and Systems I: Regular Papers*, 2009, Vol. 56, No. 7, pp. 1428–1441. DOI: 10.1109/TCSI.2008.2006217.
4. Huang W., Du J., Hua W. et al. A Hybrid Model-Based Diagnosis Approach for Open-Switch Faults in PMSM Drives [Text], *IEEE Transactions on Power Electronics*, 2022, Vol. 37, No. 4, pp. 3728–3732. DOI: 10.1109/TPEL.2021.3123144.
5. Fu R., Wille R., Yoshikawa N. et al. Efficient Cartesian Genetic Programming-Based Automatic Synthesis Framework for Reversible Quantum-Flux-Parametron Logic Circuits [Text], *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*. DOI: 10.1109/TCAD.2025.3546884.
6. Berndt A. et al. Accuracy and Size Trade-off of a Cartesian Genetic Programming Flow for Logic Optimization [Text], 2021 34th SBC/SBMicro/IEEE/ACM Symposium on Integrated Circuits and Systems Design (SBCCI), Campinas, Brazil, 2021 : proceedings, 2021, pp. 1–6. DOI: 10.1109/SBCCI53441.2021.9529968.
7. Dunaeva O., Autou S., Kudelina K. et al. Utilizing the fuzzy logic algorithm for cartesian robot control system in conditions of mechanical damage transmission [Text], *IECON 2024 – 50th Annual Conference of the IEEE Industrial Electronics Society, Chicago, IL, USA, 2024 : proceedings*, 2024, pp. 1–5. DOI: 10.1109/IECON55916.2024.10905445.
8. Khan G. M., Zafari F., Mahmud S. A. Very Short-Term Load Forecasting Using Cartesian Genetic Programming Evolved Recurrent Neural Networks (CGPRNN) [Text], 2013 12th International Conference on Machine Learning and Applications, Miami, FL, USA, 2013 : proceedings, 2013, pp. 152–155. DOI: 10.1109/ICMLA.2013.181.
9. Inglese P., Vatajelu E.-I., Di Natale G. Side Channel and Fault Analyses on Memristor-Based Logic In-Memory [Text], *IEEE Design & Test*, 2024, Vol. 41, No. 3, pp. 29–35. DOI: 10.1109/MDAT.2023.3324522.
10. Ye J., Zhu J., Cao J. et al. A Novel Parallel In-Memory Logic Array Based on Programmable Diodes [Text], *IEEE Journal of the Electron Devices Society*, 2024, Vol. 12, pp. 738–744. – DOI: 10.1109/JEDS.2024.3457021.
11. Chen X., Song T., Han Y. RRAM-based Analog In-Memory Computing: Invited Paper [Text], 2021 IEEE/ACM International Symposium on Nanoscale Architectures (NANOARCH), AB, Canada, 2021 : proceedings, 2021, pp. 1–6. DOI: 10.1109/NANOARCH53687.2021.9642235.
12. Nguyen V.-T., Trinh Q.-K., Zhang R. et al. XNOR-BSNN: In-Memory Computing Model for Deep Binarized Spiking Neural Network [Text], 2021 International Conference on High Performance Big Data and Intelligent Systems (HPBD&IS), Macau, China, 2021 : proceedings, 2021, pp. 17–21. DOI: 10.1109/HPBDIS53214.2021.9658467.
13. Hahanov V., Litvinova E., Davitadze Z. et al. Truth Table Based Intelligent Computing [Text], 2024 31st International Conference on Mixed Design of Integrated Circuits and System (MIXDES), Gdansk, Poland, 2024 : proceedings, 2024, pp. 199–204. DOI: 10.23919/MIXDES62605.2024.10614035.
14. Privman V., Solenov D. Decoherence of Dynamically Manipulated Qubits [Text], 2006 Sixth IEEE Conference on Nanotechnology, Cincinnati, OH, USA, 2006 : proceedings, 2006, pp. 842–845. DOI: 10.1109/NANO.2006.247790.
15. Durmuş M. A., Demiralay K., Khan M. M. et al. Moiré Localization Induced Enhancement in the Decoherence Time of Interlayer Excitons in WSe₂-MoSe₂ Heterobilayers [Text], 2024 Conference on Lasers and Electro-Optics (CLEO), Charlotte, NC, USA, 2024 : proceedings, 2024, pp. 1–2.
16. Zhu Shiqiang, Yu Ting, Xu Tao et al. Intelligent Computing: The Latest Advances, Challenges, and Future [Text], *Intelligent Computing*. DOI: 10.34133/icomputing.0006.
17. Tarek A., Alveed A., Farhan A. A Proof Theoretic Exploration of Mathematical Induction in Computational Paradigm [Text], 2023 Congress in Computer Science, Computer Engineering, & Applied Computing (CSCE), Las Vegas, NV, USA, 2023 : proceedings, 2023, pp. 963–972. DOI: 10.1109/CSCE60160.2023.00162.
18. Kamide N. Natural Deduction with Explosion and Excluded Middle [Text], 2023 IEEE 53rd International Symposium on Multiple-Valued Logic (ISMVL), Matsue, Japan, 2023 : proceedings, 2023, pp. 24–29. DOI: 10.1109/ISMVL57333.2023.00016.
19. Chang S.-H., Liu C.-N. J., Küster A. Behavioral Level Simulation Framework to Support Error-Aware CNN Training with In-Memory Computing [Text], 2022 18th International Conference on Synthesis, Modeling, Analysis and Simulation Methods and Applications to Circuit Design (SMACD), Villasimius, Italy, 2022 : proceedings, 2022, pp. 1–4. DOI: 10.1109/SMACD55068.2022.9816307.
20. Li S., Zhang N., Zhang W. et al. Framework Independent Modeling for SRAM-Based In-Memory Computing [Text], 2024 2nd International Symposium of Electronic Design Automation (ISED), Xi'an, China, 2024 : proceedings, 2024, P. 777. DOI: 10.1109/ISED62518.2024.10617799.
21. Westphal J., Hardy J. Logic as a Vector System [Text] // *Journal of Logic and Computation*, 2005, Vol. 15, No. 5, pp. 751–765. DOI: 10.1093/logcom/exi040.
22. Mizraji E. Vector Logic: A Natural Algebraic Representation of the Fundamental Logical Gates [Text], *Journal of Logic and Computation*, 2008, Vol. 18, No. 1, pp. 97–121. DOI: 10.1093/logcom/exm057.
23. Mizraji E. Vector logic allows counterfactual virtualization by the square root of NOT [Text], *Logic Journal of the IGPL*, 2020, Vol. 29, No. 5, pp. 859–870. DOI: 10.1093/jigpal/jzaa026.
24. Hahanov V., Gharibi W., Chumachenko S. et al. Vector Synthesis of Fault Testing Map For Logic [Text], *IAES International Journal of Robotics and Automation (IJRA)*, 2024, Vol. 13, No. 3, pp. 293–306. DOI: 10.11591/ijra.v13i3.pp293-306.
25. Hahanov V., Litvinova E., Hahanova H. et al. Vector-Logical In-Memory Simulation of Faults as Truth Table Addresses [Text], 2024 IEEE East-West Design & Test Symposium (EWDTS), Yerevan, Armenia, 2024 : proceedings, 2024, pp. 1–6. DOI: 10.1109/EWDTS63723.2024.10873615.
26. [Hahanov V., Devadze D., Hahanov I. et al. Prompt-Testing of Logic [Text], 2024 IEEE East-West Design & Test Symposium (EWDTS), Yerevan, Armenia, 2024 : proceedings, 2024, pp. 1–5. DOI: 10.1109/EWDTS63723.2024.10873774.
27. Hahanov V., Chumachenko S., Litvinova E. et al. Faults-as-address simulation [Text], *IAES International Journal of Robotics and Automation*, 2024, Vol. 13, No. 4, pp. 452–468. DOI: 10.11591/ijra.v13i4.pp452-468.

УПРАВЛІННЯ У ТЕХНІЧНИХ СИСТЕМАХ

CONTROL IN TECHNICAL SYSTEMS

UDC 004.94

DEEP REINFORCEMENT LEARNING OPTIMIZATION METHODS FOR TRAFFIC LIGHTS AT SIGNALIZED INTERSECTIONS

Boyko N. I. – PhD, Associate Professor, Associate Professor of the Department of Artificial Intelligence Systems, Lviv Polytechnic National University, Lviv, Ukraine.

Mokryk Y. L. – Post-graduate student, Department of Artificial Intelligence Systems, Lviv Polytechnic National University, Lviv, Ukraine.

