



Національний університет «Запорізька політехніка»

Радіоелектроніка, інформатика, управління

Науковий журнал

Виходить чотири рази на рік

№ 1(72) 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

National University Zaporizhzhia Polytechnic

Radio Electronics, Computer Science, Control

The scientific journal

Published four times per year

№ 1(72) 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 посилатися на повні тексти своїх статей. Журнал дозволяє повторне використання його вмісту у відповідності Creative Commons ліцензією 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>.

Журнал розповсюджується за Каталогом періодичних видань України (передплатний індекс – 22914).

Тематика журналу: телекомунікації та радіоелектроніка, програмна 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;

Бодянский 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 № 7 від 25.02.2025.

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 Льменау, 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 Олеса Гончара, Україна;

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

Камеяма Мічітака – доктор наук, професор, професор факультету науки та інженерії, Університет Ішіномакі Сеншу, 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;

Марковска-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 Льменау, 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'язку, 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. Іллі Чавчавадзе, Тбілісі, Грузія;

Уреутью 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.

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 is distributed by the Catalogue of Ukrainian periodicals (the catalog number is 22914).

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 Oliinyk – 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 Tabunshchyk – 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 Ursufiu – 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 № 7 dated 25.02.2025.

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

РАДІОЕЛЕКТРОНІКА ТА ТЕЛЕКОМУНІКАЦІЇ.....	6
<i>Atamanskiy D. V., Vasylyshyn V. I., Klymchenko V. Y., Stovba R. L., Prokopenko L. V.</i> THE RESERVES FOR IMPROVING THE EFFICIENCY OF RADAR MOVING TARGET INDICATION SYSTEM WITH BURST-TO-BURST PROBING PULSE REPETITION FREQUENCY STAGGER.....	6
<i>Hryzo A. A., Kostyria O. O., Fedorov A. V., Lukianchykov A. A., Biernik Ye. V.</i> ASSESSMENT OF THE QUALITY OF DETECTION OF A RADAR SIGNAL WITH NONLINEAR FREQUENCY MODULATION IN THE PRESENCE OF A NON-STATIONARY INTERFERING BACKGROUND.....	18
МАТЕМАТИЧНЕ ТА КОМП'ЮТЕРНЕ МОДЕЛЮВАННЯ.....	30
<i>Ausheva N. M., Sydorenko Iu. V., Kaleniuk O. S., Kardashov O. V., Horodetskyi M. V.</i> IMPLICIT CURVES AND SURFACES MODELING WITH PSEUDO-GAUSSIAN INTERPOLATION.....	30
<i>Gorodnov V. P., Druzhynin V. S.</i> THE STATES' FINAL PROBABILITIES ANALYTICAL DESCRIPTION IN AN INCOMPLETELY ACCESSIBLE QUEUEING SYSTEM WITH REFUSALS AND WITH INPUT FLOW OF REQUIREMENTS' GROUPS.....	38
<i>Pysarchuk O. O., Tuhanskykh O. A., Baran D. R.</i> METHOD FOR DETERMINING THE STRUCTURE OF NONLINEAR MODELS FOR TIME SERIES PROCESSING...53	53
<i>Фурсенко О. К., Чернопол Н. М.</i> МАТЕМАТИЧНЕ МОДЕЛЮВАННЯ БОЙОВИХ ДІЙ З МОЖЛИВІСТЮ ПЕРЕРОЗПОДІЛУ БОЙОВИХ РЕСУРСІВ МІЖ ДІЛЯНКАМИ ЗІТКНЕННЯ ТА РОЗПОДІЛУ РЕЗЕРВІВ.....	63
НЕЙРОІНФОРМАТИКА ТА ІНТЕЛЕКТУАЛЬНІ СИСТЕМИ.....	75
<i>Didenko Andrii Y., Didenko Artem Y., Subbotin S. A.</i> LIGHTWEIGHT MULTI-SCALE CONVOLUTIONAL TRANSFORMER FOR AIRCRAFT FAULT DIAGNOSIS USING VIBRATION ANALYSIS.....	75
<i>Molchanova M. O., Didur V. O., Mazurets O. V.</i> APPROACH TO DATA DIMENSIONALITY REDUCTION AND DEFECT CLASSIFICATION BASED ON VIBRATION ANALYSIS FOR MAINTENANCE OF ROTATING MACHINERY.....	84
<i>Prykhodko S. B., Trukhov A. S.</i> KEYSTROKE DYNAMICS RECOGNITION USING NINE-VARIATE PREDICTION ELLIPSOID FOR NORMALIZED DATA.....	96
<i>Sobko O. V., Dydo R. A., Mazurets O. V.</i> METHOD OF NEURAL NETWORK DETECTION OF DEFECTS BASED ON THE ANALYSIS OF ROTATING MACHINES VIBRATIONS.....	106
<i>Subbotin S. A., Bechhoefer E.</i> DATA-DRIVEN DIAGNOSTIC MODEL BUILDING FOR HELICOPTER GEAR HEALTH AND USAGE MONITORING.....	116
<i>Pukach A. I., Teslyuk V. M.</i> METHOD OF FORMING MULTIFACTOR PORTRAITS OF THE SUBJECTS SUPPORTING SOFTWARE COMPLEXES, USING A MULTILAYER PERCEPTRON.....	130
<i>Zalutska O. O., Hladun O. V., Mazurets O. V.</i> METHOD OF PREVENTING FAILURES OF ROTATING MACHINES BY VIBRATION ANALYSIS USING MACHINE LEARNING TECHNIQUES.....	142
<i>Zbezhkhovska U. R.</i> DEEPPAKE AUDIO DETECTION USING YOLOV8 WITH MEL-SPECTROGRAM ANALYSIS: A CROSS-DATASET EVALUATION.....	153
ПРОГРЕСИВНІ ІНФОРМАЦІЙНІ ТЕХНОЛОГІЇ.....	164
<i>Akhmetshina L. G., Yegorov A. A., Fomin A. A.</i> SEGMENTATION OF LOW-CONTRAST IMAGES IN THE BASIS OF EIGEN SUBSPACES OF TYPE-2 FUZZY MEMBERSHIP FUNCTIONS.....	164
<i>Касерес А., Глоба Л. С.</i> АНАЛІЗ ПІДХОДІВ ДОСТУПУ ДО ДАНИХ У МУЛЬТИ-КЛАУД СЕРЕДОВИЩІ.....	175
<i>Свінцицька О. М., Пулеко І. В., Граф М. С., Петросян Р. В.</i> МЕТОД USE CASE В УПРАВЛІННІ ІТ-ПРОЄКТОМ НА ОСНОВІ МЕТОДОЛОГІЇ AGILE.....	186
<i>Шпортко О. В., Бомба А. Я.</i> ЗАСТОСУВАННЯ БІНАРНОГО ДЕРЕВА ПОШУКУ З ФІКСОВАНОЮ ВИСОТОЮ ДЛЯ ПРИСКОРЕННЯ ОБРОБКИ ОДНОВИМІРНИХ МАСИВІВ.....	199
УПРАВЛІННЯ У ТЕХНІЧНИХ СИСТЕМАХ.....	209
<i>Holinko O. V., Alekseev M. O., Holinko V. I., Zabelina V. A.</i> APPLICATION OF SINGULAR SPECTRAL ANALYSIS IN CONTROL SYSTEMS OF TECHNOLOGICAL PROCESSES AND EXPLOSION SAFETY CONTROL OF FACILITIES.....	209

CONTENTS

RADIO ELECTRONICS AND TELECOMMUNICATIONS.....	6
<i>Atamanskiy D. V., Vasylyshyn V. I., Klymchenko V. Y., Stovba R. L., Prokopenko L. V.</i>	
THE RESERVES FOR IMPROVING THE EFFICIENCY OF RADAR MOVING TARGET INDICATION SYSTEM WITH BURST-TO-BURST PROBING PULSE REPETITION FREQUENCY STAGGER.....	6
<i>Hryzo A. A., Kostyria O. O., Fedorov A. V., Lukianchykov A. A., Biernik Ye. V.</i>	
ASSESSMENT OF THE QUALITY OF DETECTION OF A RADAR SIGNAL WITH NONLINEAR FREQUENCY MODULATION IN THE PRESENCE OF A NON-STATIONARY INTERFERING BACKGROUND.....	18
MATHEMATICAL AND COMPUTER MODELING.....	30
<i>Ausheva N. M., Sydorenko Iu. V., Kaleniuk O. S., Kardashov O. V., Horodetskyi M. V.</i>	
IMPLICIT CURVES AND SURFACES MODELING WITH PSEUDO-GAUSSIAN INTERPOLATION.....	30
<i>Gorodnov V. P., Druzhynin V. S.</i>	
THE STATES' FINAL PROBABILITIES ANALYTICAL DESCRIPTION IN AN INCOMPLETELY ACCESSIBLE QUEUING SYSTEM WITH REFUSALS AND WITH INPUT FLOW OF REQUIREMENTS' GROUPS.....	38
<i>Pysarchuk O. O., Tuhanskykh O. A., Baran D. R.</i>	
METHOD FOR DETERMINING THE STRUCTURE OF NONLINEAR MODELS FOR TIME SERIES PROCESSING...53	
<i>Fursenko O. K., Chernovol N. M.</i>	
MATHEMATICAL MODELLING OF COMBAT OPERATIONS WITH THE POSSIBILITY OF REDISTRIBUTING COMBAT RESOURCES BETWEEN THE AREAS OF CONTACT AND DISTRIBUTING RESERVES.....	63
NEUROINFORMATICS AND INTELLIGENT SYSTEMS.....	75
<i>Didenko Andrii Y., Didenko Artem Y., Subbotin S. A.</i>	
LIGHTWEIGHT MULTI-SCALE CONVOLUTIONAL TRANSFORMER FOR AIRCRAFT FAULT DIAGNOSIS USING VIBRATION ANALYSIS.....	75
<i>Molchanova M. O., Didur V. O., Mazurets O. V.</i>	
APPROACH TO DATA DIMENSIONALITY REDUCTION AND DEFECT CLASSIFICATION BASED ON VIBRATION ANALYSIS FOR MAINTENANCE OF ROTATING MACHINERY.....	84
<i>Prykhodko S. B., Trukhov A. S.</i>	
KEYSTROKE DYNAMICS RECOGNITION USING NINE-VARIATE PREDICTION ELLIPSOID FOR NORMALIZED DATA.....	96
<i>Sobko O. V., Dydo R. A., Mazurets O. V.</i>	
METHOD OF NEURAL NETWORK DETECTION OF DEFECTS BASED ON THE ANALYSIS OF ROTATING MACHINES VIBRATIONS.....	106
<i>Subbotin S. A., Bechhoefer E.</i>	
DATA-DRIVEN DIAGNOSTIC MODEL BUILDING FOR HELICOPTER GEAR HEALTH AND USAGE MONITORING.....	116
<i>Pukach A. I., Teslyuk V. M.</i>	
METHOD OF FORMING MULTIFACTOR PORTRAITS OF THE SUBJECTS SUPPORTING SOFTWARE COMPLEXES, USING A MULTILAYER PERCEPTRON.....	130
<i>Zalutska O. O., Hladun O. V., Mazurets O. V.</i>	
METHOD OF PREVENTING FAILURES OF ROTATING MACHINES BY VIBRATION ANALYSIS USING MACHINE LEARNING TECHNIQUES.....	142
<i>Zbezhkivska U. R.</i>	
DEEPPAKE AUDIO DETECTION USING YOLOV8 WITH MEL-SPECTROGRAM ANALYSIS: A CROSS-DATASET EVALUATION.....	153
PROGRESSIVE INFORMATION TECHNOLOGIES.....	164
<i>Akhmetshina L. G., Yegorov A. A., Fomin A. A.</i>	
SEGMENTATION OF LOW-CONTRAST IMAGES IN THE BASIS OF EIGEN SUBSPACES OF TYPE-2 FUZZY MEMBERSHIP FUNCTIONS.....	164
<i>Caceres A., Globa L.</i>	
ANALYSIS OF DATA ACCESS APPROACHES IN A MULTI-CLOUD ENVIRONMENT.....	175
<i>Svintsycka O. M., Puleko I. V., Graf M. S., Petrosian R. V.</i>	
USE CASE METHOD IN IT PROJECT MANAGEMENT BASED ON AGILE METHODOLOGY.....	186
<i>Shportko A. V., Bomba A. Ya.</i>	
APPLICATION OF BINARY SEARCH TREE WITH FIXED HEIGHT TO ACCELERATE PROCESSING OF ONE-DIMENSIONAL ARRAYS.....	199
CONTROL IN TECHNICAL SYSTEMS.....	209
<i>Holinko O. V., Alekseev M. O., Holinko V. I., Zabelina V. A.</i>	
APPLICATION OF SINGULAR SPECTRAL ANALYSIS IN CONTROL SYSTEMS OF TECHNOLOGICAL PROCESSES AND EXPLOSION SAFETY CONTROL OF FACILITIES.....	209

РАДІОЕЛЕКТРОНІКА ТА ТЕЛЕКОМУНІКАЦІЇ

RADIO ELECTRONICS AND TELECOMMUNICATIONS

UDC 621.396.96:551.501.815

THE RESERVES FOR IMPROVING THE EFFICIENCY OF RADAR MTI SYSTEM WITH BURST-TO-BURST PROBING PULSE REPETITION FREQUENCY STAGGER

Atamanskiy D. V. – Dr. Sc., Professor, Professor at Ivan Kozhedub Kharkiv National Air Force University, Kharkiv, Ukraine.

Vasylyshyn V. I. – Dr. Sc., Professor, Head of Department at Ivan Kozhedub Kharkiv National Air Force University, Kharkiv, Ukraine.

Klymchenko V. Y. – PhD, Associate Professor, Leading Researcher at Ivan Kozhedub Kharkiv National Air Force University, Kharkiv, Ukraine.

Stovba R. L. – Teacher at Ivan Kozhedub Kharkiv National Air Force University, Kharkiv, Ukraine.

Prokopenko L. V. – Junior Researcher at Ivan Kozhedub Kharkiv National Air Force University, Kharkiv, Ukraine.

ABSTRACT

Context. The development and improvement of technologies for creating unmanned aerial vehicles (UAVs) and their use in the military conflicts, particularly in the war in Ukraine, pose the task of effectively counteraction to UAVs. The most difficult targets for radar detection are small, low-speed UAVs flying at low altitudes. Therefore, the search for efficient methods of detecting, tracking, and identifying UAVs using both existing and new promising tools is a relevant task for scientific research.

Objective. The analysis of the operation algorithm of the moving target indication (MTI) system based on the discrete Fourier transform in radars with burst-to burst probing pulse repetition frequency stagger and to propose the modernisation of the MTI system to increase the efficiency of UAV detection against passive interferences

Method. The effectiveness of the methods is determined experimentally based on the results of simulation and their comparison with known results presented in the open literature.

Results. It is shown that in the MTI system with burst-to burst probe pulse repetition frequency stagger, a non-adaptive filter for suppressing reflections from ground clutters (GC) and incoherent energy accumulation of pulses of the input burst are realized. These circumstances cause the losses in the ratio signal/(interference + inner noise). The proposals for improving the efficiency of the MTI system by transition to the construction of the MTI system with the structure “suppression filter and integration filter” are substantiated. They consist in the inclusion of a special filter for suppressing reflections from GC and fully coherent processing of the input burst pulses. The latter is realized by using the standard discrete Fourier transform (DFT) only as a integrating filter with a slight correction of the DFT algorithm. An algorithm for energy accumulation of the burst pulses using the current estimate of the inter-pulse phase incursion of the burst pulses reflected from the target is proposed. It is shown that this accumulation algorithm is close to the optimal one. The effectiveness of these proposals is analyzed in terms of the achievable signal-to-(interference+inner noise) ratio and the detection area compression ratio. It is shown that their implementation potentially leads to an increase in the detection range and an improvement in the measurement of UAV coordinates by about two times. The proposed ways are quite simply realized by digital processing used in this MTI system

Conclusions The conducted research is a development of the existing theory and technique of radar detection and recognition of air targets. The scientific novelty of the obtained results is that the algorithms of inter-period signal processing in radar with burst-to burst probing pulse repetition frequency stagger, namely the accumulation of a burst by correcting the algorithm of the standard DFT, have been further developed. The practical value of the research lies in the fact that the implementation of the proposed proposals provides approximately twice the efficiency of detecting the signal reflected from the target, compared to the standard processing device

KEYWORDS: Unmanned aerial vehicle, Radar, UAV detection, Technical requirement, Suppression filter, Optimal processing, Passive interference.

ABBREVIATIONS

AP is an adaptive processing;

ACF is an adaptive cancellation filte;

APR is an amplitude-phase response;

AFR is an amplitude-frequency response;

UAV is an unmanned aerial vehicle;

WF is a whitening filter;

SINR is a signal/(interference+receiver internal noise) ratio;

DFT is a discrete Fourier transform;

DF is a Doppler filter;
RCS is a radar cross section;
EPR is a energy phase response;
PFS is a power frequency spectrum;
EFR is an energy frequency response;
IR is an impulse response ;
CM is a correlation matrix;
GC is a ground (fixed) clutter;
IBP is an inter-burst processing (inter-train processing) ;
MF is a meteorological formation.
MP is a matched processing;
PI is a passive interference (background clutter) ;
RP is a radio pulse;
MTI is a moving target indication;
FR is a frequency response.

NOMENCLATURE

f is a normalized frequency;
 V_r is a radial velocity;
 T_i is an i -th period repetition period of probing pulses;
 T_i is a periods of pulses probing in the each;
 λ is a radar wavelength;
 K_i is a i -th weighting coefficient of “FR smoothing”;
 $|\dot{K}(\varphi)|$ is an APR;
 $|\dot{K}(\varphi)|^2$ is an EPR;
 $|\dot{K}(F_d)|$ is an AFR;
 $|\dot{K}(F_d)|^2$ is an EFR;
 $|\dot{K}_{st}(\varphi)|$ is an APR of the standard MTI system;
 $\dot{K}_{WF}(\varphi)$ is an APR of the whitening filter;
 $|\dot{K}_{CI}(\varphi)|$ is an APR of the coherent integrator;
 $|\dot{K}_{MTI}(\varphi)|$ is an APR of MTI system;
 $|\dot{K}_{st}^{(0)}(\varphi)|$ is an APR of standard MTI system with the interference suppression function turned off
 $|\dot{K}_{st}^{(1)}(\varphi)|$ is an APR of corrected DFT of the standard MTI system;
 h_i is a relation of RCS of interference and target in the resolution element;
 R_{pic} is a maximal radar range in the conditions of PI;
 K_{impr} is a coefficient of SINR improvement by MTI system
 K_{is} is a coefficient of PI suppression by MTI system;
 K_{cr} is a compression ratio of scan area
 K_{impr} is a coefficient of SINR improvement by MTI system;

k_L is a coefficient of change of losses when the MTI system is turned on;
 $\overline{k_{mean}}$ is an average power transmission efficiency of a useful signal
 F_d is a Doppler frequency;
 F_{dmin} is a minimal value of Doppler frequency;
 F_{dmax} is a maximum value of Doppler frequency;
 F_{ds} is a signal Doppler frequency;
 $F_{d,i}$ is a doppler frequency of interference;
 F_{av} is an average repetition rate of the probing pulses;
 $\overline{F_p}$ is an average repetition rate of the probing pulses;
 \mathbf{h} is a vector-column of the coefficients of IR of the system;
 \mathbf{h}_{WF} is an impulse response of whitening filter;
 $s_s(\varphi_{d,s})$ is a values of the interference spectrum and signal with phase $\varphi_{d,s}$;
 $s_s(F_{d,s})$ is an energy spectrum of the signal at the system input;
 $s_{out}(f)$ is a power spectrum process at the system output;
 $\mathbf{x}(f)$ is an acolumn vector of equidistant samples of a complex harmonic with a normalized frequency f ;
 $\mathbf{x}^*(F_d)$ is a M – dimensional row vector complex amplitudes of the RP;
 M_n is a voltage at the output of detector of n -th phase;
 M is a size of a burst of sensing radio pulses;
 γ is a SINR;
 γ_{1match} is a current values of SINR for the case when the MTI system is turned off and PI is available;
 γ_{1opt} is a current values of SINR for the case when the MTI system is turned on and PI is available;
 γ_1 is a required signal-to-noise ratios at the output;
 γ_1 is a signal-to-noise ratio at the output of the receiving path required to ensure the specified detection quality for turned off MTI system;
 γ_{1pic} is a signal-to-noise ratio at the output of the receiving path required to ensure the specified detection quality for turned on MTI system and ability of passive interference;
 γ_{1pic} is a required signal-to-noise ratio at the receiving path;
 μ is a value of the SINR for the optimal processing on the passive interference background ;
 \mathbf{R} is a CM of interperiod fluctuations of input process samples;
 R_{pic} is a maximal range of radar in the conditions of PIs (in the conditions of PIs);

$\varphi_{s,t}$ is a current inter-pulse phase offset;
 $\widehat{\varphi}_{s,t}$ is an estimate of $\varphi_{s,t}$;
 φ_n is a rotation angle of signal vector, to which n -th filter is established;
 φ_{s_i} is an initial phase of pulses of i -th pulse burst of input signal;
 σ_s^2 is a power of signal;
 σ_0^2 is a power of pulse in the RP burst;
 Z is a number of probe pulse packets;
 $\overline{\sigma_{pi}}$ is a mean value of the total effective surface of PI sources.
 $\overline{\sigma}$ is an average value of the total effective surface of the target in the pulse volume;
 $\overline{\sigma}$ is a mean values of the total effective surface of PI sources and effective scattering surface of the target, that are in the same pulse volume.

INTRODUCTION

The massive use of UAVs is a characteristic feature of modern armed conflicts [1 – 5]. The most common today are tactical UAVs and battlefield UAVs. The capabilities of airspace radar to detect and track these UAVs are extremely limited. As noted in [6, 7], one of the reasons for the difficulty of detecting and tracking tactical (small, mini and micro) UAVs is their low flight speed, which causes entering reflections from UAVs in the rejection zone of the radar MTI system. In [6], this circumstance is directly related to the imperfection of MTI systems.

However, the development of radar methods and the digital element base makes it possible to create reliable, economical and small-sized digital equivalents of existing analog signal processing systems, and fundamentally new more complex digital systems with significantly higher efficiency, close to the limit. This reveals reserves for improving the performance of both some radar systems and the tactical capabilities of radar in general. Therefore, the search for relatively simple improvements to existing processing algorithms in existing radars to increase the efficiency of UAV detection is an actual problem.

Paper purpose: to analyze the algorithm of the MTI system on a DFT device in a radar with burst-to burst probing pulse repetition frequency stagger and to propose the modernization of the MTI system to improve the efficiency of UAV detection against passive interference.

1 PROBLEM STATEMENT

Let in the S-band radar the $Mz = Z \cdot M$ -dimensional coherent bursts of the probing radio pulses are periodically transmitted, that are composed from $Z = 4$ M -element bursts, where $M = 8, 12, 16$. The periods T_i (or frequencies $F_i = 1/T_i$) of pulses probing in the each i -th ($i \in 1, Z$) burst are constant, but they are different in the different bursts. The processing in the radar with classical MTI system is performed as burst-by-burst and

can be divided on the coherent inner-burst and noncoherent binary IBP (see Fig. 1).

The package of input influences (a vector of complex signal amplitudes received from one resolution cell in the Mz sensing periods) is divided into burst of 8 (12 or 16) elements. Each burst is consequently supplied to the input of the block of $N = M$ filters performing DFT.

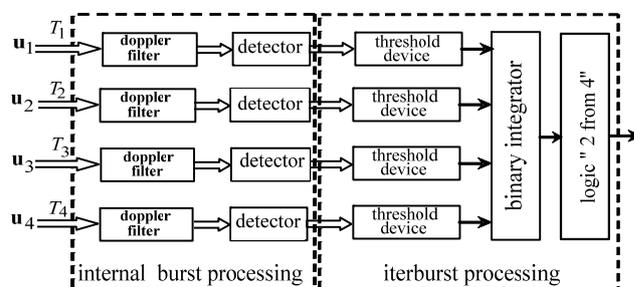


Figure 1 – Block diagram of IBP of radar signals [8]

The procedure of signal processing in n -th phase (frequency) filter is in the compensation of interperiod phase shifts of signals by corresponding phase rotation, i.e. by reduction to the same start phase with further their integration and definition of the absolute value of the obtained sum signal.

The DFs are tuned on the frequencies, that uniformly distributed on the interval $(0, F_z)$, $z \in 1, Z$ and have low level of sidelobes (i.e. “smoothed” FR). Moreover, for suppression of the reflection from the fixed scatters (or ground scatters) the voltage in the zero filter is blanked. Furthermore, the subtraction of the half value of integrated signal in the zero filter from the modulus of signals of the 1-th and 7-th filters is performed, i.e.

$$M_1 - 0,5 M_0; M_7 - 0,5 M_0.$$

The voltage at the output of detector of n -th phase filter is presented in a such way

$$M_n = \begin{cases} 0, & n = 0; \\ M_{0,n} - 0,5 M_{0,0} \cdot 0,5, & n = 1, \text{ or } n = N - 1; \\ M_{0,n}, & n \neq \{0, 1, N - 1\}, \\ & n = N - 1, \end{cases} \quad (1)$$

$$M_{0,n} = \left[\sum_{i=0}^{N-1} K_i \cdot U_i \cdot e^{j \cdot i \cdot \varphi_{s_z}} \cdot e^{-j \cdot i \cdot \varphi_n} \right], \quad (2)$$

where $\varphi_n = n(2\pi/N)$, $n \in (0, N - 1)$; $\varphi_{s_z} = 4\pi V_r T_z / \lambda$, $z \in (1, Z)$.