ABSTRACT

Context. Intersections are the most critical areas of a road network, where the largest number of collisions and the longest waiting times are observed. The development of optimal methods for traffic light control at signalized intersections is necessary for improving the flow of traffic at existing urban intersections, reducing the chance of traffic collisions, the time it takes to cross the intersection, and increasing the safety for drivers and pedestrians. Developing such an algorithm requires simulating and comparing the work of different approaches in a simulated environment.

Objective. The aim of this study is to develop an effective deep reinforcement-learning model aimed at optimizing traffic light control at intersections.

Method. A custom simulation environment is designed, which is compatible with the OpenAI Gym framework, and two types of algorithms are compared: Deep Q-Networks and Proximal Policy Optimization. The algorithms are tested on a range of scenarios, involving ones with continuous and discrete action spaces, where the set of actions the agent may take are represented either by different states of the traffic lights, or by the length of traffic light signal phases. During training, various hyperparameters were also tuned, and different reward metrics were considered for the models: average wait time and average queue length. The developed environment rewards the agent during training according to one of the metrics chosen, while also penalizing it for any traffic rule violations.

Results. A detailed analysis of the test results of deep Q network and Proximal Policy Optimization algorithms was provided. In general, the Proximal Policy Optimization algorithms show more consistent improvement during training, while deep Q network algorithms suffer more from the problem of catastrophic forgetting. Changing the reward function allows the algorithms to minimize different metrics during training. The developed simulation environment can be used in the future for testing other types of algorithms on the same task, and it is much less computationally expensive compared to existing solutions. The results underline the need to study other methods of traffic light control that may be integrated with real-life traffic light systems for a more optimal and safer traffic flow.

Conclusions. The study has provided a valuable comparison of different methods of traffic light control in a signalized urban intersection, tested different ways of rewarding models during training and reviewed the effects this has on the traffic flow. The developed environment was sufficiently simple for the purposes of the research, which is valuable due to the large computational requirements of the models themselves, but can be improved in the future by expanding it with more complex simulation features, such as various types of intersections that aren't urban, creating a road network of intersections that would all be connected to each other, adding pedestrian crossings, etc. Future work may be done to refine the simulation environment, expand the range of considered algorithms, consider the use of models for vehicle control in addition to traffic light control.

KEYWORDS: reinforcement learning, signalized intersection, traffic control, proximal policy optimization, deep Q-learning.

ABBREVIATIONS

DRL is a deep reinforcement learning;
SUMO is a Simulation of Urban MObility;
SOTL is a self-organizing traffic light;
DQN is a deep Q network;
VRCIS is a Vehicle-Road-Cloud Integration System;
CAV is a connected autonomous vehicle;
HDV is a Human-driven vehicle;
RL is a reinforcement learning;

AWT is an average waiting time;
AQL is an average queue length;
PPO is a Proximal Policy Optimization;
TRPO is a Trust Region Policy Optimization;
V2X is a Vehicle-to-everything;
V2I is a Vehicle-to-infrastructure.

NOMENCLATURE

S is a state space of the environment;

A is an action space;
 $P(s_{t+1} | s_t, a_t)$ is a transition probability function;
 $R(s_t, a_t)$ is a reward function;
 α, β, λ are weight coefficients;
 γ is a discount factor;
 a_t are actions;
 s_{t+1} are transitions;
 $r_t(\theta)$ is a probability ratio;
 $\hat{A}_t$ is an aestimated advantage function;
 $Q(s_t, A_t)$ represents the Q -values in the S_t state if action A_t was chosen;
 α parameter represents the learning rate;
 γ is a discount factor.

INTRODUCTION

As cities expand and the number of vehicles on the roads continues to climb, the challenge of managing traffic efficiently has become more urgent than ever. Ensuring that road networks are not only safe but also capable of handling increasing traffic volumes is a top priority for modern traffic management authorities. In this context, smart solutions for optimizing traffic flow – especially at intersections where congestion and accidents are most common – are gaining critical importance.

Statistics show that intersections are the most accident-prone areas of any road network, with a higher frequency of collisions compared to other road segments. [1] Notably, intersections without traffic lights experience more accidents than signalized ones. Traffic congestion in cities is an especially acute problem in the morning and evening when people commute to and from work. [2] Traffic signals help to regulate the flow and reduce the likelihood of human error, contributing to safer traffic environments.

Object of the study: the phases of traffic lights and their effect on the flow of traffic at signalized urban intersections.

Subject of the study: deep reinforcement learning-based algorithms for controlling the phases of traffic lights according to the state of the road network around an intersection.

The aim of this study is to develop a reinforcement learning agent that is capable of controlling the phases of traffic lights at a signalized intersection to optimize a chosen evaluation metric, such as the average wait time of the vehicles at the intersection, or their average queue length. This algorithm needs to also minimize the danger to the vehicles on the road by avoiding collisions between them. The goal is for these findings to contribute to the robustness of existing urban infrastructure, to make it safer and more efficient.

Tasks of the research:

- Analyze existing methods and algorithms for traffic light control optimization at a signalized intersection.
- Develop or select a suitable simulation environment that would simulate the flow of traffic, providing the agents with rewards and penalties.

- Create deep learning algorithms that are capable of controlling traffic lights in the simulated environment.
- Analyze the results and compare them with existing traffic light control methods.

1 PROBLEM STATEMENT

The primary goal of this study is to develop an effective DRL model aimed at optimizing traffic light control at intersections. To achieve this, a simulated environment will be created to replicate vehicular movement through intersections. Various DRL algorithms will be implemented and evaluated based on key performance metrics related to traffic efficiency and safety.

The rationale behind this research lies in the increasing demand for intelligent traffic control systems that can manage traffic flows dynamically and adaptively, particularly in large urban areas. Such systems have the potential to reduce congestion, improve traffic throughput, and minimize accidents resulting from human mistakes.

To fulfill this objective of designing a method for traffic light timing optimization, several tasks are required:

1. Analyze existing machine learning tools for traffic light optimization at intersections.
2. Develop an environment capable of simulating traffic flows at signalized intersections.
3. Design and implement a reinforcement learning algorithm to optimize traffic light timing within the simulation.
4. Tune the parameters of the optimization algorithm and compare its performance to that of traditional or existing control strategies.

Before a simulation environment or algorithm can be developed however, we need to discuss several issues that plague the development of such models:

- One major challenge is creating a realistic and scalable virtual environment for training agents that would control traffic lights using reinforcement learning. This environment must support extensive simulations and provide different performance metrics. A key difficulty is the automated creation of rare but critical traffic scenarios, as manually modeling all possible events is computationally infeasible.
- Building an algorithm that will be able to efficiently control the movement of traffic through an intersection, while avoiding any traffic collisions or other traffic rule violations, and at the same time minimize the waiting time of passengers. While traffic rules can be programmed, replicating nuanced human decision-making remains difficult.

- High-precision vehicle detection and movement prediction. Autonomous systems require highly accurate sensing and prediction capabilities, far exceeding those in other industries. These systems must function reliably under varying conditions, including poor weather and low visibility.

– Scalability and data demands: deep learning models used for traffic control are computationally intensive and require vast amounts of diverse training data. To ensure robustness across all possible driving scenarios, these models must be trained on extremely large, representative datasets – often measured in petabytes.

All of the above points make it clear that creating a hand-crafted algorithm for controlling traffic lights is challenging. This is the reason why machine learning approaches have often been considered for this task instead. However, as noted, the method of using neural networks or any other machine learning tool is also not without issues. Particularly when it comes to modelling the necessary training data and rewards.

The problem of optimizing traffic light control using reinforcement learning can be formally described as a Markov Decision Process (MDP) defined by the tuple:

$$M = (S, A, P, R, \gamma), \quad (1)$$

1. A state $s_t \in S$ includes:

- the current traffic light phases;
- the number of vehicles approaching the intersection;
- the number of vehicles that have already passed;
- queue lengths on each approach to the intersection.

2. A includes:

- Discrete: predefined combinations of traffic light phases (e.g., {*green-red*, *yellow-red*, ...});
- Continuous: the duration of the current traffic light phase $a_t \in [\underline{a}_{\min}, \underline{a}_{\max}] \subset \mathbb{R}$.

3. $P(s_{t+1} | s_t, a_t)$ defines the probability of moving to state s_{t+1} after taking action a_t in state s_t . This is modeled by the simulator.

4. $R(s_t, a_t) \in \mathbb{R}$ defined as:

$$R(s_t, a_t) = -\alpha AWT(s_t, a_t) - \beta \cdot AQL(s_t, a_t) - \lambda \times \text{penalties}(s_t, a_t). \quad (2)$$

The goal is to find an optimal policy $\pi^* : S \rightarrow A$ that maximizes the expected cumulative discounted reward:

$$\pi^* = \arg \max_{\pi} E_{\pi} \left[\sum_{t=0}^{\infty} \gamma^t R(s_t, a_t) \right]. \quad (3)$$

Methods for Finding the Optimal Policy:

– Deep Q-Network: approximates the Q-function $Q(s, a) \approx Q_{\theta}(s, a)$ using a neural network and updates it using the Bellman equation:

$$Q(s_t, a_t) \leftarrow (1 - \alpha) Q(s_t, a_t) + \alpha \left[R_t + \gamma \max_{a'} Q(s_{t+1}, a') \right]. \quad (4)$$

– PPO: a policy gradient method that optimizes the clipped surrogate objective:

$$L^{CLIP}(\theta) = E_t \left[\min \left(r_t(\theta) \hat{A}^t, \text{clip} \left(r_t(\theta), 1 - \epsilon, 1 + \epsilon \right) \hat{A}^t \right) \right]. \quad (5)$$

To solve this optimization problem, modern DRL algorithms such as DQN or PPO are employed. In particular, PPO is used for both discrete and continuous action spaces, while DQN is suitable for discrete cases.

2 LITERATURE REVIEW

Before choosing or designing our own simulation environment and reinforcement learning algorithm for training the traffic light control agents, an analysis of existing literature on the topic was performed.

This section has been split into two parts: one in which we analyze existing environments for running traffic simulations, comparing their advantages and disadvantages. In the second part, we looked at various other methods that have been proposed for vehicle and traffic light management using machine learning techniques.

One of the first tasks in building an effective algorithm for traffic light control is to build a virtual simulated environment for the training and validation of the developed algorithms. To begin with, it is worth considering existing solutions to this problem. Such an environment should allow for a large number of simulations to be run quickly to allow for training many iterations of the reinforcement learning algorithm. Also, the environment should provide certain metrics by which to evaluate and compare the performance of different algorithms.

There are a number of such environments that have been created for simulating road traffic. It is worth evaluating these environments according to several criteria: speed, ease of implementation of the machine learning algorithm in the environment, and simulation accuracy.

CARLA Simulator [3] is a simulator for training and validation of autonomous driving systems. CARLA provides open digital data (city plans, buildings, vehicles) that have been created for this purpose and can be freely used (Figure 1). The modeling platform supports flexible specification of sensor sets, environmental conditions, full control of static and dynamic agents, mapping, and much more. The disadvantages of the environment are the same as in the previous environment: large system requirements that can slow down training and difficulty in integrating with the machine learning model development environment. Many studies have been done using the CARLA Simulator related to autonomous traffic control [4, 5].

Flow [6] is a traffic simulation platform designed to evaluate driving algorithms for autonomous vehicles in various critical scenarios. It allows integration with machine learning software development environments. The main problem can be too much simulation complexity, which will have a bad effect on the length of training.

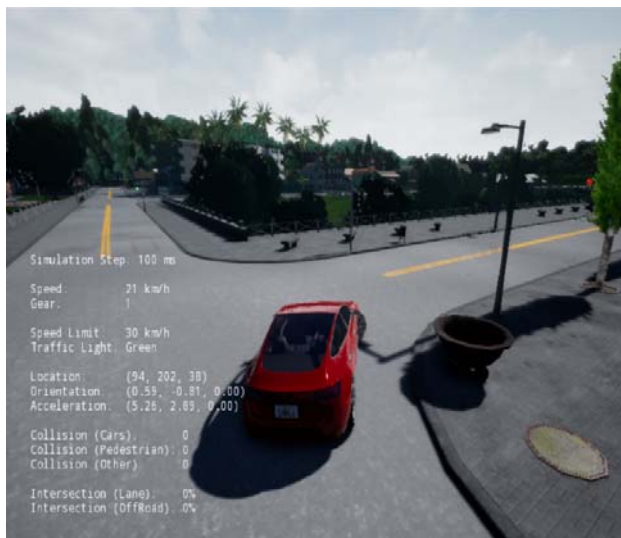


Figure 1 – Visualization of the CARLA Simulator environment

Eclipse SUMO [7] is another open source environment. This environment focuses on traffic simulation tasks with many vehicles to plan and predict different traffic situations (Figure 2). One of the key disadvantages of this environment is the high requirements for the system on which the simulation takes place. Training many agents will be much more difficult due to the high complexity of the simulator itself. In addition, it is difficult to connect this simulator to a Python-based machine learning model development environment, such as Jupyter.



Figure 2 – Visualization of the Eclipse SUMO environment

DriverGym [8] is an open-source reinforcement learning model training environment that allows you to train models to learn driving rules on real data (Figure 3). The algorithm can access semantic maps to control it. Other agents can be modeled from existing data or controlled by a special behavioral model previously trained on real-world data. This environment provides an extensive evaluation system with easily customizable metrics to evaluate trained algorithms.



Figure 3 – Visualization of the DriverGym environment

There are also some simulators that are customized for the OpenAI Gym platform [9], such as DeepCars, gym_trafficlight, and gym_graph_traffic. Such environments are usually very easy to integrate with the Python development environment and do not require large computing resources, but they are often too simple. For example, some of them only simulate movements in discrete steps, or have no way to evaluate the algorithm's performance using metrics.

After looking at existing applications for traffic simulation, we have decided to implement our own environment. The advantages of this approach include:

- The ability to easily connect it to a machine learning algorithm development environment.

- The ability to define any simulation parameters and any metrics to evaluate the performance of each algorithm

- The ability to fully customize performance and optimize the simulation for the tasks we need during development

Existing Optimization Approaches.

Aside from the simulation environment, we also need to choose the algorithms for training agents for the task of controlling traffic lights. There are a few things to consider here, including how these algorithms are known to perform at similar tasks, how well they scale and whether they support discrete or continuous actions spaces and observation spaces. Because of this, a few studies have been chosen for analysis to determine which algorithms are the most suitable for the task. The results of this analysis are summarized in Table 1. A further detailed description of each paper is given after the table.

Table 1 – Summary of existing research into traffic and traffic signal control with deep reinforcement learning

Name	Task	Tools	Year	Ref.
Deep Reinforcement Learning for Traffic Light Timing Optimization	Controlling traffic light timing	Q-learning, MaxPressure, Self-organizing traffic lights	2022	[10]
Deep Q-network-based traffic signal control models	Controlling traffic signals	Q-learning, Synchro model	2021	[11]
An Efficiency Enhancing Methodology for Multiple Autonomous Vehicles in an Urban Network Adopting Deep Reinforcement Learning	Controlling a network of autonomous vehicles at an unsignalized intersection	Reinforcement learning, Proximal policy optimization	2021	[1]
A Control Method with Reinforcement Learning for Urban Un-Signalized Intersection in Hybrid Traffic Environment	Controlling autonomous vehicles at an unsignalized intersection	Reinforcement learning, Proximal policy optimization, VRCIS, V2X, V2I	2022	[12]

In the first paper [10], the authors propose a traffic signal synchronization optimization method based on a dual dueling deep Q-network, MaxPressure, and SOTL, namely EP-D3QN, which controls traffic flows by dynamically adjusting the duration of traffic lights in a cycle and whether the phase is switched based on predefined rules and lane pressure. In EP-D3QN, each intersection corresponds to an agent, and the road entering the intersection is divided into grids, each grid stores the speed and position of the vehicle, thus forming an information matrix of the vehicle as well as the state of the agent. An agent action is a set of traffic light phases in a signal cycle that has four values. The effective duration of the traffic light is 0–60 seconds, and the switching of the traffic light phases depends on its pressing and the established rules. The agent's reward is the difference between the sum of the accumulated waiting times of all vehicles in two consecutive signal cycles. SUMO is used to model two traffic scenarios. Two types of evaluation indicators are selected, and four methods are compared to verify the effectiveness of EP-D3QN. The experimental results show that EP-D3QN has excellent performance in light and heavy traffic flow scenarios, which can reduce the waiting time and travel time of vehicles and improve the traffic efficiency of the intersection.

The authors of the next study [11] investigated a solution-finding method using artificial intelligence, which has recently gained widespread attention because it can solve complex problems such as traffic signal control.