It is necessary for this IBP system to substantiate the proposals for modernization to improve the efficiency of UAV detection against passive interferences (clutters).

2 REVIEW OF LITERATURE

There are various methods used to detect and track unmanned aerial vehicles. A brief guide to a wide range of methods can be found, for example, in [9, 10, 11]. As shown in these works, UAV detection is carried out against the background of reflections from local objects, which is one of the reasons for the difficulty of detecting and tracking tactical (small, mini- and micro-) UAVs. Reducing the negative impact of reflections from GC (hereinafter also referred to as “passive jamming”) is based on the principle of Doppler (speed) signal selection. The suppression is realized by creating sufficiently narrow “dips” in the amplitude-frequency responses of the MTI system in the areas of passive interference. The width of these dips should correspond to the width of the energy spectrum of PI fluctuations $S_{1i}(\varphi)$.

At this moment of time, in the MTI systems of surveillance radar usually non-adaptive filter MTI systems or compensation MTI systems are used [11, 12, 13]. In the filter MTI, a non-adaptive filter of passive interference suppression can be created by zeroing the voltage of the zero filter and correcting the amplitude of the filters adjacent to it [8, 13, 14, 15]. In general, the characteristics of such filters are not consistent with the parameters of the PI energy spectrum. In addition, the low flight speed causes the reflected signals from UAVs to fall into the rejection zone of the radar MTI system [6, 15–18].

UAV detection efficiency also decreases due to the burst-to burst probe pulse repetition frequency stagger. As noted in [19], this causes signal-to-noise ratio losses, since the pulses of different packets cannot be accumulated coherently.

The following quite obvious suggestions for improving the standard MTI system under consideration follow from the brief analysis. They are in the realization of the following:

1. The fully coherent processing of the input packet. As shown in [20], such a transition will increase the signal-to-noise ratio in proportion to the increase in the size (Mz) of the pulse packet;

2. A special filter for suppressing reflections from GC using DFT only as a storage filter.

The first proposal is generally based on the a priori certainty of the current inter-pulse phase offset $\varphi_{s,t}$ of the packet pulses reflected from the target. In practice, it is absent. It is proposed to eliminate a priori uncertainty of $\varphi_{s,t}$ by estimating $\varphi_{s,t}$ based on the input packet.

This estimate of $\hat{\varphi}_{s,t}$ is not used directly for its intended purpose, but only to determine the phase of the filter number that is closest to the one in which all the pulses of the packet should be coherently accumulated.

Subsequently, the packet pulse phases are corrected to the phase value corresponding to the phase of the maximum of the selected filter. This approach almost completely reduces the negative impact of estimation errors of phase $\hat{\varphi}_{s,t}$ on the processing result.

The first suggestion is quite simply realized through digital processing. The digital signal processing used in this MTI system allows algorithmic formation of suppression filters with EFRs of the required shapes, which theoretically increases the “window of transparency” in the area of the first and seventh phase filters.

This fact gives hope for increasing the detection range and improving the measurement of UAV coordinates

3 MATERIALS AND METHODS

MTI systems are described by the following characteristics:

- the amplitude-phase $|\dot{K}(\varphi)|$ [amplitude-frequency $|\dot{K}(F_d)|$], which is numerically equal to the modulus of voltage at the output of the processing device when a reference signal of a unit voltage with an initial phase φ ($\varphi \in \varphi_{\min}, \varphi_{\max}$) [Doppler frequency F_d ($F_d \in F_{d.\min}, F_{d.\max}$)] is applied to its input;

- the energy phase (frequency) response $|\dot{K}(\varphi)|^2$ [$|\dot{K}(F_d)|^2$];

- the IR \mathbf{h} of system which is related with filter EFR by relation:

$$|\dot{K}(F_d)|^2 = |\mathbf{x}^*(F_d) \cdot \mathbf{h}|^2, F_d \in F_{d.\min}, F_{d.\max}. \quad (3)$$

Further presentation of the article’s material is based on the following well-known basic relations [21]:

$$s_{\text{out}}(f) = s_{\text{in}}(f) \cdot |\dot{K}(f)|^2, \quad (4)$$

which describes the relationship between the energy spectrum of the process $s_{\text{in}}(f)$ at the input of the system and the energy spectrum of the process at its output $s_{\text{out}}(f)$;

$$\mathbf{R} = \{r_{l,i}\}_{l,i=1}^M = \int_{-0,5}^{0,5} s(f) \cdot \mathbf{x}(f) \cdot \mathbf{x}^*(f) df, \quad (5)$$

which describes the matrix form of the Wiener-Hinchin inequality [22], that connects the correlation matrix (CM) \mathbf{R} and spectrum $s(f)$ of inter-period fluctuations of the input process samples.

Here

$$\mathbf{x}(f) = \left\{ e^{j2\pi f \ell} \right\}_{\ell=1}^M, f \in -0,5, 0,5. \quad (6)$$

The effectiveness of processing options is analyzed by the following quality indicators:

- the SINR;

– the coefficient of reduction of the detection range against the interference background, which in [23] called the compression ratio of the detection zone.

The SINR $\gamma(F_{d,s})$ is defined in the following way [24]:

$$\gamma(F_{d,s}) = P_{s,out}(F_{d,s})/P_{i,out} \quad (7)$$

where

$$P_{i,out} = \int_{-\infty}^{\infty} s_i(F_d) \cdot |\dot{K}(F_d)|^2 dF_d = \mathbf{h}^* \cdot \Phi_i \cdot \mathbf{h},$$

$$P_{s,out}(F_{d,s}) = \int_{-\infty}^{\infty} s_s(F_{d,s}) \cdot |\dot{K}(F_d)|^2 dF_d \quad (8)$$

is the average power of the interfering signal with energy spectrum $s_i(F_d)$ and the signal from the target with the $F_d = F_{d,s}$ at the output of the MTI system with energy characteristic $|\dot{K}(F_d)|^2$;

$$s_s(F_{d,s}) = \sigma_s^2 \delta(F_{d,s} - F_d), \quad \sigma_s^2 = \sigma_0^2 M \quad (9)$$

is the energy spectrum of the signal at the system input.

Compression ratio K_{cr} of scan area is defined by equation [23, eq. 2.44]

$$K_{cr}^4 = \left(\frac{R_{pic}}{R} \right)^4 = \frac{\gamma_1 \cdot \overline{k_{mean}}}{\gamma_{1pic} \cdot k_L} \cdot \left(1 - \frac{\gamma_{1pic} \cdot \overline{\sigma_{pi}}}{K_{impr} \cdot \overline{\sigma}} \right), \quad (10a)$$

$$K_{impr} = K_{is} \cdot \overline{k_{mean}} \quad (10b)$$

where γ_1 and γ_{1pic} are required signal-to-noise ratios at the output of the receiving path to ensure the specified quality of detection when the MTI system is turned off (amplitude mode) and when the MTI system is turned on and PI is available (i.e. in conditions of PI); $\overline{\sigma_{pi}}$ i $\overline{\sigma}$ are the average values of the total effective surface of the PI sources and the effective surface of the target located in the same pulse volume;

It should be noted that the above formula uses the values γ_1 , γ_{1pic} that are average values for all Doppler frequencies.

According to opinion of the authors of the article, this approach is not entirely correct, since the value of the SINR is highly dependent on the difference in absolute Doppler frequencies of the signal from the target and the interferer.

Let's estimate the value of the coefficient K_c under the influence of PI. The current values of SINR for the case when the MTI system is turned off and PI is available γ_{match} (i.e. MP case) and for the case when the MTI system is turned on and PI is available γ_{opt} (i.e. AP mode) depends on the current values of Doppler signal

frequencies $F_{d,s}$ and interference $F_{d,i}$: $\gamma_{match} = \gamma_{match}(F_{d,s}, F_{d,i})$, $\gamma_{opt} = \gamma_{opt}(F_{d,s}, F_{d,i})$. In the general case $\gamma_{adap}(F_{d,s}, F_{d,i})$ and $\gamma_{match}(F_{d,s}, F_{d,i})$ depend on the power of the packet at the input of the MTI system which depends on the range to the target with the same technical characteristics of the radar

However, the attitude of the γ_1 to γ_{1pic} in (10) eliminates this dependence, so it is possible to use any power values when modeling the compression ratio.

Obviously, to detect a signal from a target the inequalities $\gamma_{match}(F_{d,s}, F_{d,i}) \geq \gamma_1$ and $\gamma_{adap}(F_{d,s}, F_{d,i}) \geq \gamma_{1pic}$ must be followed. This ratio $\gamma_{match}(F_{d,s}, F_{d,i}) \ll \gamma_1$ is small for the frequencies of interference spectrum and, therefore, to ensure detection, the value γ_1 in (10) must satisfy the relation $\gamma_1 \geq 1/\gamma_{match}(F_{d,s}, F_{d,i})$. Similarly, for γ_{1pic} we can write $\gamma_{adap} = 1/\mu(F_{d,s}, F_{d,i})$, where $\mu(F_{d,s}, F_{d,i})$ is the current value of SINR with optimal processing against the PI background that depends on frequencies $F_{d,s}$ and $F_{d,i}$

Then expression (10) will take the following form:

$$K_{cr}^4(F_{d,s}, F_{d,i}) = \left(\frac{R_{pic}(F_{d,s}, F_{d,i})}{R} \right)^4 = \frac{\gamma_{adap}(F_{d,s}, F_{d,i}) \cdot \overline{k_{mean}}}{\gamma_{match}(F_{d,s}, F_{d,i}) \cdot k_L} \cdot \left(1 - \frac{h_i}{\gamma_{adap}(F_{d,s}, F_{d,i}) \cdot K_{impr}} \right) \quad (11)$$

Here $h_i = \overline{\sigma_{pi}}/\overline{\sigma}$ is the relation of RCS of interference and target in the resolution element. In the case when the interference is absent $h_i = 1$, and for the case when it is available provided that the interference is completely suppressed $h_i = K_{pi}$.

From the analysis of formula (11), it follows that an increase in the input signal power will increase the absolute values of ranges R_{pic} and R , but not their ratio.

The compression ratio depends on the values k_L and the noise suppression factor K_{pi} .

Graphs of compression ratio values $K_{cr}(F_{d,s}, F_{d,i})$ are used as follows. First, according to the given threshold value of the SINR γ_{th} and the graphs of the SINR (8), the argument $F_{d,s}(\varphi_{d,s})$ is determined, at which $\gamma(F_{d,s}) \geq \gamma_{th}$. This argument and the graphs of $K_{cr}(F_{d,s}, F_{d,i})$ are used to determine the specific value of the compression ratio.

4 EXPERIMENTS

The following options for processing an incoming packet by MTI systems implemented by the following devices are investigated:

- option 1: using adaptive whitening filter to suppress the reflections from the GC [25, 26] with a burst integrator of RP;
- option 2: using DFT with burst-to burst probe pulse repetition frequency stagger [13] (hereinafter “standard MTI”);
- option 3: using an adaptive whitening filter to suppress reflections from the GC with an integrator of RP burst based on a standard DFT radar [13] (without the function of zeroing the amplitude of the zero filter and correcting the amplitude of the remaining filters);
- option 4: using adaptive whitening filter of GC suppression and a integrator of RP burst based on a modification of the standard DFT radar.

First, let’s model the APR of the regular MTI system. Expressions (1), (2) describe the algorithm of the standard MTI system and allow us to determine its APR. Indeed, if the pulse packets with a unit amplitude and an inter-pulse phase shift φ_{s_i} ($\varphi_{s_i} \in \varphi_{s_{\min}}, \varphi_{s_{\max}}$) is passed to the input of the MTI system, then the modulus of the signal amplitude at the device output is proportional to the device APR $|\dot{K}_{st}(\varphi)|$, i.e.

$$|\dot{K}_{st}(\varphi)| = \max \{M_n\}_{n=0}^{N-1}. \quad (12)$$

Fig. 2 shows, for a range of phases $\varphi_i \in -90^0, 90^0$, the squares of the APR $(|\dot{K}_{st,n}(\varphi_s)|^2)$ normalized to value of M (hereinafter referred to as “APRs”) of the standard MTI system with the functions of “FR smoothing” and suppression of reflections from GC (curve 2) and without the function of “FR smoothing” (curve 3).

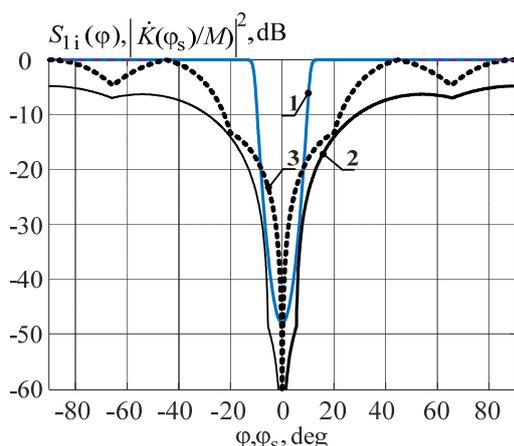


Figure 2 – Power spectrum of interference of the “GC” type and EPR of the standard MTI system

For comparison, curve 1 shows the energy anti-spectrum $S_{1i}(\varphi) = 1/[S_i(\varphi)/S_{i,\max}(\varphi)]$ of the interference of the “reflection from GC” type normalized to its

maximum value with a relative (relative to the receiver noise power) power $h_{1i} = 30$ dB, Gaussian spectrum shape ($p \rightarrow \infty$), zero phase (radial velocity) ($\varphi_{1i} = 0$ ($V_{r,1i} = 0$ m/s)), with the correlation coefficient of the radial velocity fluctuations of the interferer $\rho_i = 0.9992$, which corresponds to the width of the velocity fluctuation spectrum $W_{1i} = 0.5$ m/s for $\lambda = 0.1$ m and the average repetition rate of the probing pulses $F_s = \overline{F}_i = 1.54$ kHz.

It can be seen that “FR smoothing” not only widens the main filter lobes, but also reduces the maximum possible value of the output voltage. The finite dimension M of the discrete Fourier transform causes the fluctuations in the output amplitude of the MTI system and is the cause of additional losses in the output signal amplitude at the inter-pulse shift, which does not correspond to the phases of the maximums of FR filter main lobes.

The analysis of Fig. 2 and expressions (1), (2) shows that only individual bursts of pulses in a packet are processed coherently, and their results are integrated incoherently, i.e., the burst of pulses is not processed optimally in general, that is expected to lead to losses in the signal-to-noise ratio.

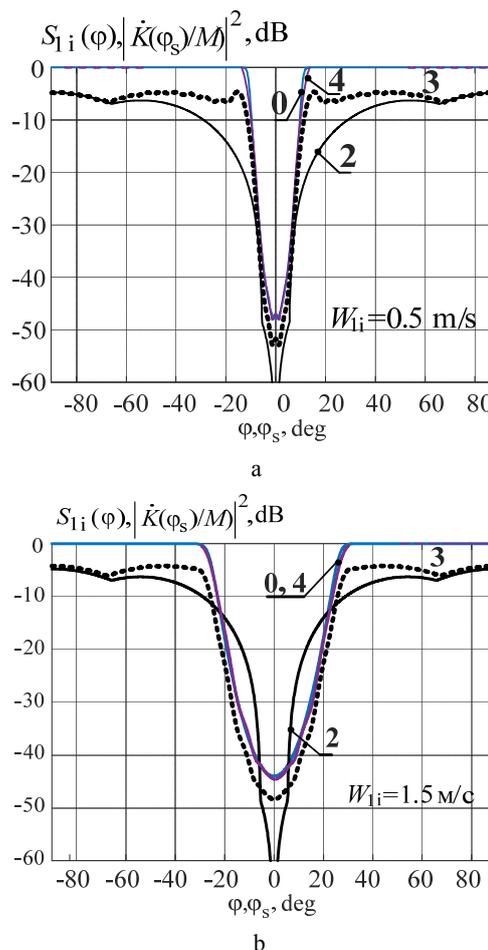
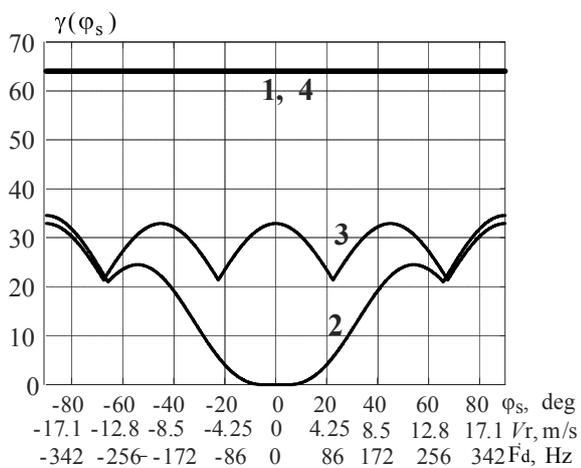


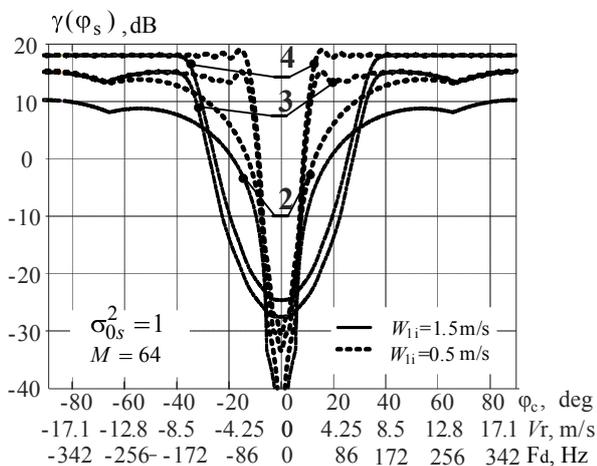
Figure 3 – EPR of MTI system

Now, by simulating the EPR of MTI systems, the values of the SINR μ (7) and the compression ratio K_c (11), the graphs of which are shown in Figures 3, 4 and 5, respectively, we will substantiate the feasibility of the proposals formulated in paragraph 2.

The curves 0 in Fig. 3 show the jammer antispectrum $S_{i1}(\varphi) = 1/[S_i(\varphi)/S_{i,max}(\varphi)]$ normalized to the maximum value with a spectral width of jammer velocity fluctuations $W_{i1} = 0.5$ m/s (Fig. 3a) and $W_{i1} = 1.5$ m/s (Fig. 3b). Curves 2, 3, 4, respectively, reflect the EPR of the standard MTI system with 8-element DFT (MTI system according to variant 2), curves 3 – MTI system according to variant 3, curves 4 – MTI system according to variant 4.



a



b

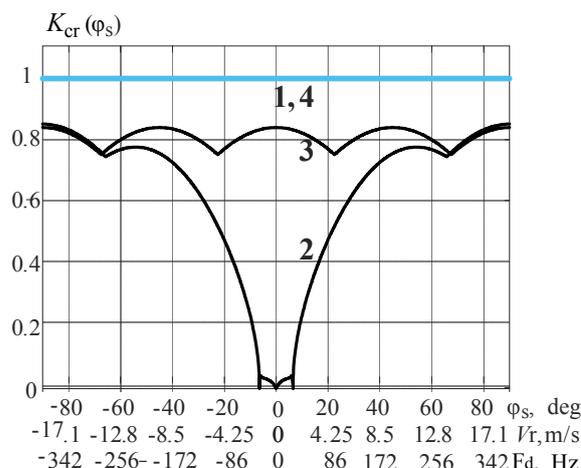
Figure 4 – SINR at the outputs of the MTI system during signal processing against the background of receiver noise (a) and GC noise (b)

The graphs in Figs. 4 and 5 correspond to the situation of processing a $Mz = 64$ -element packet (8) of a mixture of the reflected signal from the target with $\sigma_0^2 = 1$ and the receiver's internal noise with power $\sigma_n^2 = 1$ (Figs. 4a, 5a)

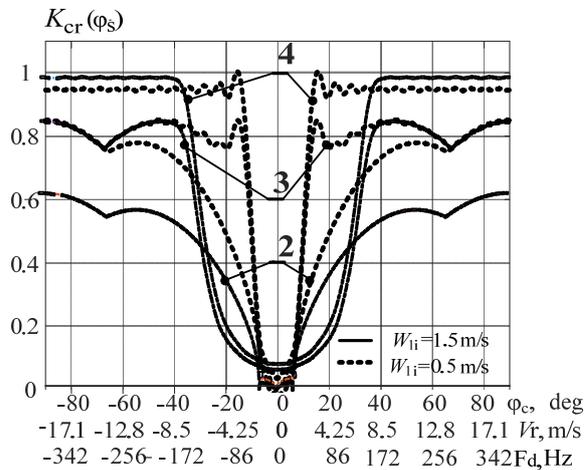
and the interference of the “GC” type (Figs. 4b, 5b) with the parameters specified for the model situation in Fig. 2.

Curves 1 illustrate the efficiency of the MTI system implemented according to variant 1. They correspond to the efficiency of package optimal processing according to the criterion of maximum SINR and determine the maximum possible values of SINR and compression ratio for the specified conditions. In addition, they perform the function of controlling the compliance of model values of performance indicators with the results known in the scientific literature.

Curves 2–4 show the effectiveness of the MTI system implemented according to the respective design options.



a



b

Figure 5 – Compression ratio of the detection zone during signal processing against the background of receiver noise (a) and GC noise (b)

The compression ratio K_c (11) was calculated under the conditions $k_L = 1$ and interference suppression ratio $K_{si} = h_i$.

Fig. 6 shows the graphs of the SINR at the output of the standard MTI system when processing the signal against the background of a mixture of reflections from GC and MF. The package parameters correspond to the model situation of Figs. 4, 5.

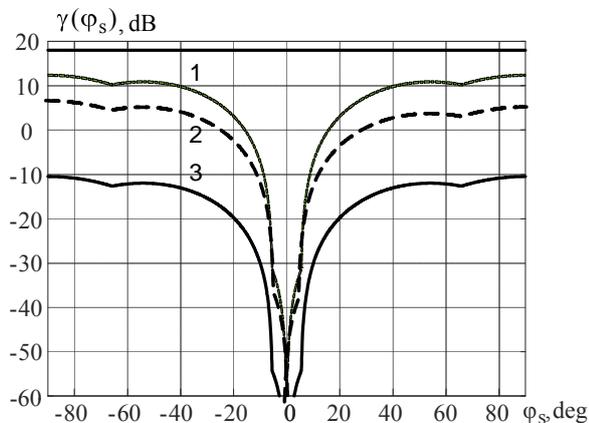


Figure 6 – SINR at the output of the standard MTI during signal processing against the background of reflections from GC and MF

Here, curve 1 corresponds to the reflections from the GC with a width $W_{1i} = 0.5$ m/s, curve 2 corresponds to one with $W_{1i} = 1.5$ m/s (Fig. 3b), and curve 3 corresponds to the reflections from the GC with a width $W_{1i} = 0.5$ m/s and MF with a width $W_{1i} = 0.5$ m/s, power $h_{2i} = 30$ dB, Gaussian spectrum shape ($p \rightarrow \infty$), radial velocity $V_{r,2i} = 15$ m/s.

5 RESULTS

1. As follows from Fig. 3, the shape and characteristics of the EPR of the standard system are fixed. The shape of the MTI system’s amplitude-frequency response is uneven over the entire range of possible target Doppler velocities. The degree and depth of the unevenness is determined by the number of DFT filters. This circumstance leads to a decrease in the maximum possible value of the received signal from the target when the target Doppler frequency and the maximums of the DFT filters do not match.

In general, the shape of the EPR does not coincide with the PFS of the GC reflections. If the EPR of the filter is narrower than the PFS of the GC reflections, then uncompensated interference residuals appear. Otherwise, the excessive expansion of the interference suppression zone reduces the level of the useful signal at the output of the EPR filter at Doppler frequencies free of interference. In both cases, the mismatch of the power phase spectra of the reflections from the GC and the EPR filter leads to a decrease in the SNIR at the output of the MTI system.

The EPR parameters of the remaining MTI systems (options 1, 3 – 5) change adaptively to take into account possible changes in the position and width of the

interfering spectrum. In particular, they automatically monitor the situation of absence of interference. Such EPRs are formed under conditions of a priori certainty, or estimation of interference parameters.

2. Blanking of the null filter and correction of the output amplitude of the filters adjacent to it is one of the ways to implement a non-adaptive passive interference suppression filter of the “reflection from GC” type. Failure to disable the zero filter blanking function and correct the amplitude of the remaining filters and “smoothing” the side lobes of the EPR leads to loss of the useful signal in the absence of interference.

3. The burst-to-burst stagger of the repetition frequency of the sensing pulses causes different inter-pulse phase hops in the pulses of the packet bursts. This can lead to the distribution of the total amplitude of the packet pulses over different DFT phase filters, and as a result, the sufficiency of the burst amplitude to make a decision on target detection does not guarantee its detection at the maximum amplitude of one of the phase filters.

4. EPR features cause different efficiencies of the systems under consideration. The efficiency of the standard system decreases almost twice due to the forced implementation of the zero filter blanking function and the correction of the amplitude of the remaining filters and the “smoothing” of the EPR side lobes, which has already been noted in the scientific literature. The maximum possible values of the EPR for a given model situation, both against the background of receiver noise and against the background of interference, are achieved by the MTI system according to variants 1 and 4. The efficiency of the MTI system according to variant 3 is between the efficiency of variants 1 and 4.

5. The influence of a mixture of interference of the “reflection from GC” and “reflection from meteorological phenomena” types due to the lack of a mechanism for suppressing the latter leads to a decrease in the SNIR in proportion to the power of the uncompensated interference.

6 DISCUSSION

Let us substantiate these results. To do this, let us analyze the APR and SINR of the MTI systems under consideration. The block diagram of the MTI system of variant 1 contains series-connected ACFs of the interference of “GC” type and a coherent integrator of the RP burst. As an ACF, a whitening filter (WF) with an APR of the form

$$|\dot{K}_{WF}(\varphi)| = 1/\sqrt{s_1(\varphi)} .$$

can be used.

It can be determined by the known CM of interference [27]:

$$|\dot{K}_{WF}(\varphi)| = |\mathbf{h}_{WF} \cdot \mathbf{x}(\varphi)|, \quad \mathbf{h}_{WF} = \frac{1}{\sqrt{\Psi_{1,1}}} \boldsymbol{\Psi}_1, \quad (13)$$

where $\boldsymbol{\Psi}_1$ is the first row of CM $\boldsymbol{\Psi} = \boldsymbol{\Phi}_1^{-1}$, which is inverse matrix to interference CM $\boldsymbol{\Phi}_1$, and $\Psi_{1,1}$ is its first element; $\mathbf{x}(\varphi)$ is the reference vector with phase $\varphi = 2\pi F_d / F_p$, which by analogy with (6) is equal to

$$\mathbf{x}(\varphi) = \{\dot{x}_m(\varphi)\}_{m=1}^M, \quad \dot{x}_m(\varphi) = x_0 \cdot e^{j\varphi \cdot (m-1)}. \quad (14)$$

The integrator APR can be determined as follows:

$$|\dot{K}_{CI}(\varphi)| = \delta(\varphi - \varphi_s) = \begin{cases} 1, & \varphi = \varphi_s \\ 0, & \varphi \neq \varphi_s \end{cases},$$

which corresponds to the M -dimensional IR with the form $\mathbf{h}_{CI} = \{1, 1, \dots, 1\}$.

The final APR of the MTI system will have the following form:

$$|\dot{K}_{MTI}(\varphi)| = \left| \frac{\delta(\varphi - \varphi_s)}{\sqrt{s_i(\varphi)}} \right|.$$

It is easy to show that the powers of the interfering $P_{i.out}$ and useful signals $P_{s.out}(\varphi_{d.s})$ (9) at the output of the MTI system are equal:

$$P_{i.out} = \int_{-0.5}^{0.5} s_i(\varphi) \cdot |\dot{K}_{MTI}(\varphi)|^2 d\varphi = 1, \\ P_{s.out}(\varphi_{d.s}) = \int_{-\infty}^{\infty} s_s(\varphi_{d.s}) \cdot |\dot{K}_{MTI}(\varphi)|^2 d\varphi = \frac{M\sigma_0^2}{s_i(\varphi_{d.s})}, \quad (15)$$

where $s_i(\varphi_{d.s})$ and $s_s(\varphi_{d.s})$ are the values of the interference spectrum and signal with phase $\varphi_{d.s}$.

Therefore, the spectrum (5) at the WF output is uniform and represents the receiver noise spectrum, which creates conditions for the MP of the input signal package. In the MTI system according to variant 1, the maximum possible value of the SINR (8) is achieved [27]:

$$\gamma(\varphi) = \mathbf{x}^*(\varphi) \boldsymbol{\Psi} \mathbf{x}(\varphi) = \frac{M\sigma_0^2}{s(\varphi_{d.s})}.$$

The block diagram of the MTI system of variant 3 contains series-connected ACF jammers of the ‘‘GC’’ type and the integrator of the standard MTI radar system [10] with the jamming function disabled.

APR of such a system:

$$|\dot{K}_{MTI3}(\varphi)| = |\dot{K}_{WF}(\varphi)| \cdot |\dot{K}_{st}^{(0)}(\varphi)|,$$

where $|\dot{K}_{WF}(\varphi)|$ is the APR of WF (13);

$$|\dot{K}_{st}^{(0)}(\varphi)| = \max \{M_{0,n}\}_{n=0}^{N-1}$$

is the of APR the standard MTI system with the interference suppression function turned off, and the values $M_{0,n}$ are determined by expression (2)

This correction is obvious, since the interference is suppressed by an adaptive suppression filter. The DFT of the standard MTI system is used as an RP integrator with enabled function of the APR lobe smoothing. This circumstance reduces the maximum achievable value SINR of the MTI system’s and compression ratio, as illustrated by curves 3 in Figs. 4 and 5, respectively.

The block diagram of the MTI system of variant 4 contains the consequently connected the ACF of jammer of the ‘‘GC’’ type with the corrected DFT of the standard MTI system (see the formulation of proposal 1 of the ‘‘Problem Statement’’ section).

APR of this system can be presented as:

$$|\dot{K}_{MTI3}(\varphi)| = |\dot{K}_{WF}(\varphi)| \cdot |\dot{K}_{st}^{(1)}(\varphi)|,$$

where $|\dot{K}_{WF}(\varphi)|$ is the APR of WF (13).

Now let’s determine the potentially achievable SINR at the output of the standard MTI system when processing the signal from the resolution element against the background of reflections from the GC and MF (Fig. 6). This situation occurs when the UAV moves in cloud formations. The significant decrease in the SINR (curve 3) which is observed in Fig. 6 is due to the non-zero average velocity of MF. Powerful reflections from the MF pass through the ‘‘unprotected’’ DFT filters to the output of the MTI system, increasing the overall level of the noise component of the SINR.

The absence of a ‘‘dip’’ in the MTI plot at the MF phase (velocity), which can be observed, for example, in the classical matched processing, is explained by different ways of determining the output voltage of the MTI system. Due to the a priori uncertainty of the MF velocity, the output voltage is considered to be the maximum voltage value from the entire set of filters. With MP, the voltage for determining the SINR is determined at the output of a specific filter tuned to the phase of the detected signal. This circumstance allows for a slight reduction in the influence of interference.

In general, if the MTI system is affected by different types of interference, appropriate filters must be provided to suppress each type of interference. For the ‘‘MF’’ type interference, a digital interleaved autocompensator can be used as such a filter. The two suppression filters can be combined into one adaptive suppression filter based on the CM estimate of the interference.

CONCLUSIONS

1. The processing algorithm of the MTI system with packet-by-packet sampling of the repetition rate of sensing pulses is analyzed. It is shown that the MTI implements a non-adaptive filter for suppressing

reflections from the GC and a “semi-coherent” energy accumulation of the input packet pulses. In general, the characteristics of such filters are not matched with the parameters of the PI energy spectrum, and this integration causes losses in the SINR.

2. It is shown that there are reserves for increasing its efficiency by improving the existing algorithm of processing input signals: transition to the classical construction of the MTI system of the “suppression filter and integration filter” type. This transition consists in the inclusion of a special filter to suppress reflections from the GC and fully coherent processing of the input packet pulses. The latter is realized by using the standard DFT only as an integration filter with a slight correction of the DFT algorithm.

3. A near-optimal algorithm for integrating the energy of packet pulses is proposed, which uses the current estimate of the inter-pulse change of the phase $\phi_{s,t}$ of the packet pulses reflected from the target. This estimate $\phi_{s,t}$ of is used only to determine the phase of the filter number that is closest to the one in which all the pulses of the packet should be coherently accumulated. Subsequently, the phases of the packet pulses are corrected to the phase value corresponding to the maximum phase of the selected filter.

4. It is proposed to algorithmically form suppression filters from the EFR of the required shapes, which increases the “window of transparency” in the area of the first and seventh phase filters. This circumstance leads to an increase in the detection range and improvement of UAV coordinate measurement by about 2 times. The adaptive suppression filter can be replaced by a non-adaptive filter with a sharp decline in APR.

5. The proposed ways are quite simply realized by the digital processing used in this MTI system.

The presented research results illustrate the potential reserves that can be realized if signal processing algorithms in MTI systems with a burst-to-burst stagger of the probing pulse repetition rate are modernized. It is advisable to investigate the effectiveness of the proposed solutions in the situation of using their estimates instead of the real values of IR filters, as well as different values of the coefficients possible in practice $k_L > 1$ and the interference suppression (cancellation) factor $K_{si} \neq h_i$.

REFERENCES

1. Kunertova D. The war in Ukraine shows the game-changing effect of drones depends on the game, *Bulletin of the Atomic Scientists*, 2023, Vol. 79, No. 2, pp. 95–102. DOI: 10.1080/00963402.2023.2178180
2. Hwang W.-J. Song S.-H. The extension of Turkish influence and the use of drones, *Comparative Strategy*, 2022, pp. 1–20. DOI: 10.1080/01495933.2022.2111907
3. Calcara A. [et al.] Will the Drone Always Get through? Offensive Myths and Defensive Realities, *Security Studies*, 2022, pp. 1–35. DOI: 10.1080/09636412.2022.2153734.
4. Tuna F. Increasing drones' combat effectiveness: an alternative analysis for integration into comprehensive military and technological systems/ Fatih Tuna, *Uludağ Üniversitesi Fen-Edebiyat Fakültesi Sosyal Bilimler Dergisi*, 2023, Vol. 25, No. 46, pp. 187–204. DOI: 10.21550/sosbilder.1336111
5. Kartashov V. M. et al. Information characteristics of sound radiation of small unmanned aerial vehicles, *Telecommunications and Radio Engineering*, 2018, Vol. 77, No. 10, pp. 915–924. DOI: 10.1615/telecomradeng.v77.i10.70
6. Riabukha V. P. Radar Surveillance of Unmanned Aerial Vehicles (Review), *Radioelectronics and Communications Systems*, 2020, Vol. 63, No. 11, pp. 561–573. DOI: 10.3103/s0735272720110011.
7. Zhuk S. Ya. et al. Adaptive Radar Tracking Algorithm for Maneuverable UAV with Probabilistic Identification of Data Using Coordinate and Amplitude Characteristics, *Radioelectronics and Communications Systems*, 2022, Vol. 65, No. 10, pp. 503–516. DOI: 10.3103/s073527272212007x.
8. Riabukha V. P. et al. Comparative Experimental Investigations of Adaptive and Non-adaptive MTI Systems in Pulse Radars of Various Applications and Wave Ranges, *Radioelectronics and Communications Systems*, 2022, Vol. 65, No. 4, pp. 165–176. DOI: 10.3103/s073527272204001x.
9. Vasilii Semkin et al. Analyzing Radar Cross Section Signatures of Diverse Drone Models at mmWave Frequencies, *IEEE Access*, 2020, Vol. 8, P. 48958–48969. DOI: 10.1109/access.2020.2979339
10. Zitar R. A., Mohammad A.-B., Mohamad R., Sofian K. A review of UAV Visual Detection and Tracking Methods [Electronic resource], *arXiv.org*, 08.06.2023. <https://arxiv.org/abs/2306.05089>
11. Park S. et al. Survey on Anti-Drone Systems: Components, Designs, and Challenges, *IEEE Access*, 2021, Vol. 9, pp. 42635–42659. DOI: 10.1109/access.2021.3065926
12. Gui R. et al. FDA radar with doppler-spreading consideration: Mainlobe clutter suppression for blind-doppler target detection, *Signal Processing*, 2021, Vol. 179. pp. 107773. 10.1016/j.sigpro.2020.107773
13. Muravskiy A. P., Azerskiy M. A. Comparative evaluation of methods for compensation of passive interference in radar location stations, *Radio industry*, 201, No. 3, pp. 79–85. DOI: 10.21778/2413-9599-2017-3-79-85
14. May I. D. [et al.] Radar station 36D6. Operation and maintenance : textbook. Zaporozhye : KEMZ “Iskra”, 2006, 139 p.
15. Chornoborodova N. P., Chornoborodov M. P. Radar 35D6 Speed Characteristic Improving. // *Systems of Arms and Military Equipment*. – 2011. – Vol. 2, no. 26. – P. 178–182. <http://eir.zntu.edu.ua/handle/123456789/113>.
16. Zhang W., Ma S., Du Q. Optimization of Adaptive MTI Filter, *International Journal of Communications, Network and System Sciences*, 2017, Vol. 10, No. 08, pp. 206–217. DOI: 10.4236/ijcns.2017.108b022
17. Pevtsov G. V., Oleshchuk M. M. Analysis of the capabilities of surveillance radar RIT to detect, track and identify unmanned aerial vehicles, *Weapon Systems and Military Equipment*, 2021, No. 3(67), pp. 24–30. DOI: 10.30748/soivt.2021.67.03
18. Klymchenko V. et al. Possible directions of modernization of radar stations of radio engineering troops in order to increase the efficiency of detection of small unmanned aerial vehicles, *Weapon systems and military equipment*, 2024, № 1 (77), pp. 16–21. DOI.org/10.30748/soivt.2024.77.02
19. Xu J. et al. Clutter and Range Ambiguity Suppression Using Diverse Pulse Train in Pulse Doppler System, *Sensors*, 2018, Vol. 18, No. 7, P. 2326. DOI: 10.3390/s18072326
20. Skolnik M. I. Radar Handbook, Third Edition. 3rd ed., McGraw-Hill Professional, 2008, 1328 p.
21. Lekhovytskyi D. I. et al. Estimation of the energy spectrums of reflections in pulse doppler weather radars. Part 1. Modifications of the spectral estimation algorithms, *Radioelectronics and Communications Systems*, 2015, Vol. 58, No. 12, pp. 523–550. DOI: 10.3103/s0735272715120018.

22. Stoica P. Introduction to spectral analysis. Upper Saddle River, N.J, Prentice Hall, 1997, 319 p.
23. Blokhin V. P., Bondarenko B. F., Nesnov V. T., Ugolnikov V. E. Fundamentals construction of radars of radio engineering troops. Kiev, KHIRTS of air ference, 1987, 368 p.
24. Riabukha V. P. et al. Comparative Experimental Investigations of Adaptive and Non-adaptive MTI Systems in Pulse Radars of Various Applications and Wave Ranges, *Radioelectronics and Communications Systems*, 2022, Vol. 65, No. 4, pp. 165–176. DOI: 10.3103/s073527272204001x
25. Drone and Controller Detection and Localization: Trends and Challenges / Jawad Yousaf [et al.] // *Applied Sciences*. – 2022. – Vol. 12, no. 24. – P. 12612. DOI: 10.3390/app122412612 .
26. Lekhovytskyi D. I. et al. Lattice Filtration Theory. Part I: One-Dimensional Lattice Filters, *Telecommunications and Radio Engineering*, 2021, Vol. 80, No. 5, pp. 41–79. DOI: 10.1615/telecomradeng.2021039186 .
27. Riabukha V. P. et al. Protection of Coherent Pulse Radars against Combined Interferences. 4. Adaptive Systems of Space-Time Signal Coprocessing against Background of Combined Interference Based on Two-Dimensional ALF, *Radioelectronics and Communications Systems*, 2022, Vol. 65, No. 8, pp. 400–410. DOI: 10.3103/s0735272722070019 .
- Received 16.12.2024.
Accepted 18.02.2025.

УДК 621.396.96:551.501.815

РЕЗЕРВИ ПІДВИЩЕННЯ ЕФЕКТИВНОСТІ СИСТЕМИ СРЦ РЛС З ПОПАЧКОВОЮ ВОБУЛЯЦІЄЮ ЧАСТОТИ ПОВТОРЕННЯ ЗОНДУВАЛЬНИХ ІМПУЛЬСІВ

Атаманський Д. В. – д-р техн. наук, професор, професор Харківського національного університету Повітряних Сил імені Івана Кожедуба, Харків, Україна.

Василишин В. І. – д-р техн. наук, професор, начальник кафедри Харківського національного університету Повітряних Сил імені Івана Кожедуба, Харків, Україна.

Климченко В. Й. – канд. техн. наук доцент провідний науковий співробітник Харківського національного університету Повітряних Сил імені Івана Кожедуба, Харків, Україна.

Стовба Р. Л. – викладач Харківського національного університету Повітряних Сил імені Івана Кожедуба, Харків, Україна.

Прокопенко Л. В. – молодший науковий співробітник Харківського національного університету Повітряних Сил імені Івана Кожедуба, Харків, Україна

АНОТАЦІЯ

Актуальність. Розвиток і удосконалення технологій створення безпілотних літальних апаратів та їх застосування у військових конфліктах, зокрема у війні в Україні, ставить завдання ефективної протидії їм. Найбільш складними цілями для радіолокаційного виявлення є малорозмірні, маловидкісні малі безпілотні літальні апарати (БПЛА), що летять на малих висотах. Тому пошук ефективних способів виявлення, супроводження та ідентифікації БПЛА з використанням як існуючих, так і нових перспективних засобів є актуальним завданням наукових досліджень.

Мета. Проаналізувати алгоритм роботи системи селекції рухомих цілей (СРЦ) на пристрої дискретного перетворення Фур'є в радіолокаційних станціях (РЛС) із по-пачковою вобуляцією частоти повторення зондувальних імпульсів і запропонувати модернізацію системи СРЦ для підвищення ефективності виявлення БПЛА на тлі пасивних завад.

Метод. Ефективність методів визначається експериментально за результатами імітаційного моделювання та їхнього порівняння з відомими результатами, викладеними у відкритій літературі.

Результати. Показано, що в системі СРЦ з по-пачковою вобуляцією частоти повторення зондувальних імпульсів реалізується неадаптивний фільтр придушення віддзеркалень від місцевих предметів і некогерентне накопичення енергії імпульсів пачок імпульсів вхідного пакета. Вказані обставини обумовлюють втрати у відношенні сигнал/(завада + внутрішній шум). Обґрунтовано пропозиції підвищення ефективності системі СРЦ за рахунок переходу до побудови системи СРЦ за структурою «фільтр придушення + фільтр накопичення». Вони полягають у включенні спеціального фільтра придушення віддзеркалень від місцевих предметів (МП) і повністю когерентну обробку імпульсів вхідного пакета. Останнє реалізується використанням штатного дискретного перетворення Фур'є (ДПФ) лише як фільтра – накопичувача з незначною корекцією алгоритму роботи ДПФ. Запропоновано алгоритм накопичення енергії імпульсів пакета, що використовує поточну оцінку міжімпульсного набігу фази $\varphi_{с.ц}$ імпульсів пакета, відбитих від цілі. Показано, що такий алгоритм накопичення близький до оптимального. Проаналізовано ефективність вказаних пропозицій за критерієм досяжного відношення сигнал/(завада+внутрішній шум) і коефіцієнта стиску зони виявлення. Показано, що їх реалізація потенціально приводить до збільшення дальності виявлення і покращення виміру координат БПЛА приблизно в 2 рази. Запропоновані шляхи доволі просто реалізуються цифровою обробкою, що використовується в даній системі СРЦ.

Висновки. Проведені дослідження є розвитком існуючої теорії й техніки радіолокаційного виявлення і розпізнавання повітряних цілей. Наукова новизна отриманих результатів полягає в тому, що набули подальшого розвитку алгоритми міжперіодної обробки сигналів в РЛС з по-пачковою вобуляцією частоти повторення зондувальних імпульсів, а саме накопичення пакету імпульсів за рахунок корекції алгоритму штатного ДПФ. Практична цінність досліджень полягає в тому, що реалізація запропонованих пропозицій забезпечує приблизно в двічі більшу ефективність виявлення сигналу, відбитого від цілі, у порівнянні з штатним пристроєм обробки.

КЛЮЧОВІ СЛОВА: безпілотний літальний апарат, РЛС виявлення БПЛА, технічні вимоги, фільтр придушення, оптимальна обробка, пасивні завади.

ЛІТЕРАТУРА

1. Kunertova D. The war in Ukraine shows the game-changing effect of drones depends on the game / Dominika Kunertova // *Bulletin of the Atomic Scientists*. – 2023. – Vol. 79, No. 2. – P. 95–102. DOI: 10.1080/00963402.2023.2178180
2. Hwang W.-J. The extension of Turkish influence and the use of drones / Won-June Hwang, Seung-Hoon Song // *Comparative Strategy*. – 2022. – P. 1–20. DOI: 10.1080/01495933.2022.2111907
3. Will the Drone Always Get through? Offensive Myths and Defensive Realities / [Antonio Calcara et al.] // *Security Studies*. – 2022. – P. 1–35. DOI: 10.1080/09636412.2022.2153734.
4. Tuna F. Increasing drones' combat effectiveness: an alternative analysis for integration into comprehensive military and technological systems / Fatih Tuna // *Uludağ Üniversitesi Fen-Edebiyat Fakültesi Sosyal Bilimler Dergisi*. – 2023. – Vol. 25, No. 46. – P. 187–204. DOI: 10.21550/sosbilder.1336111
5. Information characteristics of sound radiation of small unmanned aerial vehicles / [V. M. Kartashov et al.] // *Telecommunications and Radio Engineering*. – 2018. – Vol. 77, No. 10. – P. 915–924. DOI: 10.1615/telecomradeng.v77.i10.70
6. Riabukha V. P. Radar Surveillance of Unmanned Aerial Vehicles (Review) / V. P. Riabukha // *Radioelectronics and Communications Systems*. – 2020. – Vol. 63, No. 11. – P. 561–573. DOI: 10.3103/s0735272720110011
7. Adaptive Radar Tracking Algorithm for Maneuverable UAV with Probabilistic Identification of Data Using Coordinate and Amplitude Characteristics / [S. Ya. Zhuk et al.] // *Radioelectronics and Communications Systems*. – 2022. – Vol. 65, No. 10. – P. 503–516. DOI: 10.3103/s073527272212007x
8. Comparative Experimental Investigations of Adaptive and Non-adaptive MTI Systems in Pulse Radars of Various Applications and Wave Ranges / [V. P. Riabukha et al.] // *Radioelectronics and Communications Systems*. – 2022. – Vol. 65, No. 4. – P. 165–176. DOI: 10.3103/s073527272204001x
9. Analyzing Radar Cross Section Signatures of Diverse Drone Models at mmWave Frequencies / [Vasilii Semkin et al.] // *IEEE Access*. – 2020. – Vol. 8. – P. 48958–48969. DOI: 10.1109/access.2020.2979339
10. Zitar R. A. A review of UAV Visual Detection and Tracking Methods [Electronic resource] / [Zitar Raed Abu, Mohammad Al-Betar, Mohamad Ryalat, Sofian Kassaymeh] // *arXiv.org*. – 08.06.2023. <https://arxiv.org/abs/2306.05089>
11. Survey on Anti-Drone Systems: Components, Designs, and Challenges / Park Seongjoon et al. // *IEEE Access*. – 2021. – Vol. 9. – P. 42635–42659. DOI: 10.1109/access.2021.3065926
12. FDA radar with doppler-spreading consideration: Mainlobe clutter suppression for blind-doppler target detection / [Ronghua Gui et al.] // *Signal Processing*. – 2021. – Vol. 179. – P. 107773. DOI: 10.1016/j.sigpro.2020.107773
13. Muravskiy A. P. Comparative evaluation of methods for compensation of passive interference in radar location stations / A. P. Muravskiy, M. A. Azerskiy // *Radio industry*. – 2017. – No. 3. – P. 79–85. DOI: 10.21778/2413-9599-2017-3-79-85
14. Май І. Д. Радіолокаційна станція З6Д6. Експлуатація та технічне обслуговування: навч. посіб. / [І. Д. Май та ін.]. – Запоріжжя : КЕМЗ «Іскра», 2006. – 139 с.
15. Чорнобородова Н. П. Поліпшення швидкісної характеристики РЛС З5Д6 / Н. П. Чорнобородова, М. П. Чорнобородов // *Системи озброєння і військова техніка*. – Х.: ХУПС. – 2011. – № 2(26). – С. 178–182. <http://eir.zntu.edu.ua/handle/123456789/113>.
16. Zhang W. Optimization of Adaptive MTI Filter / Wenxu Zhang, Shudi Ma, Qiuying Du // *International Journal of Communications, Network and System Sciences*. – 2017. – Vol. 10, No. 08. – P. 206–217. DOI: 10.4236/ijcns.2017.108b022
17. Певцов Г. В. Аналіз спроможностей оглядових РЛС РТВ щодо виявлення, супроводження та ідентифікації безпілотних літальних апаратів / Г. В. Певцов, М. М. Олещук // *Системи озброєння і військова техніка*. – 2021. – № 3(67). – С. 24–30. DOI: 10.30748/soivt.2021.67.03
18. Можливі напрямки модернізації радіолокаційних станцій радіотехнічних військ з метою підвищення ефективності виявлення малорозмірних безпілотних літальних апаратів / [В. Й. Климченко та ін.] // *Системи озброєння і військова техніка*. – 2024. – № 1 (77). – С. 16–21. DOI.org/10.30748/soivt.2024.77.02
19. Clutter and Range Ambiguity Suppression Using Diverse Pulse Train in Pulse Doppler System / [Jiacen Xu et al.] // *Sensors*. – 2018. – Vol. 18, No. 7. – P. 2326. DOI: 10.3390/s18072326
20. Skolnik M. I. Radar Handbook, Third Edition / Merrill I. Skolnik. – 3rd ed. : McGraw-Hill Professional, 2008. – 1328 p.
21. Estimation of the energy spectrums of reflections in pulse doppler weather radars. Part 1. Modifications of the spectral estimation algorithms / [D. I. Likhovitskiy et al.] // *Radioelectronics and Communications Systems*. – 2015. – Vol. 58, No. 12. – P. 523–550. DOI: 10.3103/s0735272715120018
22. Stoica P. Introduction to spectral analysis / Petre Stoica. – Upper Saddle River, N.J : Prentice Hall, 1997. – 319 p.
23. Blokhin V. P. Fundamentals construction of radars of radio engineering troops / [V. P. Blokhin, B. F. Bondarenko, V. T. Nesnov et al.]. – Kiev : KHIRTS of air ference, 1987. – 368 p.
24. Comparative Experimental Investigations of Adaptive and Non-adaptive MTI Systems in Pulse Radars of Various Applications and Wave Ranges / [V. P. Riabukha et al.] // *Radioelectronics and Communications Systems*. – 2022. – Vol. 65, No. 4. – P. 165–176. DOI: 10.3103/s073527272204001x
25. Drone and Controller Detection and Localization: Trends and Challenges / [Jawad Yousaf et al.] // *Applied Sciences*. – 2022. – Vol. 12, No. 24. – P. 12612. DOI: 10.3390/app122412612
26. Lattice Filtration Theory. Part I: One-Dimensional Lattice Filters / [D. I. Likhovitskiy et al.] // *Telecommunications and Radio Engineering*. – 2021. – Vol. 80, No. 5. – P. 41–79. DOI: 10.1615/telecomradeng.2021039186
27. Protection of Coherent Pulse Radars against Combined Interferences. 4. Adaptive Systems of Space-Time Signal Coprocessing against Background of Combined Interference Based on Two-Dimensional ALF / [V. P. Riabukha et al.] // *Radioelectronics and Communications Systems*. – 2022. – Vol. 65, No. 8. – P. 400–410. DOI: 10.3103/s0735272722070019

ASSESSMENT OF THE QUALITY OF DETECTION OF A RADAR SIGNAL WITH NONLINEAR FREQUENCY MODULATION IN THE PRESENCE OF A NON-STATIONARY INTERFERING BACKGROUND

Hryzo A. A. – PhD, Associate Professor, Head of the Research Laboratory, Ivan Kozhedub Kharkiv National Air Force University, Kharkiv, Ukraine.