In this study, two traffic signal control models were developed using reinforcement learning and microscopic simulation-based estimation for an isolated intersection and two coordinated intersections. To develop these models, a DQN was used, which is a promising reinforcement learning algorithm. The performance was evaluated by comparing the developed traffic signal control models in this study with a fixed time signal optimized by Synchro, which is a traffic signal optimization model. The evaluation showed that the developed traffic signal control model of the isolated intersection was validated, and the intersection coordination was better than that of the fixed-time signal control method.

The next paper [1] is devoted to advanced deep reinforcement learning. It considers the impact of leading autonomous vehicles on the urban network in mixed traffic. A set of hyperparameters is proposed to achieve better performance. First, a set of hyperparameters for reinforcement learning agents is selected. Second, experiments are conducted with an autonomous vehicle in an urban network with different levels of autonomous vehicle penetration. Third, the superiority of leading autonomous vehicles is evaluated using fully manual vehicles and leading manual vehicle experiments. Finally, proximal behavior optimization with a pruned objective is compared to proximal behavior optimization with an adaptive Kullback-Leibler penalty to test the superiority of the proposed hyperparameter. It is shown that the fully automated traffic increased the average speed by a factor of 1.27 compared to the entire experiment with a manual vehicle. The proposed method becomes significantly more efficient at higher penetration rates of autonomous vehicles. In addition, leading autonomous vehicles can help reduce congestion.

The last research we analyzed [12] is devoted to the method of coordinated control with optimization of proximal behavior in the VRCIS, where this control problem is formulated as a reinforcement learning problem. In this system, vehicles that are connected to the entire environment V2X were used to maintain communication between vehicles, and vehicle wireless communication technology can detect vehicles that use vehicle-to-infrastructure wireless communication V2I, thus achieving a cost-effective method. Then, the CAV defined in VRCIS was trained to adapt to HDV safely through intersections using reinforcement learning. A valid, scalable RL framework is developed that can convey topologies that may be dynamic traffic. Then, the state, action, and reward of RL are designed according to the problem of an urban intersection without traffic lights. Finally, how to deploy within the RL framework was described and several experiments were conducted with this framework to verify the effectiveness of the proposed method.

3 MATERIALS AND METHODS

As explained previously, a custom simulation environment has been developed. This will allow us to

better customize the simulation, make it more efficient and more easily integrate it with the tools for training RL agents.

Two algorithms have been chosen for training RL agents: Deep Q-Networks and Proximal Policy Optimization. This is because they have shown good results in the studies examined previously, and they are relatively lightweight in terms of computation, so many iterations can be trained without significant time requirements. Furthermore, such algorithms can be trained on a discrete or continuous action space (particularly PPO), which will be utilized in the experimental setup.

The simulated environment consists of an intersection where four roads meet. Each direction has a traffic light that allows traffic to move from that direction to any other direction. For example, the upper traffic light allows traffic from the lower street in any direction.

Traffic lights switch two at a time. At the beginning of the simulation, the left and right traffic lights are red. Accordingly, the upper and lower traffic lights are green. After switching, the upper and lower traffic lights first switch to yellow, then to red. After that, the left and right traffic lights switch to yellow and then to green. Figure 4 shows a simple visualization of what the environment would look like if we display all the vehicles and roads to scale.

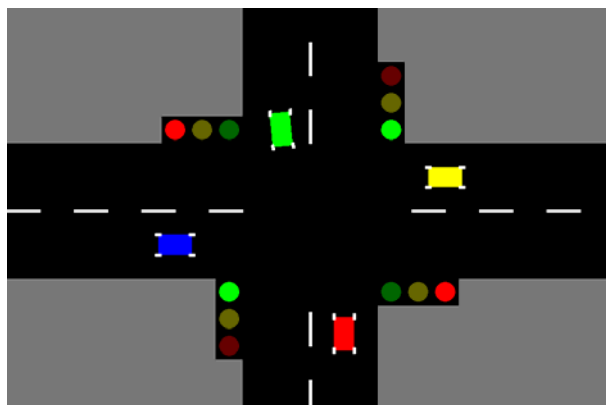


Figure 4 – Visualization of one of the scenarios in the custom simulated environment of a signalized intersection

Traffic light control is implemented using one of two approaches: discrete action spaces or continuous action spaces.

In the discrete case, traffic lights can occupy one of four predefined states, corresponding to signal phases for horizontal and vertical directions: green-red, yellow-red, red-yellow, and red-green. Every 1,000 simulation steps, the agent selects one of these possible states, determining the next phase of traffic flow. Figure 5 depicts how the discrete action space algorithm is applied during training for agents that control traffic light states.

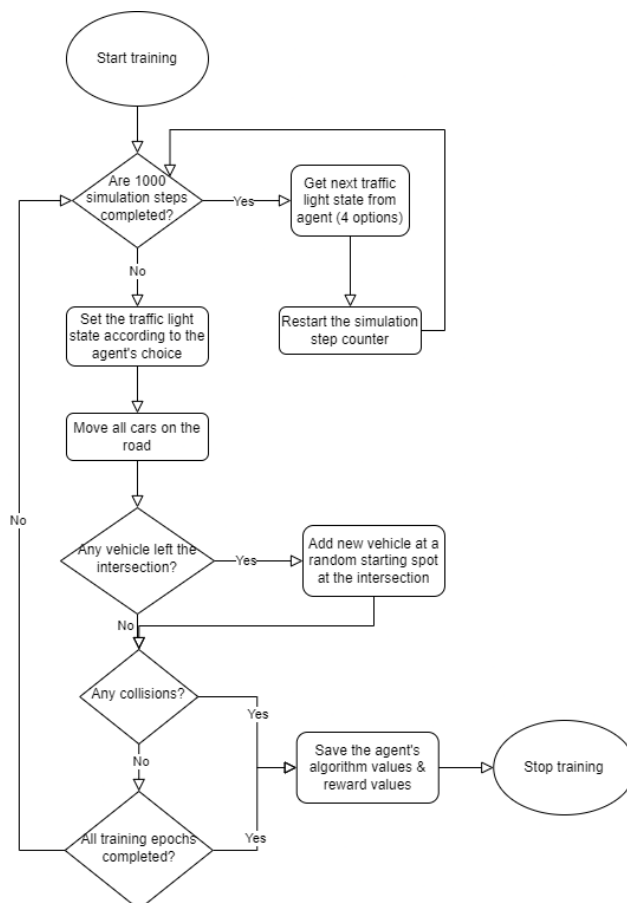


Figure 5 – Diagram of the discrete action space traffic light agent training algorithm

The continuous approach (Figure 6) allows for more dynamic control, where the agent selects the duration of each traffic light phase from a continuous range of values. Once a phase ends, the duration of the next phase is determined by the output of the reinforcement learning model, enabling finer adjustments based on traffic conditions.

The movement of cars is predetermined in this model, each car moves forward until it either stops at a red or yellow traffic light at the intersection or proceeds to cross at the intersection if the traffic light is green. While crossing, each car has a randomly determined direction that it needs to get to. The direction can be either straight ahead, left or right of the car's starting location. The car will either proceed straight or turn left or right depending on its current position, the state of the traffic light and the randomly chosen target direction.

The observation space for the traffic light control agent includes:

- The current state of the traffic signals.
- The proportion of vehicles currently on the road approaching the intersection from different directions.
- The proportion of vehicles that have already passed through the intersection from each direction.
- The proportion of vehicles still waiting in queues, i.e., the relative queue length on each side of the intersection.