Kostyria O. O. – Dr. Sc., Senior Research, Leading Research Scientist, Ivan Kozhedub Kharkiv National Air Force University, Kharkiv, Ukraine.

Fedorov A. V. – PhD, Researcher, Ivan Kozhedub Kharkiv National Air Force University, Kharkiv, Ukraine.

Lukianchykov A. A. – Senior Researcher, Ivan Kozhedub National Air Force University, Kharkiv, Ukraine.

Biernik Ye. V. – Post-graduate student, Ivan Kozhedub Kharkiv National Air Force University, Kharkiv, Ukraine.

ABSTRACT

Context. Signals with long duration frequency modulation are widely used in radar, which allows increasing the radiated energy without degrading the range resolution and with peak power limitations. Increasing the product of the spectrum width by the radio pulse duration causes the passive interference zone to stretch out from the range, which leads to an interference with a more uniform intensity distribution in space and reduces the potential signal detection capabilities. Real passive obstacles have a non-stationary power distribution in space elements, so the signal reflected from the target can be detected in the gaps of passive obstacles or in areas with a lower level of them, provided that it is assessed (mapping of obstacles) and the detection threshold is adaptively set by space elements. Therefore, it is relevant to conduct research to assess the quality of detection of signals reflected from airborne targets depending on the level of non-stationarity of the interference background.

Objective. The aim of this work is to develop a methodology for assessing the influence of the level of the side lobes of signal correlation functions on the quality indicators of their detection in the presence of a non-stationary interference background of different intensity.

Method. The quality indicators of detection of frequency-modulated signals were studied. The problem of assessing the influence of the level of the lateral lobes of the correlation function on the quality indicators of signal detection against a non-stationary passive interference was solved by determining the parameters of the generalised gamma power distribution of such an interference, depending on the shape of the autocorrelation function of the signal.

Results. It is determined that for a high level of non-stationarity of the initial interference process for all signal models, the potential gain is almost the same and has a maximum value. In the case of reducing the level of non-stationarity of this process, the gain decreases. The traditional linear-frequency modulated signal gives a slightly worse result compared to nonlinear-frequency modulated signals. For all the studied frequency modulation laws, the gain is more noticeable when the requirements for signal detection quality are reduced.

Conclusions. A methodology for estimating the quality indicators of detecting echo signals on an interfering background with varying degrees of non-stationarity is developed. To improve the energy performance of detecting small-sized airborne objects against the background of non-stationary passive interference, it is advisable to use signals with nonlinear frequency modulation and reduce the probability of correct target detection.

KEYWORDS: detection of radar signals, nonlinear frequency modulation; side lobe level, non-stationary interference background.

ABBREVIATIONS

ACF is an autocorrelation function;
CF is a consistent filter;
FMR is a frequency modulation rate;
IPM is an intra-pulse modulation;
LFM is a linear frequency modulation;
LO is a local object;
ML is a main lobe;
MM is a mathematical model;
MSLL is a maximum side lobe level;
NLFM is a nonlinear frequency modulation;
NPI is a nonstationary passive interference;
PI is a passive interference;
Radar is a radar station (Radio Detection and Ranging);
SL is a side lobe;
SLL is a side lobe level.

NOMENCLATURE

\dot{U}_2 is a total complex amplitude of oscillations of a random process, V;
 Δt is a sampling interval, s;
 i, s is a range discrete number (space element);
 \dot{U}_N is a complex amplitude of oscillations of stationary noise, V;
 σ_N^2 is a power of oscillations of stationary noise, dB;
 \dot{U}_{PI} is a complex amplitude of vibration of the PI, V;
 $\sigma_{PI,i}^2$ is a power of oscillations of the PP in the i -th element of space, dB;
 \dot{U}_S is a complex amplitude of oscillations of the useful signal, V;

σ_s^2 is a power of oscillations of the useful signal, Db;
 ξ is a weighting sum, V;
 U is a voltage amplitude at the output of the CF, V;
 D is a probability of correct target detection;
 F is a likelihood of a false alarm;
 ξ_0 is a target detection threshold level, V;
 $W_{N,PI,S}(U)$ is a density of distribution of the additive mixture of noise, PI and signal;
 $W_{N,PI}(U)$ is a distribution density of the additive mixture of noise and PI;
 $W_{PI}(U/\lambda)$ is a conditional density of PI distribution;
 λ is a vector of unknown parameters of the PI;
 $W_{PI}(\dot{U})$ is an unconditional probability density of the PI;
 $p(\lambda)$ is a density of the distribution of the vector of unknown parameters of the PI;
 $D(q^2/\lambda)$ is a partial (fixed) indicator of target detection quality;
 q^2 is a target detection parameter, dB;
 $D(q^2)$ is an average quality of target detection;
 $W(\sigma^2)$ is a generalized power distribution law of the PI;
 α, c are parameters of the power distribution form the PI;
 β is a power distribution scale parameter of the PI;
 $G(z)$ is a gamma function;
 $\langle \cdot \rangle$ is a statistical averaging operation;
 m_1 is an average value of the PI power (first initial moment of distribution), V^2 ;
 Δ is an average power fluctuation range of the PI relative to the average value, dB;
 m_2 is a second initial moment of distribution, V^4 ;
 k is a procedure of the law on power distribution of the PI;
 q_i^2 is a parameter of target detection in the i -th element of space, dB;
 ξ_i is a modular value of the weight sum in the i -th element of the space, V;
 Δ_{init} is an initial range of PI power fluctuations relative to the average value, dB.

INTRODUCTION

Attempts by manufacturers of modern radars to improve the characteristics of target detection and tracking in difficult signal and interference conditions while expanding the range of tasks they perform have led to the use of active digital antenna arrays [1–5]. The required values of mass, dimensions, reliability indicators of such systems and their functional flexibility can be achieved by using solid-state transmitting devices [1, 4, 5].

This technology has a significant disadvantage, namely, the limitation of the peak power of the sensing signals, which makes it difficult to achieve the required value of the maximum detection range. In order to overcome this drawback, the duration of signals is increased to ensure the required range of the radar, which in turn leads to an undesirable deterioration in range resolution. To overcome this contradiction, various types of IPM are used, usually linear, nonlinear frequency modulation and phase manipulation [6–8].

An analysis of the technical characteristics of radars available on the market and used by end users for airspace reconnaissance shows that they are all built on the above principle, with the actual differences being in the types of sensing signals used [7, 8].

A mandatory stage of processing complex sensing signals in the radar receiver is their compression into CF, which maximizes the signal-to-noise ratio and provides the required range resolution, which is determined by the width of the main peak of the filter response. The time-dependence of the CF response intensity is determined by its ACF, which, in addition to the ML, has SLs that can be quite long in time. In the case of detecting such compressed signals against the background of intrinsic noise, the influence of their SLs can be neglected. A different situation is observed when detecting such signals in ground surveillance radars, which almost always conduct airspace reconnaissance against the background of the SLs caused by the reflection of sounding signals from the earth's surface [9].

In the case of powerful PIs, the absolute level of the ACF SL will be significant. This leads to a stretching of the area of the PI, i.e., to the appearance of PIs in areas previously free of them, the overall level of non-stationarity of the interference background decreases. The targets previously observed in the interference gaps will be masked by the resulting background of the SL of compressed signals reflected from the PI from other areas of the range. In addition, when superimposing the SL of compressed signals reflected from long-range PIs that occupy several range discrete points, their level can reach significant values. This, in addition to an increase in the overall level of the interfering background, leads to the possibility of a multimode spectrum if reflections from different sources have different Doppler frequency shifts [6, 9].

Complex signals with NLFM have been widely used, and various authors have proposed a number of MMs of such signals with a reduced SL level of their ACF for practical application [10–33]. It seems expedient to evaluate the effect of the SL level of ACF on the quality indicators of signal detection against the background of non-stationary power PIs.

The object of study is the process of detecting NLFM signals in the background of the NPI.

The subject of study is the quality indicators for detecting NLFM signals on the background of NPI.

The purpose of the work is to development of a methodology for assessing the influence of the SLs of the ACF of complex signals on the quality indicators of their detection in the presence of a non-stationary background.

1 PROBLEM STATEMENT

We assume that a random process \dot{U}_Σ is input to the receiver sequentially in time. Then the process is sampled in time with an interval Δt . From the absolute value of time $t=i \cdot \Delta t$, we proceed to indexing by the reference number i (the number of the range discrete).

We assume that the process in each i -th range discrete is the sum of independent normally distributed oscillations of the intrinsic stationary noise \dot{U}_N with power $\sigma_N^2 = 1$, passive interference \dot{U}_{PI} with known power $\sigma_{PI,i}^2$, and possibly a useful signal \dot{U}_S with power σ_S^2 :

$$\dot{U}_\Sigma = \dot{U}_N + \dot{U}_{PI} + \dot{U}_S.$$

We also assume that the random samples of the processes have a Rayleigh distribution of amplitude and a uniform distribution of the initial phase. Such a model corresponds to a signal reflected from a large number of randomly located, independently reflecting, equivalent shiny points and is appropriate for most practical situations.

The optimal detection algorithm is reduced to calculating the likelihood ratio or its logarithm [5, 7], which are monotonically increasing functions and describe the dependence of the likelihood ratio on the modular value of the normalized weighting sum ξ . Taking this into account, the optimal detection algorithm can be implemented by comparing the modular value of the weighting sum ξ with its threshold. Under the accepted conditions, the weighting operation can be performed by an optimal filter, and the value of the modular value of the weighting sum ξ is found as the amplitude of the voltage U at the filter output, which is subject to comparison with the threshold.

The conditional probability of correct detection D and false alarm F are most often used as indicators of detection quality. The detection characteristics D and F can be calculated as follows [5, 7]:

$$F = \int_{\xi_0}^{\infty} U \cdot \exp\left(-\frac{U^2}{2}\right) dU = \exp\left(-\frac{\xi_0^2}{2}\right), \quad (1)$$

where the threshold level comes from

$$\xi_0 = \sqrt{2 \cdot \ln\left(\frac{1}{F}\right)}.$$

The conditional probability of correct detection is described by the expression:

$$D = \int_{\xi_0}^{\infty} W_{N,PI,S}(U) dU = \int_{\xi_0}^{\infty} l(U) \cdot W_{N,PI}(U) dU. \quad (2)$$

Under real conditions, the distribution of the power of the PI by space elements is arbitrary and the value of their power $\sigma_{PI,i}^2$ is an unknown parameter with a certain distribution law.

Under such a priori uncertainty in the probability characteristics of signals and interference, the probability density of, for example, interference will be described by the conditional distribution density $W_{PI}(U/\lambda)$. Such a problem belongs to the class of problems with parametric a priori uncertainty [34]. The general approach to solving such problems [35] is to use the full probability formula to move from the conditional probability density $W_{PI}(U/\lambda)$ to the unconditional probability density $W_{PI}(U)$:

$$W_{PI}(U) = \int_{(\lambda)} W_{PI}(U/\lambda) \cdot p(\lambda) d\lambda. \quad (3)$$

The a priori uncertainty is associated with signal detection in different elements of space, where the parameters of the interference background may be different.

The combination of solutions (1), (2) in any discrimination elements with similar interference statistics allows obtaining, in fact, partial (fixed) quality indicators $D(q^2/\lambda)$, where q^2 determines the signal/interference+noise ratio. Moving to the average indicators $D(q^2)$ (similar to (3)), we obtain:

$$D(q^2) = \int_{(\lambda)} D(q^2/\lambda) \cdot p(\lambda) d\lambda, \quad (4)$$

similarly, the detection threshold should be found from equation.

$$F = \int_{(\lambda)} F(\lambda) \cdot p(\lambda) d\lambda.$$

If we set the distribution density to $p(\lambda)$, the solution of a statistical problem under a priori uncertainty is reduced to solving it under known a priori conditions.

When choosing the type of distribution density of parameter λ , restrictions are imposed that are determined by the physical content and nature of the process, which narrows the range of acceptable distributions. Such restrictions may include, for example, the permissible range of parameter change, its mean value, variance, etc.

To move from the partial (conditional) to the average value of the probability of correct detection D , it is necessary to determine the distribution law that describes the statistical characteristics of the fixed parameter $\sigma_{PI,i}^2$. The generalized gamma distribution is used as the law of PI power distribution [35–39].

Further, we will take into account that in the case of a complex signal with IPM, the dependence of the CF response intensity will be determined by its ACF. The signal at each point in space will be an overlay of CF responses from neighboring range discrete samples, the number of components will be determined by the duration of the ACF signal, and their intensity – by the corresponding SL level. Obviously, the lower the SL level of the ACF, the less influence they have in the overall process, which in turn leads to a lower degree of reduction of the non-stationarity of the interference background and, accordingly, to better conditions for detecting signals reflected from targets against its background.

We will assume that the initial distribution of the PI intensity over the elements of space is determined by a gamma distribution with certain parameters, and in the case of superposition of SL reflections from different areas, taking into account the properties of the gamma function family, the total PI intensity distribution will also have a gamma distribution, but with different parameters.

Thus, the problem of assessing the effect of the SL of the ACF on the quality of signal detection against the background of the NPI can be solved by determining the parameters of the generalized gamma distribution of the PI power depending on the ACF signal shape and then averaging the partial detection rates in individual detection elements according to the law of PI intensity distribution, the parameters of which correspond to the selected IPM.

2 REVIEW OF THE LITERATURE

In [40], it is noted that radars that use compression methods for complex signals face the need to resolve conflicting requirements for pulse duration. Signals of short duration have good resolution, and long pulses are preferable in terms of ensuring maximum detection range. It is noted that when the pulse is compressed, unwanted SLs are present in the CF response, and their combined effect leads to inaccurate operation of detection systems. The authors of [41, 42] also point out that the use of complex signals avoids the problem of peak power limitation, but the presence of SLs in the CF output leads to the effect of masking weak signals, and it is noted that this effect can be quite significant and needs to be taken into account. To overcome this, algorithms for stabilizing the level of false alarms are proposed. A similar effect of SL masking weak targets by the total background of the PI of the compressed signal in the presence of powerful reflections from the PI is noted in [43]. To prevent this, methods of reducing the SLL of the compressed signal are proposed, which are based on the time analysis of the received signal and are implemented by selectively weighting the

signal samples. In [44], it is proposed to use the median filtering procedure for this purpose. The common feature of the reviewed works is that the authors did not study the influence of the shape and SLL of the ACF signals on the detection quality indicators and do not provide their numerical values.

Papers [35–39] consider the family of generalized gamma distributions and show that the generalized gamma distribution is a continuous probability distribution with three parameters. It is a generalization of the two-parameter gamma distribution and is used for parametric approximation of data sets, including approximation of experimental signal power distributions.

The authors of [10–33] proposed a number of IPN laws that allow obtaining a low SLL of the ACF, so it is advisable to develop a method for assessing the degree of influence of the shape and SLL of the ACF on the quality indicators of signal detection against the background of the NPI.

3 MATERIALS AND METHODS

Let us consider the peculiarities of using the generalized power distribution law for the statistical description of the characteristics of the PI.

If it is desirable to use the experimental intensity distributions of passive interference $\sigma_{PI,i}^2$ in the input model, it is convenient to use the generalized power distribution law proposed for the first time in [36] (for simplicity, $\sigma_{PI,i}^2 = \sigma^2$ is assumed):

$$W(\sigma^2) = \frac{c \cdot \beta^\alpha \cdot (\sigma^2)^{\alpha-1}}{G\left(\frac{\alpha}{c}\right)} \cdot \exp\left(-\beta^c \cdot (\sigma^2)^c\right), \quad (5)$$

provided that $\alpha > 0$, $c > 0$, $\beta > 0$.

This distribution generalizes a number of single-mode probabilistic models of radar reflectivity density distribution: log-normal, Weibull, K-distribution, exponential, Hoyt, Nakagami, Rayleigh, Rice, Beckman, uniform.

In practical applications, the power distribution of a power PI in space $\sigma_{PI,i}^2$ is often characterized by two parameters: the average power value $m_1 = \langle \sigma_{PI,i}^2 \rangle$ and the relative value Δ , which characterizes the average range of power fluctuations relative to the average value m_1 :

$$\Delta = \frac{\sqrt{\langle (\sigma_{PI,i}^2 - \langle \sigma_{PI,i}^2 \rangle)^2 \rangle}}{m_1} = \frac{\sqrt{m_2 - m_1^2}}{m_1}. \quad (6)$$

Usually, the value of m_1 , referred to the intrinsic noise level, and the value Δ (6) are expressed in decibels.

For practical use, we will write the gamma density (5) in terms of the parameters m_1 and Δ . The formula for the initial moment of the k -th order of the distribution (5) is as follows [35, 36]:

$$m_k = G\left(\frac{\alpha+k}{c}\right) / \left(\beta^k \cdot \Gamma\left(\frac{\alpha}{c}\right)\right). \quad (7)$$

In the case of $k=1$, this relation is converted into an expression for the mean power of a non-stationary random process $\langle \sigma_{PI,i}^2 \rangle = m_1$.

To approximate the experimental distributions of radar reflections power (selection of appropriate values of parameters α, c, β), the corresponding moments are equated, but in practice it is convenient to use the introduced values of m_1 and Δ , which is essentially equivalent to using the first two initial moments.

The probability distribution density (5) is a function of three parameters, and the number of moments to be compared is two, one parameter of the distribution law can be fixed: $c=1$.

In this case, the density (5) will be transformed into the density of the gamma distribution [35]:

$$W(\sigma^2) = \frac{\beta^\alpha \cdot (\sigma^2)^{\alpha-1}}{G(\alpha)} \cdot \exp(-\beta \cdot \sigma^2). \quad (8)$$

For the practical use of such a probabilistic model of fluctuations in the power of the PP, we define the parameters of the distribution density (8) α and β through the parameters m_1 and 3 (6).

Based on (7), taking into account that $G(\alpha+1) = \alpha \cdot G(\alpha)$, we obtain:

$$m_1 = \frac{G(\alpha+1)}{\beta \cdot G(\alpha)} = \frac{\alpha}{\beta},$$

$$\Delta = \frac{\sqrt{m_2 - m_1}}{m_1} = \sqrt{\frac{m_2}{m_1} - 1} = \sqrt{\frac{G(\alpha+2) \cdot G(\alpha)}{G(\alpha+1)^2} - 1} = \frac{1}{\sqrt{\alpha}}.$$

Invisible parameters α and β :

$$\alpha = \Delta^{-2},$$

$$\beta = \frac{\alpha}{m_1} = \frac{1}{m_1 \cdot \Delta^2}.$$

Finally, we write the density of the distribution (8) in the form (for simplicity, the index at m_1 is omitted since $m_1 \rightarrow m$):

$$W(\sigma^2, m, \Delta) = \frac{(\sigma^2)^{\Delta^2-1}}{(m \cdot \Delta^2)^{\Delta^2} \cdot G(\Delta^2)} \cdot \exp\left(-\frac{\sigma^2}{m \cdot \Delta^2}\right). \quad (9)$$

The use of this law as a probabilistic model allows for a wide range of fluctuations in the power of PI Δ while maintaining its average value m .

Let's estimate the effect of the level of obstacle non-stationarity on the quality of useful signal detection.

To simplify the calculations (by fixing the false alarm level F) under the influence of a non-stationary interference background, it is advisable to use the modular value of the weighting sum ξ_i as a sufficient statistic. It is known that in this case, the response in the absence of a useful signal and Gaussian interference has a Rayleigh distribution with unit variance $\sigma_{PI,i}^2$.

Assuming that each i -th element of the space has its own modular value of the weighting sum ξ_i , for a signal with a random amplitude distributed according to Rayleigh's law and a uniformly distributed initial phase, we have [5, 7]:

$$D_i = F \frac{1}{1 + \frac{q_i^2}{2}}, \quad (10)$$

and q_i^2 defines the signal/interference+noise ratio in the i -th element of space. For the specified conditions, the detection parameter is equal to [5, 7]:

$$q_i^2 = 2 \cdot \frac{\sigma_S^2}{\sigma_N^2 + \sigma_{PI,i}^2}. \quad (11)$$

In another similar (s -th) element of space, the detection parameter q_s^2 will differ only by a different, but known for this element, intensity value $\sigma_{PI,s}^2$.

To move from the partial (conditional) to the average value of the probability of correct detection D , it is necessary to set the distribution laws that describe the statistical characteristics of parameter $\sigma_{PI,i}^2$.

As a probabilistic model of PI power fluctuations, we use the generalized gamma power distribution (5), (8), (9). Then, the average value of the probability of correct detection, taking into account (5), (9), (10) and (11), is written as follows:

$$D = \int_0^\infty F \left(1 + \frac{\sigma_c^2}{\sigma_N^2 + \sigma_{PI,i}^2}\right)^{-1} \cdot W(\sigma_{PI,i}^2, m, \Delta) d\sigma_{PI,i}^2. \quad (12)$$

Expression (12) makes it possible to estimate the efficiency of detecting a useful echo signal under different variants of input influences. By varying the value of Δ at a fixed level of the average power of the PI m , it is possible to obtain detection quality indicators corresponding to different degrees of non-stationarity of the interfering

background. For $\Delta \rightarrow -\infty$, we obtain the detection quality indicators corresponding to a stationary interference background.

4 EXPERIMENTS

Let us evaluate the effect of the SL of the ACF on the quality of echo detection against the background of the NPI in terms of reducing the relative level of fluctuations Δ (6) of their intensity distribution in space.

In Figure 1, the solid line shows the energy relief of the PI by range, and the dashed line shows the result of modelling the response of the signal compression filter with an LFM, the first SLs of the ACF which are -13 dB relative to the level of the ML, and the duration is 120 range samples. To preserve the energy relations, the processes are normalized in such a way as to ensure the equality of the average value of the PI intensity.

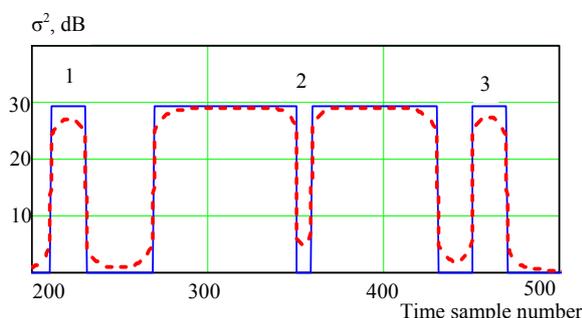


Figure 1 – Passive interference intensity distribution by range sweep

The abscissa axis is the number of range samples, and the ordinate axis is the signal intensity, referred to the intrinsic noise level in decibels. For simplicity, it is assumed that the intensity of the reflections from the LO is the same and is 30 dB (solid line in Figure 1). Figure 1 shows the situational marks 1, 2, 3, which will be discussed further.

Situation 1 corresponds to the presence of a separately located point reflection from the LO, when the distance to the next area occupied by the PI exceeds the duration of the probing signal. Situations 2, 3 correspond to the presence of closely spaced areas of the PI when the distance between them is less than the duration of the probing signal.

When observing a weak target echo, the absolute value of the compressed signal SLL will be lower than the level of the receiver's own noise, but in the case of powerful reflections from the PI, this level will be significant. For example, for the selected case, when the PI with a power of 30 dB is affected, the absolute value of the SLL will be 15...17 dB relative to the level of the receiving path's own noise. This leads to the stretching of the area of the PI action, i.e., to the appearance of PIs in areas previously free of them, in addition, when the SL superimposes compressed signals reflected from long-range PIs occupying several range discrete values, their level can reach significant values (situation 1, 2, 3 areas to the left

and right of the PI location). Situation 2 differs from Situation 1 in that the SLs of compressed signals from different areas occupied by the PI overlap, which leads to an increase in the overall level of interference background in previously free areas and the possibility of a multimode spectrum if reflections from different sources have different Doppler frequency shifts. Situation 3 corresponds to the case of the end of the range of the PI, we see that the area occupied by the PI is extended to the duration of the sensing signal, the additional increase is 120 range discrete.

Figure 1 is for illustrative purposes only and confirms that the signal at each point in space will be an overlay of CF responses from neighboring range discrete samples, the number of components will be determined by the duration of the ACF signal, and their intensity will be determined by the corresponding SL level. Obviously, the lower the level of the SL of the ACF, the less influence they have in the overall process, which in turn leads to a lower degree of reduction of the interference background non-stationarity and, accordingly, better conditions for detecting echo signals against it.

Next, let's consider the quality indicators of echo detection for three models of signals with IPM.

We will consider the MM of the NLFM signals introduced in [10–13], consisting of two and three LFM fragments, which compensate for jumps in instantaneous frequency and phase that occur at the moments of change in the FMR when moving from one signal fragment to another. The use of these models ensures the disappearance of the distortion of the shape of their spectra, which leads to a decrease in the MSLL of their ACF. To compare the results, we will use a classical LFM signal of the same duration.

We will evaluate the impact of the SLL of the ACF of the selected signal models by simulation. We will use a test implementation of the NPI intensity distribution by range elements with known parameters m and Δ . The output sequence will be determined by summing the CF responses from the corresponding range discrete elements. The CF response will be determined by the intensity of the signal and its ACF value at certain moments of time; for ease of use, the ACF characteristics of the signals to be compared are given in Table 1.

Table 1 shows the values of the MSLL and the decay rate of the SLL of their ACFs of two- and three-fragment NLFM signals (models 1 and 2, respectively) and the classical LFM signal (model 3). The values of the relative level of fluctuations Δ at the output of the CF of the corresponding type of signal for three variants of the value of the relative level of fluctuations of the initial process Δ_{init} are presented. The average power of the processes is the same $m=10$ dB. Thus, it is possible to estimate the reduction of the degree of non-stationarity of the initial process and the potential losses from this. The signals of all models have the same duration.

The ACF signals from Table 1 are shown in Figures 2, 3, 4.

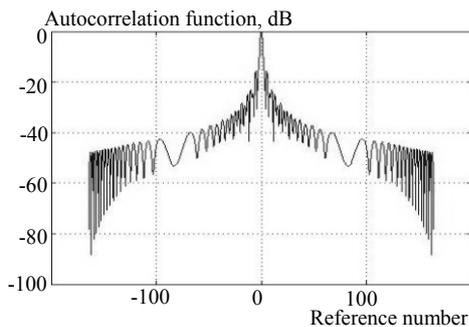


Figure 2 – ACF of a two-fragment NLCM signal (model No.1)

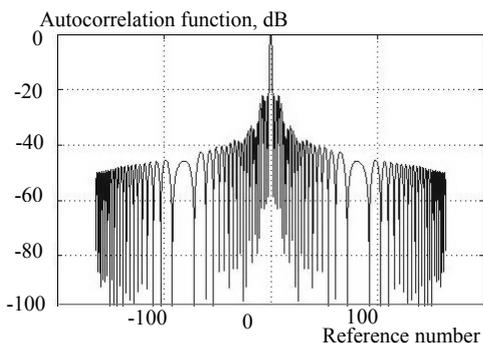


Figure 3 – ACF of a three-fragment NLCM signal (model No.2)

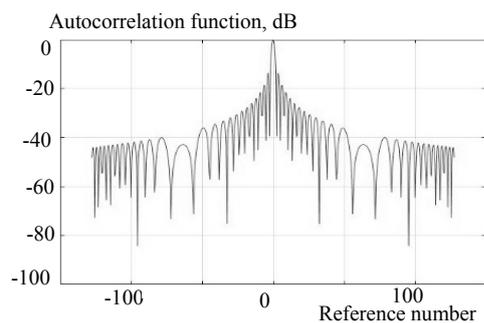


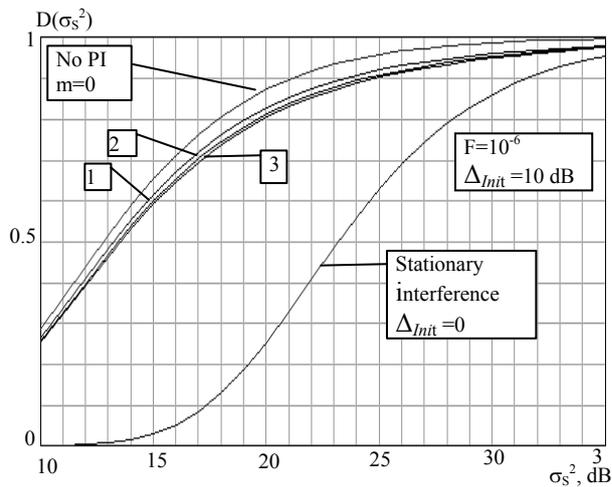
Figure 4 – ACF of the LFM signal (model No.3)

Table 1 – Characteristics of the ACF signal models and simulation results

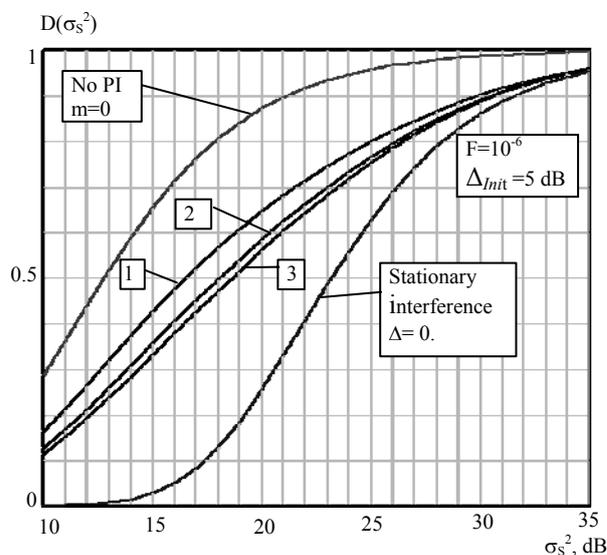
No	MM of the signal	MSLL, dB	Decrease in MRBP, dB/dec	Δ , dB		
				$\Delta_{init} = 10.23$	$\Delta_{init} = 5.03$	$\Delta_{init} = 3.05$
1.	LFM-LFM	-15.81	20.00	8.89	3.45	1.03
2.	LFM-LFM-LFM	-22.00	19.30	9.78	4.52	2.43
3.	LFM	-13.00	22.00	8.78	3.33	0.92

5 RESULTS

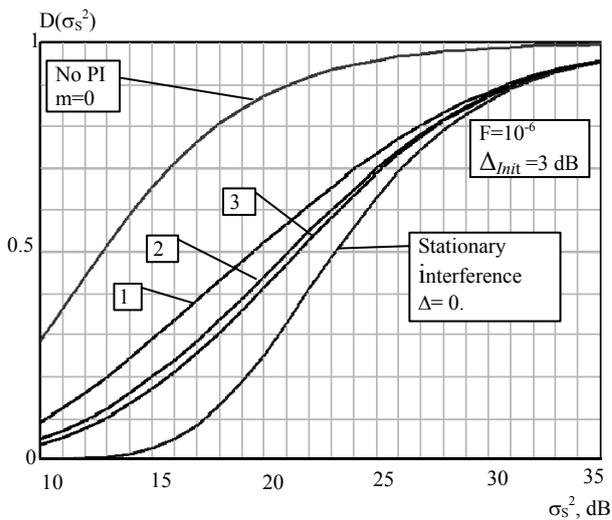
The results of the evaluation of the quality indicators of detecting fingerprints against the background of PI with different levels of non-stationarity according to (9), (12) for the signals of the selected models (Table 1) are shown in Figures 5 a,b,c as a function of the conditional probability of correct detection on the signal intensity.



a) background instability level $\Delta_{init} = 10$ dB



b) background instability level $\Delta_{init} = 5$ dB



c) background instability level $\Delta_{init} = 3$ dB

Figure 5 – Detection curves for different signal models

The corresponding model numbers are shown in the figures and correspond to the numbering in Table 1. The curves are plotted for the case of average interference power $m=10$ dB and conditional false alarm probability $F=10^{-6}$. For comparison, two extreme cases are presented – no interference and the presence of a stationary interference with an intensity of 10 dB. Model No. 3 corresponds to the use of a traditional LFM signal.

6 DISCUSSION

The analysis of the figures shows that a higher level of non-stationarity of the initial process corresponds to a greater gain in the signal/interference+noise ratio that can potentially be obtained. In Figure 5a, the level of non-stationarity of the initial process is $\Delta_{\text{init}} = 10$ dB, for all signal models (curves 1, 2, 3), the potential gain is almost the same and amounts to 9 dB. This can be explained by the fact that with such a significant level of non-stationarity, virtually all the interference power is concentrated in a few range areas, while other areas are free from its influence. Under these conditions, the shape of the ACF and its MSLL does not play a major role, in fact, it is a matter of detecting point targets against the background of their own noise. The difference in the curves is no more than 0.3 dB.

With a decrease in the level of non-stationarity of the initial process, the influence of the ACF shape and its MSLL becomes more significant. This can be explained by the fact that the areas occupied by the PIs begin to occur more frequently, and the effect of the overlapping of the SL responses from different range areas becomes more significant. For the curves corresponding to models 1 and 2, the gain is 5–6 dB for the case of $\Delta_{\text{init}} = 5$ dB and 2–3 dB for $\Delta_{\text{init}} = 3$ dB. The decrease in gain for a lower level of background non-stationarity is explained by its approach to the stationary one.

It should be noted that the traditional LFM yields a worse result by 0.5–0.7 dB compared to two- and three-fragment NLFM signals in terms of providing a potential energy gain when detecting targets against a non-stationary interference background. This is as it should be, given the higher integral SL level of its ACF.

For all cases, the gain is more noticeable with low requirements for signal detection quality ($D<0.5$) and can be 6–7 dB.

CONCLUSIONS

The paper investigates the quality indicators of detection of FM radar signals against the background of NPI formed by time-superimposed CF responses caused by the SLs of the ACF of these signals.

The scientific novelty lies in the development of a methodology for assessing the quality of detection of echo signals against an interfering background with varying degrees of non-stationarity, the essence of which, unlike the known ones, is to determine the parameters of the generalized gamma distribution of the power of the NPI depending on the shape of the ACF signal. For this purpose, we use the averaging of partial detection rates in

individual detection elements according to the law of the intensity distribution of the PP and taking into account the parameters of the law of the FM signals. This technique is applied to the study of two- and three-fragment NLFM signals and a signal with LFM.

The practical significance of The conclusion of the obtained results is the possibility of using the proposed methodology for assessing the quality indicators of detecting echo signals on an interference background with varying degrees of non-stationarity for comparative analysis of NLFM signals with different FM laws. The results obtained allow us to conclude that in order to improve the energy performance of detecting small-sized targets, it is advisable to reduce the probability of correct target detection below the value of 0.5.

Prospects for further research it is planned to study the previously developed MMs of two- and three-fragment NLFM signals for the expediency of their application in various radar models to optimise their operating modes.

ACKNOWLEDGEMENTS

We thank the management of Ivan Kozhedub Kharkiv National Air Force University for the opportunity to conduct scientific research.

REFERENCES

1. Richards M. A., Scheer J., Holm W. A. Principles of modern radar. Basic principles. Sci Tech Pub, 2010, 924 p.
2. Blackman S. S., Popoli R. F. Design and analysis of modern tracking systems. Boston, London, Artech House, 1999, 1230 p.
3. You H., Jianjuan X., Xin G. Radar data processing with applications. House of electronics industry publishers, 2016, 536 p. DOI:10.1002/9781118956878
4. Melvin W. L., Scheer J. A. Principles of Modern Radar. Vol. II: Advanced Techniques. Sci Tech Publishing, 2013, 846 p.
5. Barton D. K. Radar system analysis and modeling. Boston, London, Artech House Publishers, 2004, 566 p.
6. Van Trees H. L. Detection, estimation, and modulation theory. Part III: Radar-Sonar processing and Gaussian signals in noise. John Wiley & Sons, Inc., 2001, 643 p. DOI: 10.1002/0471221090
7. Levanon N., Mozeson E. Radar signals. New York, John Wiley & Sons, Inc., 2004, 403 p.
8. Cook C. E., Bernfeld M. Radar signals: An introduction to theory and application. Boston, Artech House, 1993, 552 p.
9. McDonough R. N., Whalen A. D. Detection of signals in noise (2nd. Ed.). San Diego, Academic Press, Inc., 1995, 495 p.
10. Kostyria O. O., Hryzo A. A., Dodukh O. M. et al. Mathematical model of a two-fragment signal with a non-linear frequency modulation in the current period of time, *Visnyk NTUU KPI Seriia – Radiotekhnika Radioaparaturbuduvannia*, 2023, Vol. 92, pp. 60–67. DOI:10.20535/RADAP.2023.92.60-67.
11. Kostyria O. O., Hryzo A. A., Dodukh O. M. et al. Improvement of mathematical models with time-shift of two- and three-fragment signals with non-linear frequency modulation, *Visnyk NTUU KPI Seriiia – Radiotekhnika Radioaparaturbudu-*

- vannia, 2023, Vol. 93, pp. 22–30. DOI: 10.20535/RADAP.2023.93.22-30.
12. Kostyria O. O., Hryzo A. A., Khizhnyak I. A. et al. Implementation of the method of minimizing the side lobe level of autocorrelation functions of signals with nonlinear frequency modulation, *Visnyk NTUU KPI Serii – Radiotekhnika Radioaparotobuduvannia*, 2024, Vol. 95, pp. 16–22. DOI: 10.20535/RADAP.2024.95.16-22
 13. Kostyria O. O., Hryzo A. A., Solomonenko Yu. S. et al. Mathematical model of shifted time of combined signal as part of fragments with linear and quadratic frequency modulation, *Visnyk NTUU KPI Serii – Radiotekhnika Radioaparotobuduvannia*, 2024, Vol. 97, pp. 5–11. DOI:10.20535/RADAP.2024.97.5-11
 14. Adithya V. N., Elizabeth R. D., Kavitha C. Modified radar signal model using NLFM, *International Journal of Recent Technology and Engineering (IJRTE)*, 2019, Vol. 8, Issue 2S3, pp. 513–516. DOI: 10.35940/ijrte.b1091.0782s319
 15. Anoocha Ch., Krishna B. T. Peak sidelobe reduction analysis of NLFM and improved NLFM radar signal with non-uniform PRI, *Aiub Journal of Science and Engineering (AJSE)*, 2022, Vol. 21, Issue 2, pp. 125–131.
 16. Zhaoa Yu., Ritchieb M., Lua X. et al. Non-continuous piecewise nonlinear frequency modulation pulse with variable sub-pulse duration in a MIMO SAR radar system, *Remote Sensing Letters*, 2020, Vol. 11, Issue 3, pp. 283–292. DOI: 10.1080/2150704X.2019.1711237
 17. Van-Zyl A. C., Wiehahn E. A., Cillers J. E. et al. Optimized multi-parameter NLFM pulse compression waveform for low time-bandwidth radar, *International Conference on Radar Systems*, 2022, pp. 289–294. DOI: 10.1049/icp.2022.2332.
 18. Fan Z., Meng H. Coded excitation with nonlinear frequency modulation carrier in ultrasound imaging system, *2020 IEEE Far East NDT New Technology & Application Forum*. Kunming, Yunnan province, China, conference paper, IEEE, 2020, pp. 31–35. DOI: 10.1109/FENDT50467.2020.9337517.
 19. Adithya V. N., Elizabeth R. D., Kavitha C. Performance analysis of NLFM signals with Doppler effect and background noise, *International Journal of Engineering and Advanced Technology (IJEAT)*, 2020, Vol. 9, Issue 3, pp. 737–742. DOI: 10.35940/ijeat. B3835.029320
 20. Zhang Y., Deng Y., Zhang Z. et al. Analytic NLFM waveform design with harmonic decomposition for synthetic aperture radar, *IEEE Geoscience and Remote Sensing Letters*, 2022, Vol. 4, Article no 4513405. DOI: 10.1109/lgrs.2022.3204351
 21. Xie Q., Zeng H., Mo Z. et al. A two-step optimization framework for low sidelobe NLFM waveform using fourier series, *IEEE Geoscience and Remote Sensing Letters*, 2022, Vol. 19, Article no 4020905. DOI: 10.1109/LGRS.2022.3141081
 22. Zhuang R., Fan H., Sun Y. et al. Pulse-agile waveform design for nonlinear FM pulses based on spectrum modulation, *IET International Radar Conference*, 2021, pp. 964–969. DOI: 10.1049/icp.2021.0700.
 23. Shuyi L., Jia Y., Liu Y. et al. Research on ultra-wideband NLFM waveform synthesis and grating lobe suppression, *Sensors*, 2022, No. 24, Article number 9829. DOI: 10.3390/s22249829
 24. Swiercz E., Janczak D., Konopko K. Estimation and classification of NLFM signals based on the time-chirp, *Sensors*, 2022, Vol. 22, Issue 21, Article no 8104. DOI: 10.3390/s22218104
 25. Saleh M., Omar S.-M., Grivel E. et al. A variable chirp rate stepped frequency linear frequency modulation waveform designed to approximate wideband non-linear radar waveforms, *Digital Signal Processing*, 2021, Vol. 109, Article no 102884, 19 p. DOI: 10.1016/j.dsp.2020.102884
 26. Xu Z., Wang X., Wang Y. Nonlinear frequency-modulated waveforms modeling and optimization for radar applications, *Mathematics*, 2022, Vol. 10, P. 3939. DOI: 10.3390/math10213939.
 27. Widyantara M. R., Suratman S.-F. Y., Widodo S. et al. Analysis of non linear frequency modulation (NLFM) waveforms for pulse compression radar, *Jurnal Elektronik dan Telekomunikasi*, Vol. 18, Issue 1, pp. 27–34. DOI: 10.14203/jet.v18.27-34
 28. Ping T., Song C., Qi Z. et al. PHS: A pulse sequence method based on hyperbolic frequency modulation for speed measurement, *International Journal of Distributed Sensor Networks*, 2024, Vol. 2024, Article no 6670576, 11 p. DOI:10.1155/2024/6670576
 29. Cheng Z., Sun Z., Wang J. et al. Magneto-Acousto-Electrical Tomography using Nonlinearly Frequency-Modulated Ultrasound, *Phys. Med. Biol.*, 2024, Vol. 69(8), PMID: 38422542. DOI:10.1088/1361-6560/ad2ee5
 30. Singh A. K., Bae K.-B., Park S.-O. NLFM pulse radar for drone detection using predistortion technique, *Journal of Electromagnetic Waves and Applications*, 2021, Vol. 35, pp. 416–429.
 31. Septanto H., Sudjana O., Suprijanto D. A novel rule for designing tri-stages piecewise linear NLFM chirp, *2022 International Conference on Radar, Antenna, Microwave, Electronics, and Telecommunications (ICRAMET), Bandung, 6–7 December 2022*. Indonesia, IEEE, 2022, pp. 62–67. DOI: 10.1109/ICRAMET.56917.2022.9991201
 32. Kostyria O. O., Hryzo A. A., Dodukh O. M. Combined two-fragment radar signals with linear and exponential frequency modulation laws, *Systems of Arms and Military Equipment*, 2024, № 4 (76), pp. 58–64. DOI: 10.30748/soivt.2023.76.06
 33. Kostyria O. O., Hryzo A. A., Dodukh O. M. Synthesis time-shifted mathematical model of a combined signal with linear and cubic frequency modulation, *Information Processing Systems*, 2024, № 1 (176), pp. 73–81. DOI:10.30748/soi.2024.176.09
 34. Kaliuzhnyi M., Semenets V., Orlenko V. et al. Synthesis of a single-channel device for receiving and primary signal processing under conditions of structural-parametric a priori uncertainty, *2022 IEEE 2nd Ukrainian Microwave Week (UkrMW)*. Ukraine, IEEE, 2022, pp. 573–578. DOI: 10.1109/UkrMW58013.2022.10037164
 35. Feller W. An introduction to probability theory and its applications, Volume 2 / W. Feller. – New York: John Wiley & Sons Inc., 1991. – 683 p.
 36. Stacy E. W. A generalization of the gamma distribution, *Annals of Mathematical Statistics*, 1962, Vol. 33(3), pp. 1187–1192. DOI: 10.1214/aoms/1177704481
 37. Alzaatreh A., Lee C. & Famoye F. Family of generalized gamma distributions: Properties and applications, *Hacettepe Journal of Mathematics and Statistics*, 2016, Vol. 45(3), pp. 869–886. DOI:10.15672/HJMS.20156610980
 38. Kiche J., Ngesa O. & Orwa G. On generalized gamma distribution and its application to survival data, *International Journal of Statistics and Probability*, 2019, Vol. 8(5), pp. 85–102. DOI: 10.5539/ijsp.v8n5p85
 39. Sportouche H., Nicolas J.-M., Tupin F. Mimic capacity of fisher and generalized gamma distributions for high-

- resolution SAR image statistical modeling, *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 2017, Vol. 10, Issue 12, pp. 5695–5711.
40. Thakur A. and Saini D. S. Transmit waveform design to avoid target masking in pulse compression radar, *2020 International Conference on Communication and Signal Processing (ICCSP)*, Chennai, India, 2020, pp. 97–101. DOI:10.1109/ICCSP48568.2020.9182221
41. Zhu X., He F., Ye F. et al. Sidelobe suppression with resolution maintenance for SAR images via sparse representation, *Sensors (Basel)*, 2018, Vol. 18(5), Article no1589. DOI: 10.3390/s18051589; PMID: 29772743; PMCID: PMC5982680.
42. Haliloglu O., Yilmaz A. O. Successive target cancelation in pulse compression radars, *2007 IEEE Radar Conference, Waltham, MA, USA, IEEE*, 2007, pp. 885–890. DOI: 10.1109/RADAR.2007.374336
43. Nepal R., Li Z., Zhang Y. et al. A simulation of study of the impact of surface clutter on spaceborne precipitation radar sensor, *In Proceedings of the 36th Conference on Radar Meteorology*. Breckenridge, CO, USA: IEEE, 2013, pp. 16–20.
44. Maki M., Kobori T. Construction of three-dimensional weather radar data from volcanic eruption clouds, *MethodsX*, 2021, Vol. 8, Article no 101535. DOI: 10.1016/j.jvolgeores.2021.107178

Received 12.11.2024.

Accepted 26.01.2025.

УДК 621.396.962

ОЦІНКА ЯКОСТІ ВИЯВЛЕННЯ РАДІОЛОКАЦІЙНОГО СИГНАЛУ З НЕЛІНІЙНОЮ ЧАСТОТНОЮ МОДУЛЯЦІЄЮ ЗА НАЯВНОСТІ НЕСТАЦІОНАРНОГО ПЕРЕШКОДОВОГО ФОНУ

Гризо А. А. – канд. техн. наук, доцент, начальник НДІ Харківського національного університету Повітряних Сил імені Івана Кожедуба, Харків, Україна.

Костира О. О. – д-р техн. наук, старший науковий співробітник, провідний науковий співробітник Харківського національного університету Повітряних Сил імені Івана Кожедуба, Харків, Україна.

Федоров А. В. – д-р філос., науковий співробітник Харківського національного університету Повітряних Сил імені Івана Кожедуба, Харків, Україна.

Лук'янчиков А. А. – старший науковий співробітник Харківського національного університету Повітряних Сил імені Івана Кожедуба, Харків, Україна.

Бернік Є. В. – ад'юнкт Харківського національного університету Повітряних Сил імені Івана Кожедуба, Харків, Україна.

АНОТАЦІЯ

Актуальність. В радіолокації широке застосування знайшли сигнали з частотною модуляцією великої тривалості, що дозволяє без погіршення роздільної здатності за дальністю збільшити випромінювану енергію при обмеженнях на пікову потужність. Збільшення добутку ширини спектру на тривалість радіоімпульсу викликає розтягування зони пасивних перешкод з дальності, що призводить до появи перешкоди з більш рівномірним розподілом інтенсивності у просторі та знижує потенційні можливості з виявлення сигналу. Реальні пасивні перешкоди мають нестационарний розподіл потужності в елементах простору, за таких умов відбитий від цілі сигнал можна виявляти у розривах пасивних перешкод або на ділянках з меншим їх рівнем за умови його оцінювання (ведення карти перешкод) та адаптивного встановлення порогу виявлення за елементами простору. Тому є актуальним проведення досліджень з оцінки якості виявлення відбитих від повітряних цілей сигналів у залежності від рівня нестационарності перешкодового фону.