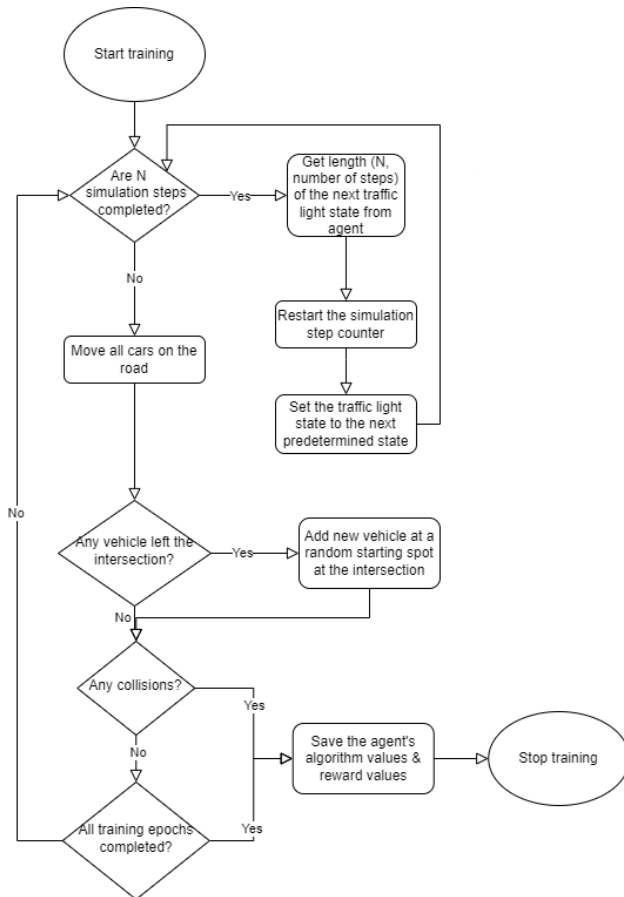


Figure 6 – Diagram of the continuous action space traffic light agent training algorithm

During training, it is possible to randomly vary traffic flow patterns by adjusting the distribution of vehicles across the intersection's lanes. Parameters control the rate at which non-agent-controlled vehicles enter the intersection from each direction. The rates can be thought of as a percentage chance that a new vehicle will come in from the top, bottom, left or right direction on the intersection. These rates are randomized at the start of each training episode, simulating asymmetric or uneven traffic conditions and adding variability to the environment. The agent is provided with these vehicle distribution ratios during training, helping it adapt its traffic light control to varying scenarios. Figure 7 shows the simple algorithm for the movement of randomized traffic on the road.

The reward function for traffic light control algorithms is based on two key performance metrics:

- Average Waiting Time (6) – the average number of simulation steps that each vehicle spends waiting at a red light before crossing the intersection.
- Average Queue Length (7) – the average number of vehicles queued at red lights from each direction at any given time.

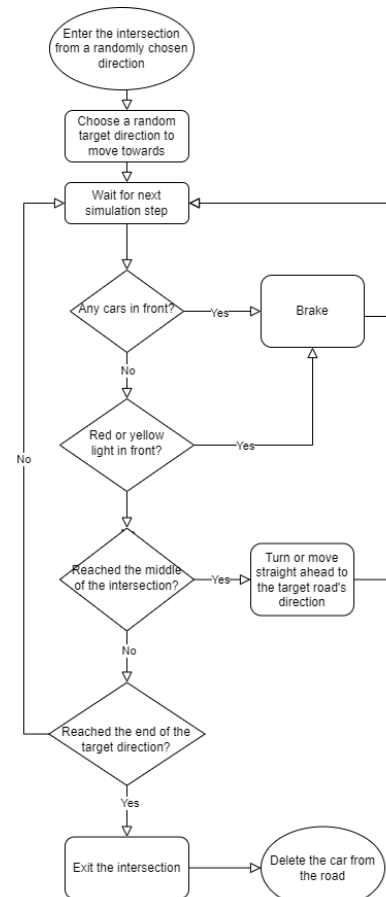


Figure 7 – Diagram of a car's movement pattern during a training episode

These metrics are evaluated using the following formulas:

$$AWT = \frac{1}{n} \sum_{i=1}^n w_i, \quad (6)$$

where n is the total number of vehicles that have been at the intersection during the training or validation episode, w_i is the time that the vehicle spend standing still due to a red light or other vehicles ahead.

$$AQL = \frac{1}{m} \sum_{i=1}^m q_i, \quad (7)$$

where m is the number of directions around the intersection (in this case, always 4) and q_i is the number of vehicles that are on that road at the current step of the simulation.

To summarize, the goal of the traffic light control agents is to minimize either the AWT or the AQL metric, chosen before training for each experiment. During validation, all algorithms can be tested on their performance with both metrics, to compare the impact of the choice of training metric on the algorithm's performance.

4 EXPERIMENTS

Both DQN and PPO models were trained on the discrete action space scenario, while PPO models were additionally trained on the continuous action space scenario. The models were trained using both AWT and AQL metrics for the reward function, one at a time. The number of vehicles at the intersection and the directions from which they came was randomly varied with a uniform distribution.

We consider DQN models built from two neural networks: A Q-network and a target network with the same architecture. The architecture of the networks consists of two hidden layers, 64 neurons each, and an output layer, where the number of neurons corresponds to the size of the action space. We consider DQN models with different values of the learning rate hyperparameter, DQN with a decreasing learning rate hyperparameter, and PPO with different values of the entropy coefficient hyperparameter. The networks used in the PPO algorithms have the same architecture as those used in the DQN models. All algorithms were trained for 8 thousand steps, the constant learning rate was set to 0.01 and 0.001, and the learning rate decreased from LR 1.0 or 0.1, which decreased linearly to zero throughout the training process. The values of the entropy coefficient of the PPO models were 0.1, 0.01, 0.001, and 0.0001.

At the core of RL is an agent's policy, a mapping from states to actions that guides behavior. The agent must balance exploration (trying new actions to discover rewards) and exploitation (choosing known actions that yield good outcomes), a trade-off fundamental to RL.

One key algorithm for learning such behavior is Q-learning, a model-free RL method. It does not require prior knowledge of the environment and learns by estimating Q-values, the expected cumulative reward of taking an action in each state and following the policy thereafter. These Q-values are stored in a Q-table, and the agent selects actions based on the highest Q-value for its current state. The Bellman equation (8) lies at the heart of Q-learning, since it allows updating the Q-values iteratively [13].

$$Q(S_t, A_t) = (1 - \alpha) Q(S_t, A_t) + \alpha (R_t + \gamma \max_a Q(S_{t+1}, a)). \quad (8)$$

However, Q-learning struggles with large or continuous state spaces, as storing and updating Q-values in a table becomes impractical. This challenge is addressed by DQN, which replaces the Q-table with a neural network that approximates the Q-function. This allows the agent to generalize from limited experience to unseen states, making it suitable for complex tasks.

DQN also uses experience replay, where past experiences are stored and randomly sampled during training. This reduces correlations between consecutive actions and stabilizes learning. Together, neural function approximation and experience replay enable DQN to scale Q-learning to real-world applications, such as traffic signal control, game playing, and robotics.

Policy gradient methods are fundamental among the recent breakthroughs in the use of deep neural networks to control agents in video games, 3D environment and even go and chess. However, getting optimal results with strategy gradient methods is challenging because they are very sensitive to the choice of hyperparameters, such as the size of the training step. Too small a step makes the learning extremely slow, and too large a step oversaturates the signal with noise, which can also cause a catastrophic drop in the learning results. Very often, such algorithms are inefficient in the use of the training set, meaning that they require millions or billions of steps to learn how to perform the simplest tasks.

The problem with strategy search methods is that they can converge slowly if the information they operate on is noisy. For example, this happens when trajectories are long and the variance of returns is large in episodic problems. As mentioned above, value function-based methods that rely on time differences can help in this case. In recent years, several actor-critic algorithms have been proposed that follow this idea and have been shown to perform well on a variety of tasks.

Recently, machine learning researchers have been looking for a way to avoid these problems with algorithms such as TRPO and ACER by limiting or otherwise optimizing the size of the strategy update step. Such methods have their own drawbacks: ACER is much more complicated to implement than PPO, requiring the addition of code to adjust the optimal strategy and the repetition buffer, and these algorithms have only a slight advantage over PPO. TRPO, while useful for problems with a continuous action space, is not easy to combine with algorithms that share parameters between the strategy and the loss function, such as PPO.

PPO is a family of model-free reinforcement learning algorithms developed by OpenAI in 2017. Such algorithms are examples of strategy gradient methods, meaning they search in the strategy space rather than assigning values to specific actions in the action space.

PPO algorithms have some advantages over TRPO methods, but their implementation is relatively simpler, and they generally use training samples more efficiently than TRPO [14]. These results are achieved by changing the objective function in PPO algorithms [15].

Various metrics can be used to evaluate vehicle and traffic signal control algorithms at an intersection, some of which have already been mentioned in the previous sections.

First, it is worth noting that the neural network loss function, which approximates the Q-function for a DQN model or the reward function for a PPO model, is not a significant metric. The problem with such an indicator is that the loss function in these models is constantly changing, because the target data for network training is not stable. Usually, the loss function in DQN and PPO models is constantly growing or fluctuating, despite the improvement of the model results in practice. That is why it is necessary to look at the results of the algorithm in a

validation environment, and at the metrics that can be obtained from it.

One such metric is the average reward function, which is defined by the training or validation environment. In this environment, which was described earlier in this paper, the reward function takes several important factors into account: traffic rules, distance to the target, and successful completion of the episode. The developed reward function gives high rewards when cars avoid colliding with each other, avoiding crossing into the oncoming lane, crossing the intersection during a red light at a traffic light, waiting during a green light, and crossing the edge of the road. The positive factors are getting closer to the goal and successfully reaching the right lane on the right side of the intersection. The fewer steps required to reach the goal, the higher the reward value.