Метою роботи є розробка методики для оцінки впливу рівня бічних пелюсток функцій кореляції сигналів на показники якості їх виявлення при наявності нестационарного перешкодового фону різної інтенсивності.

Метод. Досліджувалися показники якості виявлення частотно-модульованих сигналів. Задачу оцінки впливу рівня бічних пелюсток функцій кореляції на показники якості виявлення сигналів на фоні нестационарної пасивної перешкоди вирішено шляхом визначення параметрів узагальненого гама-розподілу потужності такої перешкоди в залежності від форми автокореляційної функції сигналу.

Результати. Визначено, що для високого рівня нестационарності початкового перешкодового процесу для усіх моделей сигналів потенційний виграш майже однаковий та має максимальне значення. У разі зниження рівня нестационарності цього процесу виграш зменшується. Традиційний лінійно-частотно модульований сигнал дає у порівнянні з нелінійно-частотно модульованими сигналами дещо гірший результат. Для всіх досліджених законів частотної модуляції виграш більш відчутний за зниження вимог до показників якості виявлення сигналу.

Висновки. Розроблено методику оцінки показників якості виявлення луна-сигналів на перешкодовому фоні з різним ступенем нестационарності. Для покращення енергетичних показників виявлення малорозмірних повітряних об'єктів на фоні нестационарних пасивних перешкод доцільно застосовувати сигнали з нелінійною частотною модуляцією та знижувати значення ймовірності правильного виявлення цілей.

КЛЮЧОВІ СЛОВА: виявлення радіолокаційних сигналів, нелінійна частотна модуляція; рівень бічних пелюсток, нестационарний перешкодовий фон.

ЛІТЕРАТУРА

1. Richards M. A. Principles of modern radar. Basic principles / M. A. Richards, J. Scheer, W. A. Holm. – Sci Tech Pub, 2010. – 924 p.
2. Blackman S. S. Design and analysis of modern tracking systems / S. S. Blackman, R. F. Popoli. – Boston; London: Artech House, 1999. – 1230 p.
3. You H. Radar data processing with applications / H. You, X. Jianjuan, G. Xin. – House of electronics industry publishers, 2016. – 536 p. DOI:10.1002/9781118956878
4. Melvin W. L. Principles of Modern Radar. Vol. II: Advanced Techniques / W. L. Melvin, J. A. Scheer. – Sci Tech Publishing, 2013. – 846 p.
5. Barton D. K. Radar system analysis and modeling / D. K. Barton. – Boston; London: Artech House Publishers, 2004. – 566 p.
6. Van Trees H. L. Detection, estimation, and modulation theory. Part III: Radar-Sonar processing and Gaussian signals in noise / H. L. Van Trees. – John Wiley & Sons, Inc., 2001. – 643 p. DOI: 10.1002/0471221090
7. Levanon N. Radar signals / N. Levanon, E. Mozeson. – New York: John Wiley & Sons, Inc., 2004. – 403 p.
8. Cook C. E. Radar signals: An introduction to theory and application / C. E. Cook, M. Bernfeld. – Boston: Artech House, 1993. – 552 p.
9. McDonough R. N. Detection of signals in noise (2nd. Ed.) / R. N. McDonough, A. D. Whalen. – San Diego : Academic Press, Inc., 1995. – 495 p.
10. Mathematical model of a two-fragment signal with a non-linear frequency modulation in the current period of time / [O. O. Kostyria, A. A. Hryzo, O. M. Dodukh et al.] // Visnyk NTUU KPI Serii – Radiotekhnika Radioaparotobuduvannia. – 2023. – Vol. 92. – P. 60–67. DOI: 10.20535/RADAP.2023.92.60-67.
11. Improvement of mathematical models with time-shift of two- and tri-fragment signals with non-linear frequency modulation / [O. O. Kostyria, A. A. Hryzo, O. M. Dodukh et al.] // Visnyk NTUU KPI Serii – Radiotekhnika Radioaparotobuduvannia. – 2023. – Vol. 93. – P. 22–30. DOI: 10.20535/RADAP.2023.93.22-30.
12. Implementation of the method of minimizing the side lobe level of autocorrelation functions of signals with nonlinear frequency modulation / [O. O. Kostyria, A. A. Hryzo, I. A. Khizhnyak, et al.] // Visnyk NTUU KPI Serii – Radiotekhnika Radioaparotobuduvannia. – 2024. – Vol. 95. – P. 16–22. DOI: 10.20535/RADAP.2024.95.16-22
13. Mathematical model of shifted time of combined signal as part of fragments with linear and quadratic frequency modulation / [O. O. Kostyria, A. A. Hryzo, Yu. S. Solomonenko, et al.] // Visnyk NTUU KPI Serii – Radiotekhnika Radioaparotobuduvannia. – 2024. – Vol. 97. – P. 5–11. DOI:10.20535/RADAP.2024.97.5-11
14. Adithya V. N. Modified radar signal model using NLFM / V. N. Adithya, R. D. Elizabeth, C. Kavitha // International Journal of Recent Technology and Engineering (IJRTE). – 2019. – Vol. 8, Issue 2S3. – P. 513–516. DOI: 10.35940/ijrte.b1091.0782s319
15. Anoosha Ch. Peak sidelobe reduction analysis of NLFM and improved NLFM radar signal with non-uniform PRI / Ch. Anoosha, B. T. Krishna // Aiub Journal of Science and Engineering (AJSE). – 2022. – Vol. 21, Issue 2. – P. 125–131.
16. Non-continuous piecewise nonlinear frequency modulation pulse with variable sub-pulse duration in a MIMO SAR radar system / [Yu. Zhaoa, M. Ritchieb, X. Lua et al.] // Remote Sensing Letters. – 2020. – Vol. 11, Issue 3. – P. 283–292. DOI: 10.1080/2150704X.2019.1711237
17. Optimized multi-parameter NLFM pulse compression waveform for low time-bandwidth radar / [A. C. van-Zyl, E. A. Wiehahn, J. E. Cillers et al.] // International Conference on Radar Systems, 2022. – P. 289–294. DOI: 10.1049/icp.2022.2332.
18. Fan Z. Coded excitation with nonlinear frequency modulation carrier in ultrasound imaging system / Z. Fan, H. Meng // 2020 IEEE Far East NDT New Technology & Application Forum. – Kunming, Yunnan province, China: conference paper: IEEE, 2020. – P. 31–35. DOI: 10.1109/FENDT50467.2020.9337517.
19. Adithya V. N. Performance analysis of NLFM signals with Doppler effect and background noise / V. N. Adithya, R. D. Elizabeth, C. Kavitha // International Journal of Engineering and Advanced Technology (IJEAT). – 2020. – Vol. 9, Issue 3. – P. 737–742. DOI: 10.35940/ijeat.B3835.029320
20. Analytic NLFM waveform design with harmonic decomposition for synthetic aperture radar / [Y. Zhang, Y. Deng, Z. Zhang et al.] // IEEE Geoscience and Remote Sensing Letters, 2022. – Vol. 4, Article no 4513405. DOI: 10.1109/lgrs.2022.3204351
21. A two-step optimization framework for low sidelobe NLFM waveform using fourier series / [Q. Xie, H. Zeng, Z. Mo et al.] // IEEE Geoscience and Remote Sensing Letters, 2022. – Vol. 19, Article no 4020905. DOI: 10.1109/LGRS.2022.3141081
22. Pulse-agile waveform design for nonlinear FM pulses based on spectrum modulation / [R. Zhuang, H. Fan, Y. Sun et al.] // IET International Radar Conference, 2021. – P. 964–969. DOI: 10.1049/icp.2021.0700.
23. Research on ultra-wideband NLFM waveform synthesis and grating lobe suppression / [L. Shuyi, Y. Jia, Y. Liu et al.] // Sensors. – 2022. – No. 24, Article number 9829. DOI: 10.3390/s22249829
24. Swiercz E. Estimation and classification of NLFM signals based on the time-chirp / E. Swiercz, D. Janczak, K. Konopko // Sensors. – 2022. – Vol. 22, Issue 21, Article no 8104. DOI: 10.3390/s22218104
25. A variable chirp rate stepped frequency linear frequency modulation waveform designed to approximate wideband non-linear radar waveforms / [M. Saleh, S.-M. Omar, E. Grivel et al.] // Digital Signal Processing. – 2021. – Vol. 109, Article no 102884. – 19 p. DOI: 10.1016/j.dsp.2020.102884
26. Xu Z. Nonlinear frequency-modulated waveforms modeling and optimization for radar applications / Z. Xu, X. Wang, Y. Wang // Mathematics. – 2022. – Vol. 10. – P. 3939. DOI: 10.3390/math10213939.
27. Analysis of non linear frequency modulation (NLFM) waveforms for pulse compression radar / [M. R. Widyantara, S.-F. Y. Suratman, S. Widodo et al.] // Jurnal Elektronika dan Telekomunikasi. – Vol. 18, Issue 1. – P. 27–34. DOI: 10.14203/jet.v18.27-34
28. PHS: A pulse sequence method based on hyperbolic frequency modulation for speed measurement / [T. Ping, C. Song, Z. Qi et al.] // International Journal of Distributed Sensor Networks. – 2024. – Vol. 2024, Article no 6670576. – 11 p. DOI:10.1155/2024/6670576
29. Magneto-Acousto-Electrical Tomography using Nonlinearly Frequency-Modulated Ultrasound / [Z. Cheng, Z. Sun, J. Wang et al.] // Phys. Med. Biol. – 2024. – Vol. 69(8), PMID: 38422542. DOI:10.1088/1361-6560/ad2ee5

30. Singh A. K. NLFM pulse radar for drone detection using predistortion technique / A. K. Singh, K.-B. Bae, S.-O. Park // *Journal of Electromagnetic Waves and Applications*. – 2021. – Vol. 35. – P. 416–429.
31. Septanto H. A novel rule for designing tri-stages piecewise linear NLFM chirp / H. Septanto, O. Sudjana, D. Suprijanto // *2022 International Conference on Radar, Antenna, Microwave, Electronics, and Telecommunications (ICRAMET)*, Bandung, 6–7 December 2022. – Indonesia: IEEE, 2022. – P. 62–67. DOI: 10.1109/ICRAMET56917.2022.9991201
32. Kostyria O. O. Combined two-fragment radar signals with linear and exponential frequency modulation laws / O. O. Kostyria, A. A. Hryzo, O. M. Dodukh // *Systems of Arms and Military Equipment*. – 2024. – № 4 (76). – P. 58–64. DOI: 10.30748/soivt.2023.76.06
33. Kostyria O. O. Synthesis time-shifted mathematical model of a combined signal with linear and cubic frequency modulation / O. O. Kostyria, A. A. Hryzo, O. M. Dodukh // *Information Processing Systems*. – 2024. – № 1 (176). – P. 73–81. DOI:10.30748/soi.2024.176.09
34. Synthesis of a single-channel device for receiving and primary signal processing under conditions of structural-parametric a priori uncertainty / [M. Kaliuzhnyi, V. Semenets, V. Orlenko et al.] // *2022 IEEE 2nd Ukrainian Microwave Week (UkrMW)*. – Ukraine: IEEE, 2022. – P. 573–578, DOI: 10.1109/UkrMW58013.2022.10037164
35. Feller W. An introduction to probability theory and its applications, Volume 2 / W. Feller. – New York: John Wiley & Sons Inc., 1991. – 683 p.
36. Stacy E. W. A generalization of the gamma distribution / E. W. Stacy // *Annals of Mathematical Statistics*. – 1962. – Vol. 33(3). – P. 1187–1192. DOI: 10.1214/aoms/1177704481
37. Alzaatreh A. Family of generalized gamma distributions: Properties and applications / A. Alzaatreh, C. Lee, & F. Famoye // *Hacettepe Journal of Mathematics and Statistics*. – 2016. – Vol. 45(3). – P. 869–886. DOI:10.15672/HJMS.20156610980
38. Kiche J. On generalized gamma distribution and its application to survival data / J. Kiche, O. Ngesa, & G. Orwa // *International Journal of Statistics and Probability*. – 2019. – Vol. 8(5). – P. 85–102. DOI: 10.5539/ijsp.v8n5p85
39. Sportouche H. Mimic capacity of fisher and generalized gamma distributions for high-resolution SAR image statistical modeling / H. Sportouche, J.-M. Nicolas, F. Tupin // *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*. – 2017. – Vol. 10, Issue 12. – P. 5695–5711
40. Thakur A. Transmit waveform design to avoid target masking in pulse compression radar / A. Thakur and D. S Saini // *2020 International Conference on Communication and Signal Processing (ICCSP)*, Chennai. – India, 2020. – P. 97–101. DOI:10.1109/ICCSP48568.2020.9182221
41. Sidelobe suppression with resolution maintenance for SAR images via sparse representation / [X. Zhu, F. He, F. Ye et al.] // *Sensors (Basel)*. – 2018. – Vol. 18(5), Article no1589. DOI: 10.3390/s18051589; PMID: 29772743; PMCID: PMC5982680.
42. Haliloglu O. Successive target cancelation in pulse compression radars / O. Haliloglu, A. O. Yilmaz // *2007 IEEE Radar Conference*, Waltham. – MA, USA: IEEE, 2007. – P. 885–890. DOI: 10.1109/RADAR.2007.374336
43. A simulation of study of the impact of surface clutter on spaceborne precipitation radar sensor / [R. Nepal, Z. Li, Y. Zhang et al.] // *In Proceedings of the 36th Conference on Radar Meteorology*, Breckenridge. – CO, USA: IEEE, 2013. – P. 16–20.
44. Maki M. Construction of three-dimensional weather radar data from volcanic eruption clouds / M. Maki, T. Kobori // *MethodsX*. – 2021. – Vol. 8, Article no 101535. DOI: 10.1016/j.jvolgeores.2021.107178

МАТЕМАТИЧНЕ ТА КОМП'ЮТЕРНЕ МОДЕЛЮВАННЯ

MATHEMATICAL AND COMPUTER MODELING

UDC 004.94

IMPLICIT CURVES AND SURFACES MODELING WITH PSEUDO-GAUSSIAN INTERPOLATION

Ausheva N. M. – Dr. Sc., Professor, Head of the Department of Digital Technologies in Energy, National Technical University of Ukraine “Igor Sikorsky Kyiv Polytechnic Institute”, Kyiv, Ukraine.

Sydorenko Iu. V. – PhD, Associate Professor of the Department of Digital Technologies in Energy, National Technical University of Ukraine “Igor Sikorsky Kyiv Polytechnic Institute”, Kyiv, Ukraine.

Kaleniuk O. S. – PhD, Senior lecturer of the Department of Digital Technologies in Energy, National Technical University of Ukraine “Igor Sikorsky Kyiv Polytechnic Institute”, Kyiv, Ukraine.

Kardashov O. V. – Post-graduate student of the Department of Digital Technologies in Energy, National Technical University of Ukraine “Igor Sikorsky Kyiv Polytechnic Institute”, Kyiv, Ukraine.

Horodetskyi M. V. – Post-graduate student of the Department of Digital Technologies in Energy, National Technical University of Ukraine “Igor Sikorsky Kyiv Polytechnic Institute”, Kyiv, Ukraine.

ABSTRACT

Context. With the contemporary development of topological optimization, and parametric and AI-guided design, the problem of implicit surface representation became prominent in additive manufacturing. Although more and more software packages use implicit modeling for design, there is no common standard way of writing, storing, or passing a set of implicit surfaces or curves over the network. The object of the study is one of the possible ways of such representation, specifically: modeling implicit curves and surfaces using pseudo-Gaussian interpolation.

Objective. The goal of the work is the development of a modeling method that improved the accuracy of the implicit object representation without significant increase in memory used or processing time spent.

Method. One of the conventional ways to model an implicit surface would be to represent its signed distance function (SDF) with its values defined on a regular grid. Then a continuous SDF could be obtained from the grid values by means of interpolation. What we propose instead is to store not SDF values but the coefficients of a pseudo-Gaussian interpolating function in the grid, which would enable picking the exact interpolation points before the SDF model is written. In this way we achieve better accuracy in the regions we're interested the most in with no additional memory overhead.

Results. The developed method was implemented in software for curves in 2D and validated against several primitive implicit curves of different nature: circles, squares, rectangles with different parameters of the model. The method has shown improved accuracy in general, but there were several classes of corner cases found for which it deserves further development.

Conclusions. Pseudo-Gaussian interpolation defined as a sum of radial basis functions on a regular grid with points of interpolation defined in the proximity of the grid points generally allows to model an implicit surface more accurately than a voxel model interpolation does. The memory intake or computational toll isn't much different in these two approaches. However, the interpolating points selection strategy and the choice of the best modeling parameters for each particular modeling problem remain an open question.

KEYWORDS: surface representation, curve representation, implicit representation, pseudo-Gaussian function, regular grid, implicit surface modeling, implicit surface data format.

ABBREVIATIONS

3MF is a data format called 3D Manufacturing Format;

DICOM is a data format called Digital Imaging and Communications in Medicine;

HDF5 is a data format called Hierarchical Data Format;

HRBF a a Hermite radial basis function;

NRRD is a data format called Nearly Raw Raster Data;

RBF is a radial basis function;

SDF is a signed distance function;

SOP is a Service-Object Pair.

NOMENCLATURE

k is a distance between grid points;

M is a first dimension of a 3D or 2D grid;

n is an order of the pseudo-Gaussian sum continuity;

N is a second dimension of a 3D or 2D grid;

P is a third dimension of a 3D grid;

$P(x)$ is an algebraic polynomial that models a Gaussian function;

$pg(x)$ is a pseudo-Gaussian function of a single variable;

$pgr_{ij}(x, y)$ is a radial basis function based on a pseudo-Gaussian function;

r is a range of pseudo-Gaussian basis functions;

$SDF(x, y)$ is a signed-distance function representing an implicit curve to model;

$Sp_g(x, y)$ is a pseudo-Gaussian interpolating function;

(x_a, y_b) is an interpolating point in 2D;

(x_a, y_b, z_c) is an interpolating point in 3D;

(x_j, y_i) is a point of a 2D grid;

(x_j, y_b, z_l) is a point of a 3D grid.

INTRODUCTION

With the contemporary development of topological optimization, and parametric and AI-guided design, the problem of implicit surface representation became prominent in additive manufacturing. Although more and more software packages use implicit modeling for design, there is no common standard way of writing, storing, or passing a set of implicit surfaces or curves over the network. The straightforward way to represent an implicit surface would be to write down all the operations necessary for its signed distance function computation, but this would be too tedious to formalize and implement, besides the producer of the model design might not even want to disclose its computation method as it might be by itself a trade secret.

One of the usual ways to represent a signed distance function (SDF) of an implicit surface would be to write down its values defined on a regular grid. This is sometimes called a 3D image, a 3D bitmap, or a voxel model. Then a continuous SDF could be obtained from the grid values by means of interpolation. What we propose instead is to store not SDF values but the coefficients of a pseudo-Gaussian interpolating function, which allows us to pick the exact interpolation points before the SDF model is written. In this way we achieve better accuracy in the regions we're the most interested in (see Fig. 1.) with no additional memory overhead.

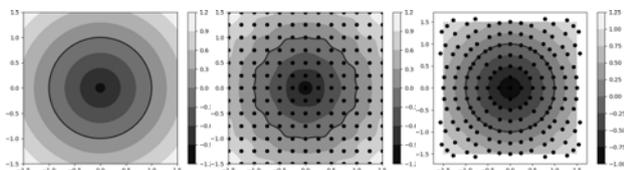


Figure 1 – An SDF of a circle, its model with values set on a regular grid, its model with interpolation points adjusted

The object of study is the modeling of implicit curves and surfaces with data on regular grids.

The subject of study is the pseudo-Gaussian interpolation in the context of implicit curves and surfaces modeling.

The purpose of the work is to increase the accuracy of implicit surface modeling while not increasing the memory intake of the model.

1 PROBLEM STATEMENT

Let $SDF(x, y, z)$ be a function that represents an implicit surface $SDF(x, y, z) = 0$. The value of the function, when not zero, is the signed distance to the abovementioned implicit surface.

Suppose we have a regular grid of points (x_j, y_b, z_l) , where $j = 1..M, i = 1..N, l = 1..P$. Let distance between the closest grid points be k , and the coordinate-wise minimal point of the grid: (x_0, y_0, z_0) .

The problem of the implicit surface modeling $SDF(x, y, z) = 0$ with pseudo-Gaussian interpolation would be to define a pseudo-Gaussian interpolation function $Sp_g(x, y, z)$ that is close to the $SDF(x, y, z)$ within some range of tolerance on the 3D range $[x_0, x_0 + (M-1)k] \times [y_0, y_0 + (N-1)k] \times [z_0, z_0 + (P-1)k]$.

2 REVIEW OF THE LITERATURE

Since normally storing or transmitting the whole computation path of an implicit surface's SDF is impractical, the computation may even require going beyond the algebra of real numbers [1], we have to resort to some sort of an SDF model instead. The common way to represent an SDF with an easily serializable homogeneous data structure fit for transmission would be to use a voxel grid, also known as 3D-image, where each voxel is assigned a value or a set of values that help the receiving side model the initial surface. In this case, the problem of representation concerns the voxel access optimization [2] or the specific data assigned to each voxel, which can not only be an SDF value in a point but a gradient of the SDF in that very point as well [3].

In recent times, another common approach for implicit surface modeling is using a neural network trained to recreate the surface [4], [5]. For this approach, the problem of representation is then reduced to the problem of representation of the model's coefficients. It is not a solved problem, different networks allow users to balance the model precision and the model size differently.

We propose to use an approach that, in a way, combines the previous two. We store the coefficients of the model, but the model is not a neural network but an interpolating function, consisting of weighted radial functions, defined on the regular grid, similar to a voxel grid.

Using radial functions is a common practice in another adjacent problem – the problem of implicit surface reconstruction from a point cloud. For instance, a meta-analysis of implicit surface reconstruction via RBF interpolation methods performed by Mo Jiahui, Shou Huahao, and Chen Wei in 2022 [6] lists 125 published sources. A comparative analysis of RBF approximations has also been performed by Majdisova and Skala [7].

Often, along with the values of modeled functions to approximate, RBFs also take a function's gradient into account. In this case, we usually refer to them as Hermite radial basis functions (HRBF) [8], [9].

In this work, we use a very specific sort of RBF approximation – Gaussian interpolation [10], [11]. However, to enable localized computation, and thus enable practical application of the global interpolating function, we must resort to a substitution. In this work, we'll use pseudo-Gaussian functions instead of Gaussian functions.

3 MATERIALS AND METHODS

Let's define a pseudo-Gaussian function as a symmetric function $pg(x)$ of a real argument such as:

$$\begin{aligned} pg(0) &= 1; \\ pg(r) &= 0; \\ pg(r)^{(i)} &= 0, i = 1..n; \\ pg(x) &= 0, x \leq -r; \\ pg(x) &= 0, x \geq r. \end{aligned} \quad (1)$$

In this way, the function is non-zero on $(-r, r)$, zero elsewhere, and it has its first n derivatives in $-r$ and r , specified as 0 too.

The function is symmetric and, as such, mimics the Gaussian function near 0 (see Fig. 2).

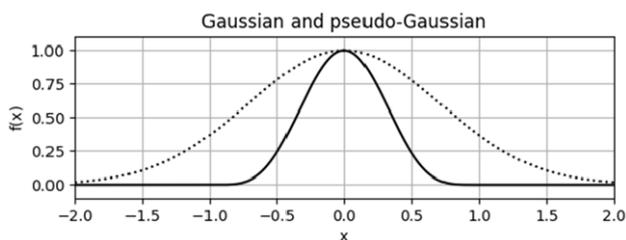


Figure 2 – A Gaussian function (dotted) and a pseudo-Gaussian with $n = 4, r = 1$ (solid)

The amount of non-zero derivatives on $(-r, r)$ that also become 0 in r and $-r$ condition the continuity class C^n of the pseudo-Gaussian sum and therefore the smoothness degree of the implicit surface (or a contour in 2D) that we're aiming to model.

To specify a pseudo-Gaussian $pg(x)$ for any given r and n , we can use a polynomial representation. A polynomial that represents $pg(x)$ will be a symmetrical polynomial, so it will only consist of even degrees of x , and its degree then should be at least $2(n+2)$. To obtain all the coefficients of a minimal degree polynomial that satisfy the conditions from above, we should comprise a system of equations:

$$\begin{cases} P(0) = 1, \\ P(r) = 0, \\ P'(r) = 0, \\ P''(r) = 0, \\ \dots \\ P^{(n)}(r) = 0. \end{cases} \quad (2)$$

Since r and n are known constants and we only need to determine the coefficients of the polynomial, the system is linear. We can use any solver to get the coefficients for any degree n and any r .

Interestingly, we might not even have to solve the system at all. The coefficients of the polynomial regardless of r seem to follow the binomial coefficients. For instance, here are the generalized formulas for the $P_n(x)$ for $n=1..4$:

$$pg_1(x) = 1 - \frac{2x^2}{r^2} + \frac{x^4}{r^4}, \quad (3)$$

$$pg_2(x) = 1 - \frac{3x^2}{r^2} + \frac{3x^4}{r^4} - \frac{x^6}{r^6}, \quad (4)$$

$$pg_3(x) = 1 - \frac{4x^2}{r^2} + \frac{6x^4}{r^4} - \frac{4x^6}{r^6} + \frac{x^8}{r^8}, \quad (5)$$

$$pg_4(x) = 1 - \frac{5x^2}{r^2} + \frac{10x^4}{r^4} - \frac{10x^6}{r^6} + \frac{5x^8}{r^8} - \frac{x^{10}}{r^{10}}. \quad (6)$$

The core difference between the Gaussian and the pseudo-Gaussian function is that the latter is localized, meaning that all non-zero values of $pg(x)$ lie within the $(-r, r)$ interval. Since we're planning to use this function as a basis radial function determined on a regular grid, this means that for any point in space, we can compute the sum of all the functions in the whole grid by only computing the ones that are defined in the proximity of the point. Every RBF defined further than r from the point of computation will be 0 by definition and, therefore will not contribute to the sum.

This allows us to use a global interpolation function – the pseudo-Gaussian interpolation function – such as it would have been local.

A pseudo-Gaussian interpolating function is a weighted sum of pseudo-Gaussian radial functions determined in all the points of a finite regular grid.

Let's now focus on 2D space for brevity. If we have a grid of size $M \times N$ that starts at a point (x_0, y_0) , and has a constant distance k between adjacent points (x_{j+1}, y_j) and (x_j, y_i) as well as between (x_j, y_{i+1}) and (x_j, y_i) for all $i = 0..N-2, j = 0..M-2$, then the points of this grid would be defined as:

$$(x_j, y_i) = (x_0 + jk, y_0 + ik), i = 0..N-1, j = 0..M-1. \quad (7)$$

Let's say we have defined pseudo-Gaussian-based radial basis functions $pg_{r_{ij}}(x, y)$ for all $i = 0..n-1, j = 0..m-1$:

$$pg_{r_{ij}}(x, y) = pg_n(|(x, y) - (x_j, y_i)|). \quad (8)$$

And their respective coefficients a_{ij} . The pseudo-Gaussian interpolating function will then look like this:

$$Spg(x, y) = \sum_{i=0}^{n-1} \sum_{j=0}^{m-1} a_{ij} pgr_{ij}(x, y). \quad (9)$$

Although here we present a formula as a full sum of all the weighted functions, remember, we don't have to do the full computation for all the basis functions every time we want to compute the $Spg(x, y)$ at a point. Given that the pseudo-Gaussians $pg(x)$ we choose have an effective range of non-zero values r , meaning that $pg(x) = 0$ when $x \leq -r$ and $x \geq r$, to compute the interpolating function at a point (x, y) we only need to account for the radial basis functions that are defined in the nodes (x_j, y_i) where $||[x, y] - [x_j, y_i]|| < r$. This means that in order to compute an $Spg(x, y)$ at any point, only a relatively small and predefined number of basis functions should be computed, which reduces the computation complexity of such a computation from $O(N^2)$ to $O(1)$.

The coefficients a_{ij} are exactly the data that characterize the $Spg(x, y)$ function. To make the pseudo-Gaussian sum interpolate an arbitrary function we want to model: $SDF(x, y)$, we should select exactly $M \times N$ points (x_a, y_b) where $a = 1..M, b = 1..N$ such as they are all defined within the range of the corresponding radial functions meaning that $|(x_a, y_b) - (x_j, y_i)| < r$. Then for the defined grid and defined values $SDF(x_a, y_b)$, we can define a system of equations:

$$Spg(x_a, y_b) = SDF(x_a, y_b). \quad (10)$$

In regard to the coefficients a_{ij} the system is linear. Moreover, for $N > r/d$ and $M > r/d$, which is expected given the presumed application of this interpolation function, the system will appear sparse. Methods of solution for these types of systems are well known. By solving the system with a chosen method we get the array of a_{ij} coefficients each defined for a grid point (x_j, y_i) and for its radial basis function respectively.

The pseudo-Gaussian interpolation function generalizes easily to 3D. We still use the same radial basis functions, it's only now there are 3 arguments in the function, and the grid is, of course, also 3-dimensional.

If we have a grid of size $M \times N \times P$ that starts at a point (x_0, y_0, z_0) , and has a constant distance k between adjacent points (x_{j+1}, y_b, z_l) and (x_j, y_b, z_l) , as well as between (x_j, y_{i+1}, z_l) and (x_j, y_b, z_l) , and now also (x_j, y_b, z_{l+1}) and (x_j, y_b, z_l) , for all $i = 0..N-2, j = 0..M-2, l = 0..P-2$ then the points of this grid would be:

$$(x_j, y_i, z_l) = (x_0 + jk, y_0 + ik, z_0 + lk), \quad (11)$$

$$i = 0..N-1, j = 0..M-1, l = 0..P-1.$$

The radial basis functions:

$$pgr_{ijl}(x, y, z) = pg_n(|(x, y, z) - (x_j, y_i, z_l)|). \quad (12)$$

And the pseudo-Gaussian interpolating function will then be:

$$Spg(x, y, z) = \sum_{i=0}^{n-1} \sum_{j=0}^{m-1} \sum_{l=0}^{p-1} a_{ijl} pgr_{ijl}(x, y, z). \quad (13)$$

The interpolation points are then defined as (x_a, y_b, z_c) , where $a = 1..M, b = 1..N, c = 1..P$.

And the system to get the coefficients a_{ijl} is:

$$Spg(x_a, y_b) = SDF(x_a, y_b). \quad (14)$$

Note that in both 2D and 3D, the interpolation points may or may not be equal to the grid points. In this work, we argue that selecting interpolation points different from the grid points is a valuable feature we get by storing interpolation function coefficients instead of SDF values.

In both 2D and 3D, the result of interpolation relies heavily on the choice of the interpolation points (x_a, y_b) or (x_a, y_b, z_c) respectively. Of course, other factors impact the interpolation result too: the length of the non-zero range r of the pseudo-Gaussian function $pg(x)$, the distance between the grid points k , and the order of continuity of the basis functions sum n . Still, while all these parameters shape the resulting function differently, and as such deserve studying on their own, they mostly help balance the computational complexity and the accuracy of the model. The interpolation points choice, however, is crucial for successful SDF modeling.

As we mentioned before, we can forfeit the choice to the default and set interpolation points equal to the points of the grid: $(x_a, y_b) = (x_j, y_i)$.

Let's take a 2D SDF of a rotated square (Fig. 3), and show how the result of its interpolation looks like when the interpolation points coincide with the points of the grid (Fig. 4).

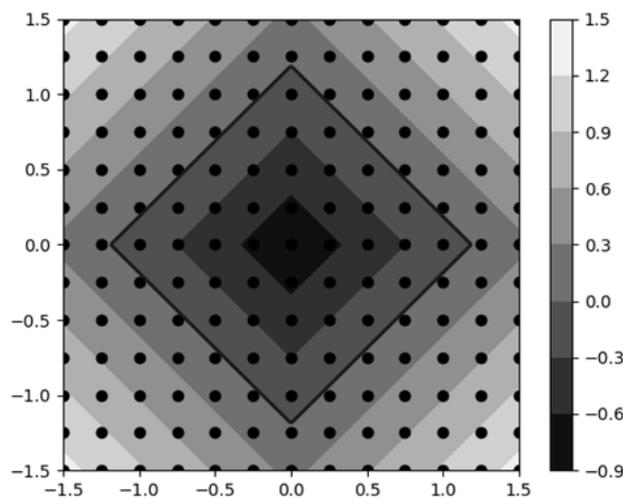


Figure 3 – An SDF of a square rotated by 45 degrees, and a regular 13x13 grid

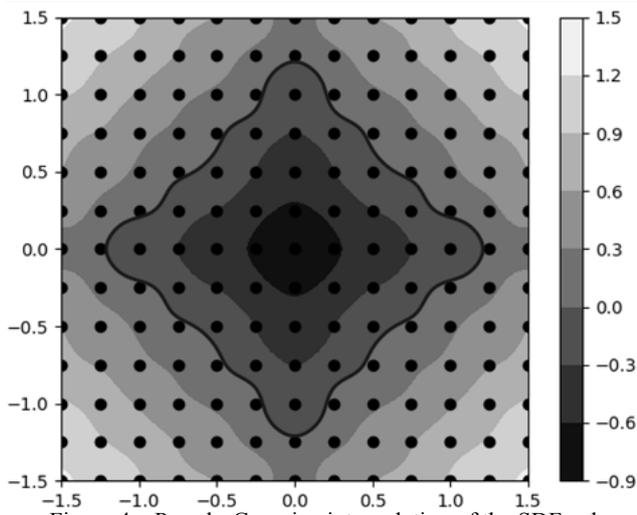


Figure 4 – Pseudo-Gaussian interpolation of the SDF values in grid points

This interpolation models the SDF and we already can store the coefficients of the interpolating function in a 2D image instead of SDF values. The resulting image will take exactly as much memory as the 2D image of SDF values, but we wouldn't have to compute the coefficients afterward.

But if we define each point of interpolation (x_a, y_b) as the point of the isoline $SDF(x, y) = z$, nearest to the grid point (x_j, y_i) , where $z = z_{step}q$, q is integer, and the isoline step $z_{step} \leq k$, then with this choice of interpolation points, we make our model much more precise around the select isolines, and, what's even more important, around the isoline $SDF(x, y) = 0$ that represents the entity we want to model (see Fig. 5).

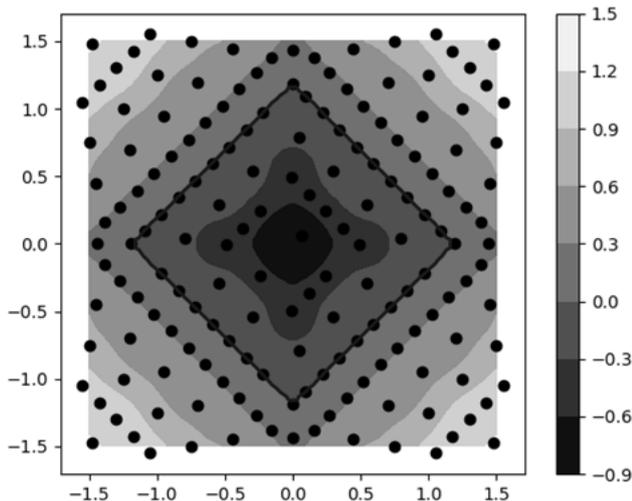


Figure 5 – Model of the rotated square SDF with interpolation points defined on $SDF(x, y) = 0.25q$

The points (x_j, y_i) where the radial functions are defined remain intact for every possible choice of the interpolating points (x_a, y_b) . Of course, the choice affects the model, it affects its coefficients a_{ij} , but not the way the interpolating function is computed from these coefficients, so this choice of points brings no impact on the

computational complexity of the whole function computation either.

The proposed method of selecting the interpolation points has not been extensively tested in real-world cases. It rather illustrates the idea of point selection itself than serves as a practical guide. The idea itself is interesting since the very possibility of point selection extends our possibilities for more precise modeling of implicit contours in 2D and surfaces in 3D, and doesn't add any memory toll to the resulting data structure.

4 DISCUSSION

With our approach, we can define the interpolating point in any point in between grid points but there are limitations to this too. Any point (x_a, y_b) should be no further from the grid point (x_j, y_i) than r . But what's more important, the total number of these points still can't exceed nm for the regular grid of size $m \times n$. In the context of storing and transmitting SDF models, this means that we can make our model of SDF more accurate in some regions only by forfeiting accuracy in others. This works well with implicit surface (and contour in 2D) representation where we want to focus on the isoline $SDF(x) = 0$, but not necessary in the more general context.

In a way, this reflects how the lossy compression works. We accentuate the important information, and by letting the less important details go, use less data to store essentially the same object as with lossless compression. But what's important and what's not depends solely on the context of the application, not on the information, or the model, itself.

For instance, the consistency of the SDF is important for contouring – turning the SDF into a set of polylines in 2D and triangle meshes in 3D. Common contouring methods like marching cubes, surface nets, or double contouring – all rely on the linearity of the SDF. Marching cubes often use linear interpolation to determine the vertices' positions, and dual contouring also requires Hermite data, so it requires the gradient of the SDF to be consistent with the distance function too.

Another example of a practical problem where the accuracy of the SDF model is equally important as on the isoline $SDF(x) = 0$ as everywhere else is the offset modeling. Signed distance functions, as their name hints are perfect for offset modeling since all you have to do to build an offset of an SDF is to subtract the offset distance from the SDF. Offsets are routinely used in 3D printing both in 3D for solving positioning and supporting problems, and in 2D, to generate toolpaths from the contours of the sliced model. Although our approach allows us to put modeling focus into multiple isolines, we don't necessarily know beforehand which isolines will be used for offsetting so we can't exploit this information to make our model more accurate for the offsetting specifically.

The interpolation points selection is a valuable option, but it's not omnipotent. We can make our model more accurate for some applications, but there will always be counterexamples when this rearrangement of data points backfires.

The other problem is that the control over interpolating points doesn't necessarily grant control over the shape of the modeled body. The exact shape of the surface or contour we achieve doesn't necessarily coincide with the initial shape let alone our expectations. In Fig. 6 you can see an SDF and the pseudo-Gaussian interpolation of this SDF with interpolation points defined on $SDF(x, y) = 0.25q$.

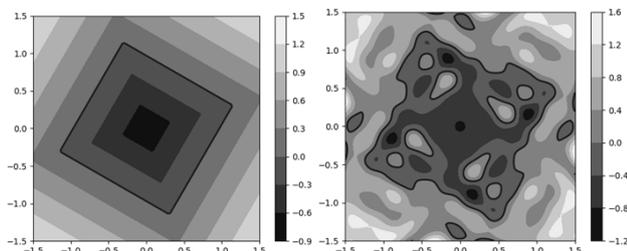


Figure 6 – An SDF of a rotated square and its pseudo-Gaussian interpolation gone wrong

Please note, that 0.25 is the same isoline step as before (see Fig. 5). We also use the same basis functions and the same grid. In Fig. 5 we have shown a similar square rotated slightly differently, and the same strategy for interpolation points selection worked there well.

The result of the modeling depends on four factors:

1. The non-zero range of pseudo-Gaussian basis functions r .
2. The order of the pseudo-Gaussian sum continuity n .
3. The distance between grid points k .
4. And the choice of interpolation points (y_j, y_i) .

How exactly these factors affect the result of the modeling has not been yet extensively studied. We have run a series of experiments modeling implicit curves in 2D to determine how r , n , and k affect the model's shape but the results have been inconclusive so far.

Sometimes, when a model of an implicit curve loses its original shape, meaning that there are extra contours in between interpolation points as in Fig. 6, increasing the non-zero range of the pseudo-Gaussian function may help establish its original contour. But for some SDFs, this is not the case. We haven't found the specific pattern here. Also, note that enlarging the r raises the computation cost of the interpolation function since for each point we compute it at, we have to take more basis functions into account. In other words, we can't raise r too much, as this will hinder the applicability of the whole approach.

Also, since at every point the resulting model depends on the contribution of all the radial functions defined within radius r from the point of computation, it is highly recommended to define the interpolating grid in a way that it bounds the zero polyline with at least r/d wide margin (see Fig. 7).

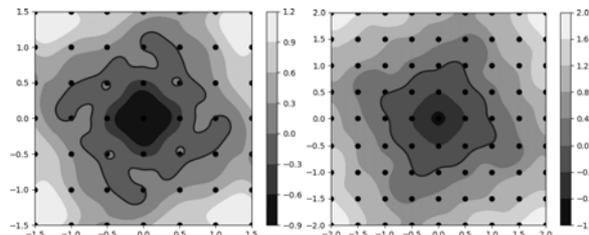


Figure 7 – An extra margin of only one cell helps preserve the shape

Surprisingly, raising the order of pseudo-Gaussian sum continuity n may also help keep the shape of the modeled curve closer to the original although within limits. E. g. if the shape is lost with $n = 2$, raising n to 3 or 4 might help. But if the shape is still lost, adding more members to the pseudo-Gaussian function will most likely elevate the computational cost with no additional benefit. We haven't tested this extensively so far, so this requires further investigation. It might even be worth considering not only experimenting with n , but taking completely different kinds of radial basis functions, or perhaps not even radial, into test.

Managing the distance between points k or, which is the same, regulating the grid density is the common way of trading computational speed and memory footprint for accuracy with traditional voxel-based modeling. Normally, but not necessarily, it works the same with our interpolation-based approach too. The problem is, with linear interpolation, the value between two interpolation points lies in between their respective values. With pseudo-Gaussian interpolation, this is not a given. Besides, the whole point of storing interpolation coefficients instead of SDF values was to increase model precision in the important area without adding extra grid points there. Raising the grid density too much diminishes this advantage.

So while selecting different r , n , and k may or may not help us ensure that the shape of the model is consistent with its original, the only consistent way to do so is by choosing the interpolation points. Moreover, this is the only choice that doesn't affect the computational speed or the memory footprint of the resulting model. Also, this is the only approach that has so far shown itself robust in the experiments.

For the model in Fig. 6 specifically, we have conducted the following computational experiment. We didn't change the method of interpolation points selection itself, we still chose the closest point on isoline $SDF(x, y) = z_{step}q$. Using numerical optimization, however, we found the isoline step value where the difference between the original SDF and its interpolated model is minimal. It was shown that for this particular model, the best z_{step} , computed up to the 6th decimal sign after the decimal point, equals 0.081384. See Fig. 8.

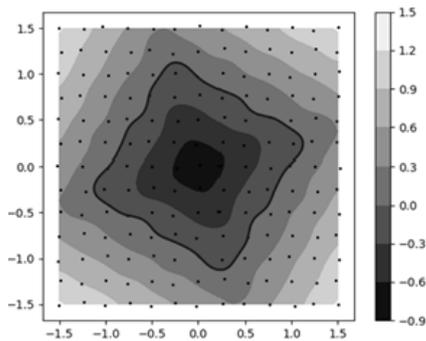


Figure 8 – A model of the square SDF with interpolating points defined on $SDF(x, y) = 0.081384q$

The experiment shows that the interpolation points selection is crucial for every modeling problem in particular. The fact that the optimal z_{step} for the model in the experiment is so low hints that it could be that for this case we need another method of points selection. Something that puts more data into the main isoline $SDF(x, y) = 0$ and not necessarily $SDF(x, y) = z_{stepq}$ where $q \neq 0$.

CONCLUSIONS

Pseudo-Gaussian interpolation defined as a set of radial basis functions on a regular grid with points of interpolation defined in the proximity of the grid points allows us to model an implicit surface more accurately than a voxel model interpolation. At the same time, the memory or computational toll isn't much different in these two approaches. Let's summarize.

Due to the pseudo-Gaussian radial basis function localization, the computational complexity of the whole interpolation function doesn't depend on the grid size and thus can be evaluated as $O(1)$.

Given that the interpolation function is a weighted sum of the radial basis functions defined in the known grid points, and to define such a function we only need to compute, store, and, if necessary transmit, its weight coefficients – 1 per grid point – the memory footprint of this model is exactly the same as a footprint of a voxel grid with SDF values.

At the same time, we retain an option to choose the points of interpolation before the weight coefficients are composed. This allows us to model select isolines more accurately than any generic interpolation of a voxel grid can.

Additionally, since the data structure of the pseudo-Gaussian interpolation function is compatible with a voxel grid, we can already store this type of data in any existing data format that allows voxel grids (also known as 3D images or grey-value floating point images). This includes 3MF, HDF5, NRRD, and even, with the introduction of a user-defined SOP extension, DICOM.

At the same time, the three mechanisms to balance the performance and accuracy of the modeling: selecting the non-zero range of pseudo-Gaussian basis functions r , picking the order of the pseudo-Gaussian sum continuity n , and choosing the distance between grid points k – all deserve additional studying.

But the most important is the problem of the interpolating points selection. The method we proposed in this paper brings only illustrative not practical value. The key to turning the pseudo-Gaussian grid interpolation into a robust technology for implicit surface storage and transmission lies in discovering the best interpolating points selection strategy.

ACKNOWLEDGEMENTS

The study was performed without financial support.

REFERENCES

1. Andrianov I. V., Ausheva N. M., Olevska Yu. B., Olevskiy V. I. Surfaces Modelling Using Isotropic Fractional-Rational Curves, *Journal of Applied Mathematics*. Hindawi, 2019, Vol. 1, pp. 1–13. DOI: 10.1155/2019/5072676
2. Li H., Yang X., Zhai H., Liu Y., Bao H., Zhang G. Vox-Surf: Voxel-Based Implicit Surface Representation, *IEEE Transactions on Visualization and Computer Graphics*, 2024, Vol. 30 (3), pp. 1743–1755. DOI: 10.1109/TVCG.2022.3225844
3. Sommer C., Sang L., Schubert D., Cremers D. Gradient-SDF: A Semi-Implicit Surface Representation for 3D Reconstruction, *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2022, pp. 6280–6289.
4. Michalkiewicz M., Pontes J. K., Jack D., Baktashmotlagh D., Eriksson A. Implicit Surface Representations As Layers in Neural Networks, *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, 2019, pp. 4743–4752.
5. Xu Q., Wang W., Ceylan D., Mech R., Neumann U. DISN: Deep Implicit Surface Network for High-quality Single-view 3D Reconstruction, *Advances in Neural Information Processing Systems*, 2019, Vol. 32. DOI: 10.48550/arXiv.1905.10711
6. Jiahui M., Huahao Sh., Wei Ch. Implicit Surface Reconstruction via RBF Interpolation: A Review, *Recent Patents on Engineering*, 2022, Vol. 16 (5), pp. 49–66. DOI: 10.2174/1872212115666210707110903
7. Majdisova Z., Skala V. Radial basis function approximations: comparison and applications, *Appl. Math. Modelling*, 2017, Vol. 51, pp. 728–743. DOI: 10.1016/j.apm.2017.07.033
8. Macêdo I., Gois J. P., Velho L. Hermite interpolation of implicit surfaces with radial basis functions, *Proceedings of the 2009 XXII Brazilian Symposium on Computer Graphics and Image Processing (SIBGRAPI)*, IEEE, 2009, pp. 1–8. DOI: 10.1109/SIBGRAPI.2009.11
9. Pan R., Meng X., Whangbo T. Hermite variational implicit surface reconstruction, *Science in China Series F: Information Sciences*, 2009, Vol. 52, pp. 308–315. DOI: 10.1007/s11432-009-0032-x
10. Sydorenko Yu. V., Horodetskyi M. V. Modification of the algorithm for selecting a variable parameter of the Gaussian interpolation function, *Control Systems and Computers*, 2020, Vol. 6 (290), pp. 21–28. DOI: 10.15407/csc.2020.06.021
11. Sydorenko Yu. V., Onysko A. I., Shaldenko O. V., Horodetskyi M. V. Interpolation of different types of spiral-like curves by gaus-interpolation methods, *Control Systems and Computers*, 2022, Vol. 3 (299), pp. 1–10. DOI: 10.15407/csc.2022.03.003

Received 27.11.2024.
Accepted 28.01.2025.

МОДЕЛЮВАННЯ ІМПЛІЦИТНИХ КРИВИХ І ПОВЕРХОНЬ ПСЕВДОГАУСОВОЮ ІНТЕРПОЛЯЦІЄЮ

Аушева Н. М. – д-р техн. наук, професор, завідувачка кафедри цифрових технологій в енергетиці Національного технічного університету України «Київський політехнічний інститут імені Ігоря Сікорського», Київ, Україна.

Сидоренко Ю. В. – канд. техн. наук, доцент кафедри цифрових технологій в енергетиці Національного технічного університету України «Київський політехнічний інститут імені Ігоря Сікорського», Київ, Україна.

Калениук О. С. – канд. техн. наук, старший викладач кафедри цифрових технологій в енергетиці Національного технічного університету України «Київський політехнічний інститут імені Ігоря Сікорського», Київ, Україна.

Кардашов О. В. – аспірант кафедри цифрових технологій в енергетиці Національного технічного університету України «Київський політехнічний інститут імені Ігоря Сікорського», Київ, Україна.

Городецький М. В. – аспірант кафедри цифрових технологій в енергетиці Національного технічного університету України «Київський політехнічний інститут імені Ігоря Сікорського», Київ, Україна.

АНОТАЦІЯ

Актуальність. Із сучасним розвитком методів топологічної оптимізації, параметричного проектування і проектування на основі штучного інтелекту, проблема запису імпліцитних поверхонь у задачах 3D-друку стала надважливою. Хоча все більше програмних продуктів використовують імпліцитні моделі у проектуванні, єдиного стандарту для запису, а значить і для збереження та передачі таких моделей засобами комп'ютерних мереж, не існує. Об'єктом цього дослідження є один з можливих способів такого запису, а саме – моделювання імпліцитних кривих і поверхонь із використанням псевдогаусової інтерполяції.

Мета роботи. Ціль роботи полягає у створенні способу моделювання імпліцитних кривих і поверхонь, із покращеною точністю без значних витрат пам'яті чи часу обчислення.

Метод. Одним із загальноприйнятих способів запису функції знакової відстані (ФЗВ) є запис її значень на регулярній сітці. Такий спосіб ще називається 3D-зображення, 3D-бітмап, або воксельна модель. Неперервна ФЗВ може бути отримана із записаних значень за допомогою інтерполяції. Натомість пропонується записувати не значення ФЗВ, а значення коефіцієнтів псевдогаусової інтерполяційної функції, що дозволяє обирати точки інтерполяції до запису коефіцієнтів моделі. Таким чином можна досягти більш точного моделювання у найважливіших регіонах (див. рисунок 1) без використання додаткової пам'яті.