Another factor that can be used as a metric is the AWT. In practice, this metric can be roughly estimated using the average length of episodes, because in the environment developed, a training or validation episode ends after successfully reaching the target intersection. Thus, if the episode is completed successfully, the average number of cars, seconds, or steps in such episodes can be taken into account to find out how optimally cars move through the intersection.

Another metric for evaluating traffic signals is the average queue length at an intersection. This metric measures the average queue length of vehicles that have not yet crossed the intersection, for example, because they are waiting for the traffic light to change, or because they are waiting for the car in front of them to move.

In addition to metrics of the quality of the algorithms in the validation environment, we can also consider metrics of the training of the algorithms themselves. While the loss function metric does not provide valuable results when training DQN models, it is also possible to obtain an estimate of the training time of the models depending on their hyperparameters, and the number of episodes or iterations or steps that are required to achieve satisfactory results. Such metrics can help compare algorithms in terms of training performance, which can be valuable for selecting the fastest or cheapest algorithm to train on much more complex environments.

5 RESULTS

Two different training approaches were used for the algorithms: one where the reward function depended on the AWT of vehicles at a red light, and another where it instead depended on the AQL of vehicles at a red light. Additionally, the PPO algorithm was trained using two distinct traffic light control methods: a discrete approach, where the algorithm selects the traffic light phase every 100 steps, and a continuous approach, where the algorithm determines the duration (in steps) of each traffic light phase. The DQN algorithm was trained only with a discrete action space, as it tends to yield better results in such settings and because most DQN implementations support only discrete action spaces.

The models were trained on these environments for tens of thousands of iterations, where each iteration represents a set number of steps. After every few iterations the parameters of the models were saved to a file, which allows comparing them individually for validation in case the final models have a larger error compared to the earlier ones. This is a common problem with reinforcement learning algorithms (especially DQN), known as catastrophic forgetting. This validation step also compares the performance of all best models together (chosen by the reward metric), to compare the AWT and AQL metrics achieved by each of them.

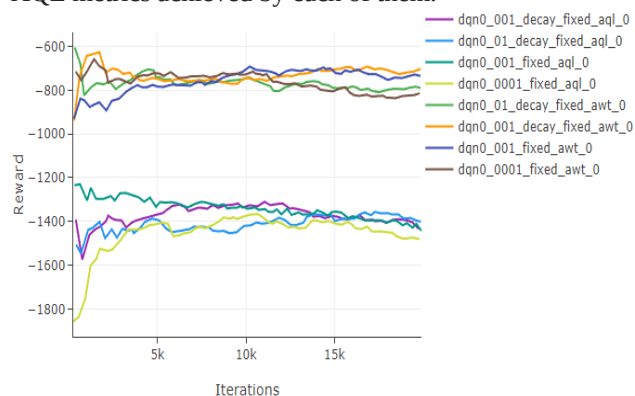


Figure 8 – Model training rewards with a discrete action space (DQN)

The first set of experiments was performed in a discrete action space where the DQN models could choose one of the available states for the traffic lights, for 20,000 iterations. As a result, Figure 8 shows that the DQN models, which were trained on the AWT metric, achieved higher rewards, but this value on its own is not meaningful, so it will be later compared during validation for a better understanding of the model's performance. Notably, the models with a learning rate of 0.001 achieved the best results, while the ones with smaller learning rates were more difficult to train. There is not a significant difference between models with a decaying learning rate and a constant learning rate. Although the ones with learning rate decay, perform slightly better on average, at least in this scenario.

The next graph demonstrates how PPO models with various learning rates performed on the same scenarios. It's important to note that on Figure 6 models with a fixed learning rate of 0.01 performed the best, regardless of whether they were using AQL or AWT as the reward metric. Once again, models trained on the AWT metric received slightly higher rewards, but this time the difference is much smaller than among the equivalent DQN models.

Finally, PPO models trained on a continuous action space show much faster improvement compared to the ones that were trained on a discrete action space. Allowing the algorithms to vary the length of the traffic light phases had a positive effect on training. This time, the learning rates did not have as significant of an impact on training, however algorithms trained on the AWT metric improve slightly less in 20,000 iterations than equivalent algorithms trained on the AQL metric.

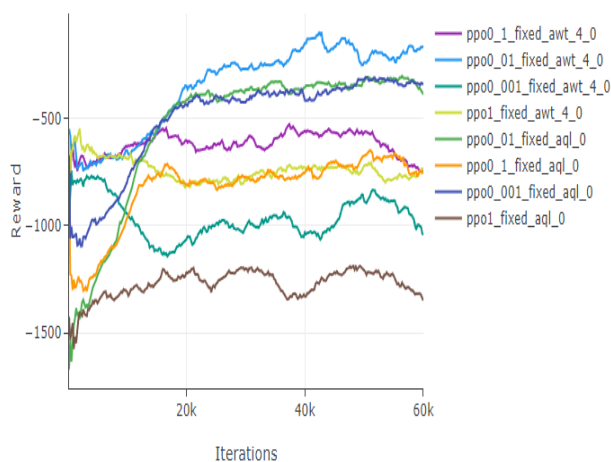


Figure 9 – Model training rewards with a discrete action space (PPO)

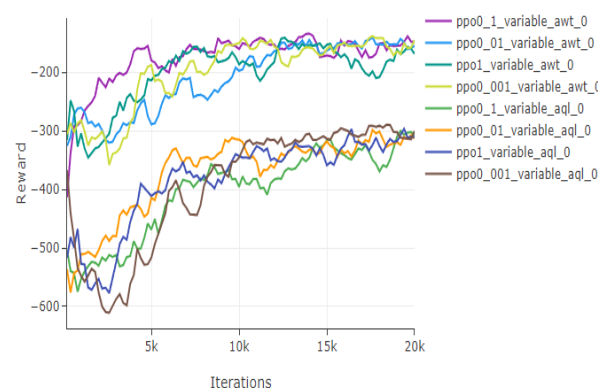


Figure 10 – Model training rewards with a continuous action space (PPO)

As illustrated in the graphs Fig. 9 and Fig. 10, PPO algorithms exhibit significantly better training performance in this scenario, with reward functions showing a more stable upward trend. This suggests that in the context of traffic light control, DQN algorithms suffer more from catastrophic forgetting, hindering their performance.

In Fig. 8 we can also observe how the chosen reward metric (AWT or AQL) impacted model performance during validation. The models were validated over 10,000 simulation steps, during which AWT and AQL were recorded. The models chosen here were the ones, which had the highest, reward value among the ones trained previously, since they were saved every few iterations, as mentioned before. The Fig. 11 shows that algorithms trained with different reward functions performed distinctly: AWT-based reward functions led to lower average waiting times, while AQL-based reward functions minimized queue lengths, as expected.

6 DISCUSSION

The study was aimed at developing an algorithm to optimize the control of traffic light phases at an intersection. The existing body of work in the subject area was analyzed, especially the papers that focused on deep

reinforcement learning methods, as well as the existing frameworks for traffic simulation.

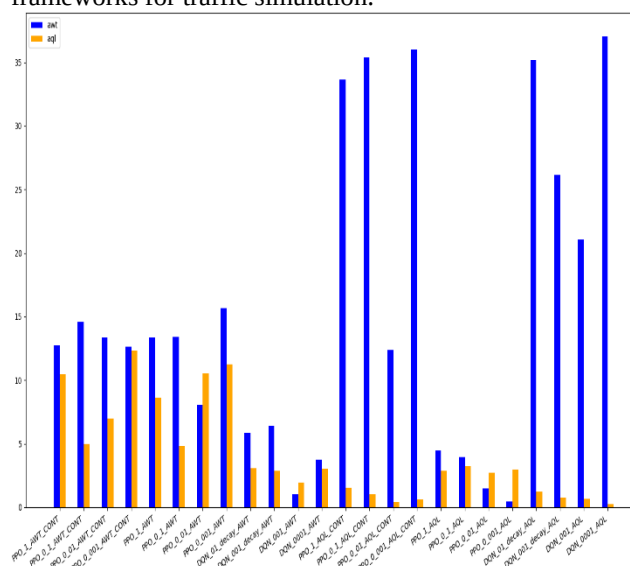


Figure 11 – Average queue length and average wait time compared for different traffic light deep learning algorithms with different reward functions

The research has underlined a few key challenges in creating an optimal traffic light control system, such as the development of a suitable training and testing environment, working with imprecise data about a vehicle's position and velocity, scaling the system to a larger real-life sized environment. To address these challenges, a novel simulation environment was created that allowed several approaches to be tested and compared.

The results indicate that the dependence of the reward function on the AWT and AQL metrics significantly impacted the results of algorithm training. Specifically, algorithms trained with the AWT metric in mind yielded lower AWT values for cars at red lights, and vice versa for those trained with the AQL metric. This highlights the importance of selecting appropriate reward metrics tailored to specific traffic management goals.

PPO algorithms trained on the traffic light control scenario demonstrated more stable reward function growth compared to DQN, indicating more consistent learning. This suggests that PPO is better suited for traffic light control tasks where training stability is critical. In this context, DQN algorithms showed more pronounced issues with catastrophic forgetting, further emphasizing the advantages of PPO for this application.

Furthermore, PPO algorithms that operated in a continuous action space, by controlling the duration of traffic light phases, performed significantly better than those using a discrete action space, where traffic light decisions occurred at fixed intervals (every 200 simulation frames). This ability to fine-tune phase durations in real-time contributes to improved traffic flow and highlights the benefits of using continuous control in traffic signal optimization.

Comparing these findings to prior studies, earlier research using DQN-based traffic light control showed improvements over fixed-time signal systems. However, this study builds on those results by demonstrating that PPO models using continuous action spaces can further improve performance in terms of AWT and queue length. The novel traffic simulation environment developed in this study enables detailed evaluation and comparison of such algorithms, providing a valuable tool for future research and potential real-world applications aimed at optimizing traffic signal operations.

CONCLUSIONS

During the training of both DQN and PPO models for traffic light control, a comparative analysis of AWT and AQL metrics was conducted when calculating reward functions. It was found that the reward function's dependence on AWT or AQL significantly influences training outcomes. Specifically, algorithms trained with AWT-based rewards achieved lower average waiting times, and vice versa for AQL-based rewards.

PPO algorithms consistently performed better in each traffic light control scenario, with more stable reward progression. Again, this confirms that DQN algorithms are more susceptible to catastrophic forgetting in this context. Moreover, PPO algorithms trained with a continuous action space, i.e., controlling the phase duration of traffic lights, outperformed all discrete control algorithms, which changed traffic light phases every 200 simulation frames.

These research results can be compared with previous studies on this topic. In the first section, we reviewed several works that utilized DQN or other Q-learning-based algorithms for traffic management at intersections. For instance, one study [12] examined DQN for traffic light control at a signalized intersection, developing a coordinated traffic signal control model that outperformed a static signal optimized using the Synchro method. In this thesis, similar DQN-based traffic light control methods were compared with PPO-based methods, including an analysis of discrete versus continuous action spaces in reinforcement learning agents for this scenario. It was found that the PPO algorithm developed by OpenAI [15], with appropriate hyperparameters, delivers superior results compared to DQN.

It is important to note that the findings presented here are limited in scope. Each algorithm was trained only once per set of hyperparameters over a fixed number of iterations. While the trends in reward functions and average episode lengths have been analyzed, these results are not definitive. To draw more robust conclusions, further studies involving more extensive experimentation and more complex scenarios are necessary.

Additionally, the simulated environment in this study focused solely on the intersection and did not account for traffic signs, pedestrians, or adjacent road segments. These simplifications represent key limitations, which future research can address.

The developed simulation environment offers a flexible platform for future research, allowing for the training and evaluation of various machine learning algorithms beyond those explored in this study

Scientific Novelty: The research contributes to the existing body of work on the subject of traffic control by creating a novel computationally cheap environment for testing several various traffic flow scenarios at a signalized intersection. It also gives insights into how the performance of PPO and DQN algorithms varies depending on the model's reward and the nature of the action space.

Practical Significance: The results of this research are relevant for transportation authorities, because they provide insight into the way signalized urban intersections can be optimized to allow for safer and more efficient traffic flow.

Prospects for Further Research: Future work could investigate alternative reinforcement learning approaches or entirely different methodologies to enhance model performance. Additionally, the simulated environment in this study focused solely on the intersection and did not account for traffic signs, pedestrians, or adjacent road segments. These simplifications represent key limitations, which future research can address.

Recommendations for Further Research: Expanding the scope to more realistic and challenging driving scenarios, including the presence of pedestrians, traffic signals, and complex road networks would provide valuable insights. Moreover, exploring enhanced versions of Q-learning or policy gradient methods tailored to these environments could lead to more effective and adaptable traffic control solutions.

ACKNOWLEDGEMENTS

The study was created within research topic "Methods and means of artificial intelligence to prevent the spread of tuberculosis in wartime" (№0124U000660), which is carried out at the Department of Artificial Intelligence Systems of the Institute of Computer Sciences and Information of technologies of the National University "Lviv Polytechnic".

REFERENCES

1. Tran Q.-D., Bae S.-H. An Efficiency Enhancing Methodology for Multiple Autonomous Vehicles in an Urban Network Adopting Deep Reinforcement Learning, *Appl. Sci.*, 2021, Vol. 11(4), P. 1514, Mode of access: <https://doi.org/10.3390/app11041514>
2. Vyklyuk Ya., Nevinskyi D., Boyko N. GeoCity – a New Dynamic-Spatial Model of Urban Ecosystem, *J. Geogr. Inst. Cvijic*, 2023, Vol. 73(2), P. 187–203, Mode of access: <https://doi.org/10.2298/IJGI2302187V>.
3. Team C. CARLA Simulator [Electronic resource]. Access mode: URL: <https://carla.org/> (access date: 01.04.2025). Title from the screen.
4. Gutiérrez-Moreno R., Barea R., López-Guillén E., Araluce J., Bergasa L.M. Reinforcement Learning-Based Autonomous Driving at Intersections in CARLA Simulator,

- Sensors, 2022, Vol. 22(21), P. 8373, Mode of access: <https://doi.org/10.3390/s22218373>.
5. Činčurak D., Grbić R., Vranješ M., Vranješ D. Autonomous Vehicle Control in CARLA Simulator Using Reinforcement Learning, *Int. Symp. ELMAR*, 2024, pp. 311–316, Mode of access: <https://doi.org/10.1109/ELMAR62909.2024.10694632>.
6. Wu C., Parvate K., Vinitsky E., Bayen A. M. Flow: A Modular Learning Framework for Mixed Autonomy Traffic, *IEEE Trans. Robot.*, 2022, Vol. 38.2, pp. 1270–1286.
7. Eclipse SUMO – Simulation of Urban Mobility [Electronic resource]. Access mode: URL: <https://eclipse.dev/sumo/> (access date: 01.04.2025). Title from the screen.
8. Kothari P., Perone C., Bergamini L., Alahi A., Ondruska P. DriverGym: Democratizing Reinforcement Learning for Autonomous Driving, *ArXiv*, 2021, Mode of access: <https://doi.org/10.48550/arXiv.2111.06889>.
9. OpenAI, Gym Beta [Electronic resource]. Access mode: URL: <https://openai.com/research/openai-gym-beta> (access date: 05.04.2025). Title from the screen.
10. Wang B., He Z., Sheng J., Chen Y. Deep Reinforcement Learning for Traffic Light Timing Optimization, *Processes* 2022, Vol. 10(11), P. 2458, Mode of access: <https://doi.org/10.3390/pr10112458>.
11. Park S., Han E., Park S., Jeong H., Yun I. Deep Q-network-based traffic signal control models, *PLOS ONE*, 2021, Vol. 16.9, Mode of access: <https://doi.org/10.1371/journal.pone.0256405>.
12. Shi Y., Liu Y., Qi Y., Han Q. A Control Method with Reinforcement Learning for Urban Un-Signalized Intersection in Hybrid Traffic Environment, *Sensors*, 2022, Vol. 22, Mode of access: <https://doi.org/doi:10.3390/s22030779>.
13. Sutton R. S., Barto A. G. Reinforcement Learning: an Introduction, second edition: MIT Press, 2018, P. 552.
14. Schulman J., Wolski F., Dhariwal P., Radford A., Klimov O. Proximal Policy Optimization Algorithms, *ArXiv*, 2017, Mode of access: <https://arxiv.org/abs/1707.06347>.
15. OpenAI, Proximal Policy Optimization [Electronic resource]. Access mode: URL: <https://openai.com/research/openai-baselines-ppo> (access date: 11.04.2025). Title from the screen.
- Received 14.07.2025.
Accepted 24.09.2025.

УДК 004.94

МЕТОДИ ОПТИМІЗАЦІЇ РОБОТИ СВІТЛОФОРІВ НА РЕГУЛЬОВАНИХ ПЕРЕХРЕСТЯХ ЗА ДОПОМОГОЮ ГЛИБИННОГО НАВЧАННЯ З ПІДКРІПЛЕННЯМ

Бойко Н. І. – канд. економ. наук, доцент, доцент кафедри Системи штучного інтелекту, Національний університет «Львівська політехніка», Львів, Україна.

Мокрик Я. Л. – аспірант кафедри Системи штучного інтелекту, Національний університет «Львівська політехніка», Львів, Україна.

АНОТАЦІЯ

Актуальність. Перехрестя є найбільш критичною ділянкою дорожньої мережі, де спостерігається найбільша кількість зіткнень та найдовший час очікування. Розробка оптимальних методів керування світлофорами на регульованих перехрестях необхідна для покращення руху транспортного потоку на існуючих міських перехрестях, зменшення ймовірності зіткнень, часу, необхідного для перетину перехрестя, та підвищення безпеки для водіїв і пішоходів. Розробка такого алгоритму вимагає моделювання руху транспорту та порівняння роботи різних підходів у змодельованому середовищі.

Мета роботи є розробка ефективної моделі глибинного навчання з підкріпленням (DRL), спрямованої на оптимізацію керування фазами світлофорів на перехрестях.

Метод. Розроблено власне симуляційне середовище, сумісне з OpenAI Gym, та проведено порівняння двох типів алгоритмів: глибинні Q-мережі та метод оптимізації близьких стратегій. Алгоритми протестовано на низці сценаріїв, включаючи сценарії з неперервним та дискретним просторами дій, де набір дій, які може виконати агент, представлений або різними станами світлофора, або тривалістю фаз сигналу світлофора. Під час навчання також налаштовувалися різні гіперпараметри та розглядалися різні метрики винагороди для моделей: середній час очікування та середня довжина черги. Розроблене середовище винагороджує агента під час навчання відповідно до однієї з обраних метрик, а також штрафує його за порушення правил дорожнього руху.

Результати. Проведено детальний аналіз результатів тестування алгоритмів DQN та PPO. Загалом, алгоритми PPO демонструють більш стабільне покращення під час навчання, тоді як алгоритми DQN більше страждають від проблеми катастрофічного забування. Зміна функції винагороди дозволяє алгоритмам мінімізувати різні метрики під час навчання. Розроблене моделююче середовище може бути використане в майбутньому для тестування інших типів алгоритмів на тій самій задачі, і воно є значно менш затратним в обчислювальному плані порівняно з існуючими рішеннями. Отримані результати підкреслюють необхідність дослідження інших методів керування світлофорами, які можуть бути інтегровані з реальними світлофорними системами для більш оптимального та безпечного руху транспортних потоків.

Висновки. Дослідження надало порівняння різних методів управління світлофорами на регульованому міському перехресті, протестувало різні способи заохочення моделей під час навчання та проаналізувало вплив, який це має на транспортний потік. Розроблене середовище було досить простим для цілей дослідження, що є цінним через великі обчислювальні вимоги самих моделей, але в майбутньому його можна вдосконалити, розширивши його більш складними функціями моделювання, такими як різні типи перехресть, які не є міськими, створення дорожньої мережі перехресть, які були б з'єднані між собою, додавання пішохідних переходів тощо. У майбутньому планується вдосконалити середовище моделювання, розширити спектр розглянутих алгоритмів, розглянути можливість використання моделей для керування транспортними засобами на додаток до керування світлофорами.

КЛЮЧОВІ СЛОВА: навчання з підкріпленням, сигналізовані перехрестя, керування транспортом, метод оптимізації близьких стратегій, глибинне Q-навчання.

ЛІТЕРАТУРА

1. Tran Q.-D. An Efficiency Enhancing Methodology for Multiple Autonomous Vehicles in an Urban Network Adopting Deep Reinforcement Learning./ Q.-D. Tran, S. -H. Bae // Appl. Sci. – 2021. – Vol. 11(4). – P. 1514, Mode of access: <https://doi.org/10.3390/app11041514>
2. Vykylyuk Ya. GeoCity – a New Dynamic-Spatial Model of Urban Ecosystem / Ya. Vykylyuk, D. Nevinskyi, N. Boyko // J. Geogr. Inst. Cvijic. – 2023. – Vol. 73(2). – P. 187–203, Mode of access: <https://doi.org/10.2298/IJGI2302187V>.
3. Team C. CARLA Simulator [Electronic resource]. Access mode: URL: <https://carla.org/> (access date: 01.04.2025). Title from the screen.
4. Reinforcement Learning-Based Autonomous Driving at Intersections in CARLA Simulator / R. Gutiérrez-Moreno, R. Barea, E. López-Guillén et al.] // Sensors. – 2022. – Vol. 22(21), – P. 8373, Mode of access: <https://doi.org/10.3390/s22218373>.
5. Autonomous Vehicle Control in CARLA Simulator Using Reinforcement Learning / [D. Činčurak, R. Grbić, M. Vranješ, D. Vranješ] // Int. Symp. ELMAR. – 2024. – P. 311–316, Mode of access: <https://doi.org/10.1109/ELMAR62909.2024.10694632>.
6. Flow: A Modular Learning Framework for Mixed Autonomy Traffic / C. Wu, K. Parvate, E. Vinitsky, A. M. Bayen // IEEE Trans. Robot. – 2022. – Vol. 38.2. – P. 1270–1286.
7. Eclipse SUMO – Simulation of Urban Mobility [Electronic resource]. Access mode: URL: <https://eclipse.dev/sumo/> (access date: 01.04.2025). Title from the screen.
8. DriverGym: Democratising Reinforcement Learning for Autonomous Driving / [P. Kothari, C. Perone, L. Bergamini et al.] // ArXiv. – 2021, Mode of access: <https://doi.org/10.48550/arXiv.2111.06889>.
9. OpenAI, Gym Beta [Electronic resource]. Access mode: URL: <https://openai.com/research/openai-gym-beta> (access date: 05.04.2025). Title from the screen.
10. Deep Reinforcement Learning for Traffic Light Timing Optimization / [B. Wang, Z. He, J. Sheng, Y. Chen] // Processes. – 2022. – Vol. 10(11). – P. 2458, Mode of access: <https://doi.org/10.3390/pr10112458>.
11. Deep Q-network-based traffic signal control models / [S. Park, E. Han, S. Park et al.] // PLOS ONE. – 2021. – Vol. 16.9, Mode of access: <https://doi.org/10.1371/journal.pone.0256405>
12. A Control Method with Reinforcement Learning for Urban Un-Signalized Intersection in Hybrid Traffic Environment / [Y. Shi, Y. Liu, Y. Qi, Q. Han] //Sensors. – 2022. – Vol. 22, Mode of access: <https://doi.org/doi:10.3390/s22030779>.
13. Sutton R. S. Reinforcement Learning: an Introduction, second edition / R. S. Sutton, A. G. Barto. – MIT Press, 2018. – P. 552.
14. Proximal Policy Optimization Algorithms / [J. Schulman, F. Wolski, P. Dhariwal et al.] // ArXiv. – 2017, Mode of access: <https://arxiv.org/abs/1707.06347>.
15. OpenAI, Proximal Policy Optimization [Electronic resource]. Access mode: URL: <https://openai.com/research/openai-baselines-ppo> (access date: 11.04.2025). Title from the screen.

Наукове видання

**Радіoeлектроніка,
інформатика,
управління**

№ 4/2025

Науковий журнал

Головний редактор – д-р техн. наук С. О. Субботін

Заст. головного редактора – д-р техн. наук Д. М. Піза

Комп'ютерне моделювання та верстання
Редактор англійських текстів

С. В. Зуб
С. О. Субботін

Оригінал-макет підготовлено у редакційно-видавничому відділі НУ «Запорізька політехніка»

Реєстрація суб'єкта у сфері друкованих медіа:
Рішення Національної ради України з питань телебачення і радіомовлення № 3040 від 07.11.2024 року
Ідентифікатор медіа: R30-05582

*Підписано до друку 02.12.2025. Формат 60×84/8.
Папір офс. Різогр. друк. Ум. друк. арк. 28,48.
Тираж 300 прим. Зам. № 1070.*

69063, м. Запоріжжя, НУ «Запорізька політехніка», друкарня, вул. Жуковського, 64

Свідоцтво суб'єкта видавничої справи
ДК № 6952 від 22.10.2019.