Результати. Запропонований спосіб був імплементований у вигляді комп'ютерної програми для моделювання плоских імпліцитних кривих і провалідований на декількох примітивних моделях різного походження: колах, квадратах, прямокутниках, – із різними параметрами моделі. Вцілому, порівняно із інтерпольованими значеннями ФЗВ у точках решітки, метод показує кращу точність, але разом із тим має декілька граничних станів, у яких він потребує подальшого вивчення.

Висновки. Псевдогаусова інтерполяція, визначена як сума радіальних базисних функцій на регулярній сітці із точками інтерполяції визначеними у ненульовому околі точок сітки в загальному випадку дозволяє моделювати імпліцитні криві і поверхні точніше ніж інтерполяція воксельної моделі. Разом з тим, оптимальна стратегія визначення точок інтерполяції і інших параметрів моделі для прикладного застосування лишається відкритою проблемою.

КЛЮЧОВІ СЛОВА: представлення поверхонь, представлення кривих, імпліцитне представлення, псевдогаусова функція, регулярна сітка, моделювання імпліцитних поверхонь, формат даних для імпліцитних поверхонь.

ЛІТЕРАТУРА

1. Surfaces Modelling Using Isotropic Fractional-Rational Curves / [I. V. Andrianov, N. M. Ausheva, Yu. B. Olevska, V. I. Olevskiy] // Journal of Applied Mathematics. – Hindawi, 2019. – Vol. 1. – P. 1–13. DOI: 10.1155/2019/5072676
2. Vox-Surf: Voxel-Based Implicit Surface Representation / [H. Li, X. Yang, H. Zhai et al.] // IEEE Transactions on Visualization and Computer Graphics. – 2024. – Vol. 30 (3), P. 1743–1755. DOI: 10.1109/TVCG.2022.3225844
3. Gradient-SDF: A Semi-Implicit Surface Representation for 3D Reconstruction / [C. Sommer, L. Sang, D. Schubert, D. Cremers] // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). – 2022. – P. 6280–6289.
4. Implicit Surface Representations As Layers in Neural Networks / [M. Michalkiewicz, J. K. Pontes, D. Jack et al.] // Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV). – 2019. – P. 4743–4752.
5. DISN: Deep Implicit Surface Network for High-quality Single-view 3D Reconstruction / [Q. Xu, W. Wang, D. Ceylan et al.] // Advances in Neural Information Processing Systems. – 2019. – Vol. 32. DOI: 10.48550/arXiv.1905.10711
6. Jiahui M. Implicit Surface Reconstruction via RBF Interpolation: A Review / M. Jiahui, Sh. Huahao, Ch. Wei // Recent Patents on Engineering. – 2022. – Vol. 16 (5). – P. 49–66. DOI: 10.2174/1872212115666210707110903
7. Majdisova Z. Radial basis function approximations: comparison and applications / Z. Majdisova, V. Skala // Appl. Math. Modelling. – 2017. – Vol. 51. – P. 728–743. DOI: 10.1016/j.apm.2017.07.033
8. Macêdo I. Hermite interpolation of implicit surfaces with radial basis functions / I. Macêdo, J. P. Gois, L. Velho // Proceedings of the 2009 XXII Brazilian Symposium on Computer Graphics and Image Processing (SIBGRAPI), IEEE. – 2009. – P. 1–8. DOI: 10.1109/SIBGRAPI.2009.11
9. Pan R. Hermite variational implicit surface reconstruction / R. Pan, X. Meng, T. Whangbo // Science in China Series F: Information Sciences. – 2009. – Vol. 52. – P. 308–315. DOI: 10.1007/s11432-009-0032-x
10. Sydorenko Yu. V. Modification of the algorithm for selecting a variable parameter of the Gaussian interpolation function / Yu. V. Sydorenko, M. V. Horodetskiy // Control Systems and Computers. – 2020. – Vol. 6 (290). – P. 21–28. DOI: 10.15407/csc.2020.06.021
11. Sydorenko Yu. V. Interpolation of different types of spiral-like curves by gaus-interpolation methods / [Yu. V. Sydorenko, A. I. Onysko, O. V. Shaldenko, M. V. Horodetskiy] // Control Systems and Computers. – 2022. – Vol. 3 (299). – P. 1–10. DOI: 10.15407/csc.2022.03.003

THE STATES' FINAL PROBABILITIES ANALYTICAL DESCRIPTION IN AN INCOMPLETELY ACCESSIBLE QUEUING SYSTEM WITH REFUSALS AND WITH INPUT FLOW OF REQUIREMENTS' GROUPS

Gorodnov V. P. – Dr. Sc., Professor, Professor of the National Academy of National Guard of Ukraine, Kharkov, Ukraine.

Druzhynin V. S. – PhD, Associate Professor of the Department of Fuel Supply of Odessa Military Academy, Odessa, Ukraine.

ABSTRACT

Context. The basis for the creation and management of real queuing systems (QS) is the ability to predict their effectiveness. For the general case of such systems with refusals, with limited approachability of service devices and with a random composition of group requirements in the input flow, the prediction of their performance remains an unsolved problem.

Objective. The research has the aim to find an analytical representation for final probabilities in the above-mentioned case of Markov QS, which allows us to predict the efficiency of its operation depending on the values of the parameters in its structure and control.

Method. For the above-mentioned types of QS, the state probabilities can be described by a system of Kolmogorov's differential equations, which for the stationary case is transformed into a homogeneous system of linearly dependent algebraic equations. For real QS in communication systems, the number of equations can be estimated by the degree set and amount to several thousand, which gives rise to the problem of their recording and numerical solution for a specific set of operating conditions parameters values. The predictive value of such a solution does not exceed the probability of guessing the numerical values of the QS operating conditions parameters set and for parameters with a continuous value, for example, for random time intervals between requests, is zero.

The method used is based on the analytical transition to the description of QS states groups with the same number of occupied devices. At the same time, the desire to obtain the final probabilities of states in a form close to the Erlang formulas remains. The influence of the above-mentioned QS properties can be localized in individual recurrent functions that multiplicatively distort Erlang formulas.

Results. For the above-mentioned types of QS, analytical calculation formulas for estimating the QS states final probabilities have been found for the first time, which makes it possible to predict the values of all known indicators of system efficiency. In this case, the deformation functions of the states groups' probability distribution in QS have a recurrent form, which is convenient both for finding their analytical expressions and for performing numerical calculations.

When the parameters of the QS operating conditions degenerate, the resulting description automatically turns into a description of one of known QS with failures, up to the Erlang QS.

Conclusions. The analytical calculation expressions found for the final probabilities of the above-mentioned QS turned out to be applicable to all types of Markov QS with failures, which was confirmed by the results of a numerical experiment. As a result, it became possible to practically apply the obtained analytical description of the considered QS for operational assessments of developed and existing QS effectiveness in the possible range of their operating conditions.

KEYWORDS: Markov queueing systems, requirements' groups.

ABBREVIATIONS

AAMF is an anti-aircraft missile fire zone;

MAA is an enemy means of air attack;

$M_L/M/n$ is a designation of a QS with a Poisson input flow of requirements groups with random composition, with maximum number (L) of requirements in a group (M_L), with an exponential distribution of random service time for each requirement (M), with number (n) of identical service channels (devices);

$M/M/n$ is a Kendall-Basharin classification for QS with failures;

No. i, j is a cell address in Table 1: i -row number, j -column number;

QS is a designation for queuing system;

SAM is a designation for an anti-aircraft missile system;

NOMENCLATURE

b_r is a probability designation of occurrence in the input flow of QS a request consisting of exactly r requirements;

C_n^m is a number of combinations from n to m ;

d_k^q is a probability of transition to the k -th level of the model graph of queuing system by jumping through q tiers of the graph;

$f_1()$ is a notation for the input density function of the demand flow;

$f_2()$ is a notation for the service duration distribution density function;

F_k is a probability deformation function, which deforms the Erlang probability P_k ;

I is a designation for the intensity of the requirements input flow in QS;

i_1, \dots, i_n is a designation for the numbers of specific service devices;

$k_1 k_2$ are the numbers of channels that become occupied when moving from previous states to the state under consideration;

L is a designation for the maximum total requirements number in one request;

$M_{b,d}$ is a designation for the mathematical expectation of devices number engaged in servicing;

M_m is a designation for the mathematical expectation of the number of requirements in one request at the QS input;

n is a designation for the number of devices/channels in QS;

N_{missed} is a designation for the aircraft number mathematical expectation that have penetrated the air defense system with impunity;

$N_{total,En}$ is an enemy aircraft total number in a blow;

P_k is a notation for the QS state probability in which are occupied exactly k devices;

p_{ijk} is a designation for such a QS state probability in which are engaged in servicing devices with numbers i, j, k ;

$P_{service}$ is a designation for the probability of requirements service in QS (QS performance indicator);

$Q_{i_1 \dots i_m}$ is a probability of the requirement falling within the service area of channels with i_1, \dots, i_m numbers;

r is a designation for requirements number in one request;

s is a number of air defense missile systems not engaged in firing at enemy aircraft;

$S_{i_1 \dots i_k}$ is one of possible QS states, in which channels with numbers i_1, \dots, i_m are busy servicing;

T is a mathematical expectation of requirement's duration service time in service device;

v_i is a designation for the requirement starting service by one of the devices probability, given that i available devices are occupied already;

v_i^1 is a designation for the intensity of requirement starting service by one of the devices, given that i available devices are occupied already;

z_j is a total area of j -fold overlap of service channel accessibility zones;

z_Σ is a total service area of requirements flow in QS;

λ is a designation for the input requests flow parameter in QS;

λ_r is a designation for the requests partial flow parameter which consists of exactly r requirements in each request at the input of the QS;

μ is a designation for channel performance in QS;

ρ is a load factor of a service device in QS of $M/M/n$ type;

α_k^q is an intensity of the transition QS into a group of states with k occupied channels by a jump over q group states QS;

γ_i is a channel number occupied in the previous and current states;

β_i is a busy channel service number in a queuing system;

π_j is a probability of a requirement falling into the total accessibility zone of exactly j service channels;

ξ_m is a designation for the channels' groups maximum number m out of n channels' total number;

ρ_q is a designation for the QS load factor by the partial λ_q incoming flow of requests.

INTRODUCTION

To ensure life, every person has to periodically satisfy their needs for food, clothing, communication services, transportation services, banking services, medical and other services. Needs for such services often arise at unexpected (random) moments in time. For a group of people within individual regions, such needs accumulate, generating continuous flows of typical demands for specific services.

For the noted conditions, the properties of such random-time flows of events were investigated by A. Ya. Khinchin [1], and he proved the statement about the asymptotic emergence of a set of mathematical features in these flows, which received the name "simplest" flow of events.

Due to the repetitive nature of service requirements, systems for processing and satisfying such requirements are created in each specific area of activity to meet them. The creation of systems requires preliminary calculations of many parameters values for each system, which determined the need for the emergence of calculation models, called queuing system (QS) models. For conditions with the simplest input flow of demands, Markov models of QS with failures are known, where if all devices are busy at the moment the next demand arrives, this demand is rejected and leaves the QS without being serviced.

There are many different conditions for the receipt and servicing of requirements for which the necessary QS models are developed. Thus, at the moment of admission, a patient at a clinic may not find a free doctor of the required specialization, a driver at a gas station may not find the required type of fuel, an enemy aircraft may fly through the air defense group's fire zone with impunity. In all such cases, at the moment of the requirement receipt, there may be free service devices in the system, but these devices are not available for service.

In some cases, the system input may receive requirements not one by one, but in groups with a composition that is not known in advance.

Thus, during an epidemic or during military operations, patients may be admitted to medical

institutions in groups. Enemy aircraft, as a rule, carry out missions as part of tactical groups that may enter the fire zone of an air defense group.

As a result, shock loads arise in mass service systems and the efficiency of the systems decreases.

When developing such systems, it becomes necessary to have a calculation scheme (a model) of its parameters in which both properties – the flow of groups of requirements and the partial availability of service channels – must be taken into account simultaneously.

For any real QS with the noted features, the development of necessary model is possible in the form of a system of Kolmogorov differential equations and the corresponding algebraic equations for the stationary operating mode of the QS. The solution of equations system is possible only by numerical methods. In this case, the structure of equations must correspond to the value of operating conditions parameters.

It is possible to guess the future values of some random parameters of working conditions with a probability strictly equal to zero, which reduces the predictive value of calculations and makes it relevant to develop an analytical description of the required model.

The first analytical description of a single-link switching non-fully accessible QS model was obtained in the theory of teletraffic [2] (the third Erlang formula). For the case of group requests entering the system, an analytical model was developed in [3]. For a multi-channel non-fully accessible $M/M/n$ QS, an analytical model was developed in [4] and can be useful for developing an analytical model that simultaneously takes into account two noted features of the QS operating conditions.

The research object is a stationary service process in a queuing system $M_L/M/n$ with refusals, with the entry into the system requirements groups with a previously unknown composition and with incomplete accessibility of service devices for the incoming flow of requirements.

The research subject is the probability distribution law of states groups in incompletely accessible QS of the $M_L/M/n$ type in a stationary mode.

The research purpose is to obtain an analytical description of the final probabilities of states groups in incompletely accessible QS of $M_L/M/n$ type, with a simultaneous assessment of its correctness.

1 PROBLEM STATEMENT

A flow of requests (requirement groups) with intensity I and density $f_1(t) = Ie^{-It}$ enters non-fully accessible QS of $M_L/M/n$ type.

Each service channel in non-fully accessible QS can be a part of one or more channel groups. Each group of r ($r=1, \dots, L$) requirements represents a service request and must be serviced using r service devices.

One of the accessible device groups is selected to service the next request in the input flow. In the selected

channel group, any free channel is assigned to service each requirement in the request with equal probability.

In case of insufficient number of free channels some of requirements from this request are denied service and leave the system. The service duration of each requirement is random and has an exponential distribution $f_2(t) = \mu e^{-\mu t}$.

The considered distribution densities allow us to assert the possibility of describing the desired model in the class of Markov processes with discrete states and continuous time.

For the sake of brevity, we will further use the well-known statements [1, pp. 14, 40, 41] on the properties of a stationary non-ordinary flow of requirements.

A stationary flow of time moments without after effects, in which groups of events appear, is called a non ordinary or universal stationary flow and has the properties of the simplest flow. A non ordinary flow includes requests of r requirements ($r=1, 2, \dots, L$) in a request.

Such a flow can be defined by specifying the probability distribution law (b_r) of the occurrence of exactly r requirements in any group (in any request) of the input flow. The intensity of requirements groups flow turns out to be greater than the flow parameter ($I, \lambda < I$), which contains piecemeal flows with parameters λ_r :

$$\lambda_r = \lambda b_r, \quad r = \overline{1, L}; \quad M_{in} = \sum_{r=1}^L r \cdot b_r; \quad (1)$$

$$\lambda = \sum_{r=1}^L \lambda_r; \quad I = \sum_{r=1}^L r \cdot \lambda_r.$$

In a non ordinary flow of requirements at the QS input, time intervals between requests are random and satisfy the conditions of A. Ya. Khinchin limit theorem [1], have an exponential distribution with the parameter λ :

$$f(t) = \lambda e^{-\lambda t}, \quad t > 0. \quad (2)$$

The following well-known [4] logic of operation of the same QS, but with the simplest input flow of requirements, can be used as the basis for constructing the $M_L/M/n$ model with incomplete accessibility of service devices, an example of which, for the variant of a city polyclinic operation, is given in [4].

To demonstrate the physical processes, involved in constructing a model for servicing the simplest flow of requirements by a non-fully accessible system, [4] examines a simplified example of the operation of three SAM systems group, located on the ground (Fig. 1) when firing at the simplest flow of enemy aircraft.

Each SAM system (No. 1, No. 2, No. 3) has a firing zone, projected onto the earth's surface, in the form of a

circle (Fig. 1 Q_1, Q_2, Q_3). In the general fire zone of SAM systems there are areas with mutual overlapping fire zones of neighboring SAM systems, which in the service system form a group of devices.

Each SAM system (service device) can be a member of several service groups. To indicate the probability that the next requirement (enemy aircraft) will be accessible to a specific group of m devices, a designation indicating the numbers of these devices is used $Q_{i_1...i_m}$.

In the example (Fig. 1) such probabilities for SAM systems fire zones are indicated by symbols $Q_1, Q_2, Q_3, Q_{12}, Q_{23}, Q_{13}, Q_{123}$.

When an aircraft enters such an area, the aircraft can be fired upon (servicing the requirement) by any of the adjacent SAM systems.

The group's maximum number ξ_j from j service devices out of the total number of n service devices coincides with the number of combinations C_n^j from n by j (3):

$$\xi_j = C_n^j = \frac{n!}{j! \cdot (n-j)!}, \quad 0 \leq j \leq n. \quad (3)$$

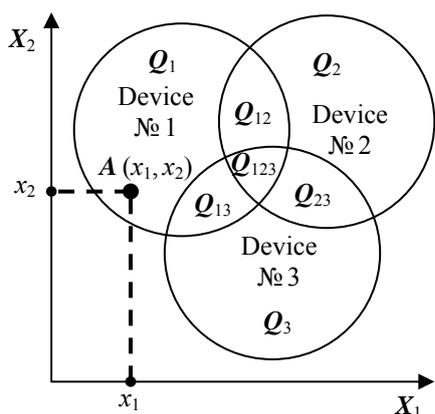


Figure 1 – Visualization of the principles for choosing an affordable device to serve the next requirement [4]

Within a constant number j , the observed probability remains approximately the same for different groups of devices:

$$Q_{i_1...i_j} = l_j. \quad (4)$$

The probability l_j estimation method for the example of a city polyclinic operation is given in [4]. For a simplified example of SAM systems grouping, the relative value π_j of the total area z_j of the region with j -fold overlap of fire zones can be used:

$$\pi_j = \frac{z_j}{z_\Sigma}; \quad 0 \leq j \leq n. \quad (5)$$

On the other hand, the probability that the next requirement (enemy aircraft) will be accessible for servicing by any of groups of j devices (SAM systems), is equal to the probabilities sum:

$$\pi_j = \sum_{i=1}^{\xi_j} l_j = \xi_j \cdot l_j = C_n^j \cdot l_j, \quad 0 \leq j \leq n. \quad (6)$$

From expressions (3) and (4) we find the probability estimate l_j :

$$l_j = \frac{z_j}{C_n^j \cdot z_\Sigma} = \frac{\pi_j}{C_n^j}, \quad 0 \leq j \leq n. \quad (7)$$

In [4] an example of a $M/M/n$ QS graph (Fig. 2) with incomplete availability of service channels is considered in detail and a basic equality [4, formulas (23), (29)] is found for finding the final probabilities of group states of a non-fully accessible QS with refusals (8):

$$\begin{aligned} k \cdot P_k &= P_{k-1} \cdot C_{n-k+1}^1 \cdot \rho \cdot v_{k-1}, \quad k=1, 2, \dots, n; \\ \rho &= \frac{I}{\mu}; \text{ then } k\mu P_k = P_{k-1} \cdot C_{n-k+1}^1 \cdot \alpha_k, \quad k=\overline{1, n}; \\ v_k^1 &= I \cdot \sum_{j=1}^n \left(l_j \cdot \sum_{i=0}^{j-1} \frac{1}{C_{i+1}^1} C_{k-1}^{j-1-i} C_{n-k}^i \right), \quad 0 \leq k \leq n. \end{aligned} \quad (8)$$

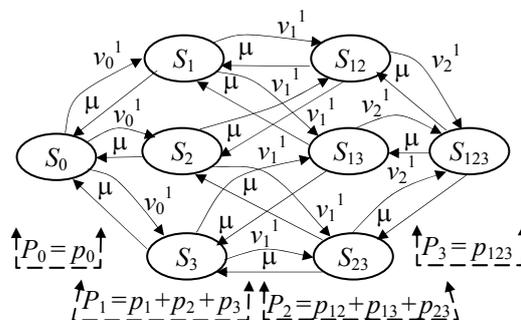


Figure 2 – Graph of $M/M/3$ QS model with refusals and with incomplete accessibility of service devices [4]

In [4], the correctness of the obtained expressions for the final probabilities was also proven by their automatic transition to the well-known Erlang formulas [5] for the $M/M/n$ system with refusals.

In [3, formula (17)], a basic equality was found for searching for the final probabilities P_k of group states in QS with a non ordinary requirements groups' input flow:

$$\rho_i = \frac{1}{\mu} \sum_{j=1+i}^L \lambda_j, \quad 0 \leq i < L;$$

$$kP_k = \rho_0 P_{k-1} + \rho_1 P_{k-2} + \dots + \rho_{k-1} P_0; \quad (9)$$

$$\mu k P_k = \sum_{j=1}^k P_{k-j} \cdot \sum_{q=j}^L \lambda_q.$$

In [3] the correctness of the obtained expressions for the final probabilities is also shown by their automatic transition to the well-known Erlang formulas [5] for the $M/M/n$ system with refusals.

The problem is to find analytical expressions for final probabilities P_k , ($k = 1, 2, \dots, n$) of states groups' in the service process, in which in the $M_L/M/n$ system, with incomplete accessibility of service devices there are exactly k requirements. The marked probabilities of states group's in QS, in which exactly k channels are occupied with servicing, allow us to find the servicing probability $P_{service}$ of requirements.

Thus, the problem of predicting the efficiency of real queuing systems with partially accessible devices with refusals arises in the presence of a requirements groups' flow with random composition at the input of the system.

2 REVIEW OF THE LITERATURE

The first model of a queuing system was developed by A.K. Erlang [5] in 1909 to describe the operation of a telephone station. The process of each requirement servicing from the next subscriber consisted of connecting his communication channel to the communication channel with another subscriber. After the end of the communication session, the channels were freed and it became possible to use them for service requests from other subscribers. If a requirement was received from a subscriber at a time when all communication channels were busy, such a requirement from the subscriber was denied service. The moments of requirements receipt time and end of communication sessions were not known in advance and were considered random in the model.

Processes of servicing requirements with an unknown start and end time also occur in other systems and areas of human activity. These include logistic systems [6], production systems [7], telecommunication networks [2], systems for management in medicine [8], systems for traffic management [9], and systems for the defense of objects from air blows [10] as well as socio-economic systems [11]. In each of these areas, the Erlang model could be used either directly or with modifications that were necessitated by the peculiarities of the processes in a particular area. Thus, the closest to the description of processes in the management of medical institutions and systems for protecting objects from air strikes were models [4] and [3], which separately took into account the properties of incomplete accessibility of service devices and the group composition of requirements in service requests, respectively. Each of these properties necessitated a transformation of the Erlang model, which

made it possible to assess the directions in changes in the efficiency of the corresponding systems. However, both of the noted properties can simultaneously occur in service processes, which necessitated further modification of the Erlang model [5].

3 MATERIALS AND METHODS

For the sake of certainty, we will consider the construction of the desired service model using the example of repelling attack of enemy aircraft of different composition and purpose (flow of requests at the input of the service system) by group of n SAM systems deployed on the ground, which, in this case under consideration, is n -channel not fully accessible queuing system. In the general zone of anti-aircraft missile fire of the SAM systems group there are areas with j -th layer of fire zone (Fig.1). At each such section, the enemy aircraft can be fired upon by any of the j SAM systems (serviced by any of the j devices). Each section can correspond to its own set of specific SAM systems (service devices).

The enemy MAA strike passes through the SAM systems grouping fire zone. The enemy MAA combat formations contain MAA groups of different purposes according to $r = 1, \dots, L$ MAA in the group (a non ordinary flow of requirements at the input of the queuing system).

If a group, consisting of r MAAs (r -group) enters a section of the AAMF zone with j -th layering, in which s SAM systems are free $s \leq j$, then at $r > s$ exactly s the MAAs will begin to be fired upon (will be serviced), and $r - s$ MAAs will pass through the AAMF zone without impact (are denied in service). There is reason to assume that the probability of the r -group MAA getting into any of the AAMF regions with the j -th layering is the same and equal to l_j . Then the probability π_j of getting into one of the zones with the j -th layering is found (3)–(7), taking into account all the accessible zones:

$$\pi_j = \xi_j \cdot l_j, \quad 0 \leq j \leq n. \quad (10)$$

In turn, from (10) and (3) it follows:

$$l_j = \frac{\pi_j}{C_n^j}, \quad 0 \leq j \leq n. \quad (11)$$

For the noted conditions, the combat model of the SAM systems group can be represented by a model of incompletely accessible queuing system of a non ordinary flow of requests.

Statement for a mathematical problem. A non ordinary Poisson flow of requirements, grouped into requests, with the parameter λ enters QS $M_L/M/n$ with non-fully accessible channels and with the vector of probabilities $\{\pi_j\}$ for suitability of requirements to service devices.

The flow of requirements is a superposition of independent partial flows, each of which is characterized

by a constant number r of requirements in a request ($r=1, \dots, L$) and a parameter λ_r of the exponential distribution of time intervals between neighboring requests:

$$\lambda_r = \lambda \cdot b_r, \quad r=1, \dots, L. \quad (12)$$

Each request (group of requirements) can be assigned to any of the service zones, which has its own individual set of channels with specific numbers j_1, \dots, j_k capable of servicing requirements in that zone. Each channel spends on average T minutes on a service cycle and has a productivity of $\mu = T^{-1}$.

Any of the accessible free channels is selected for servicing without any preferences, that is, with a probability inversely proportional to the number of free accessible service channels. Some of the requirements, for which there are free service channels in the zone under consideration, are started to be serviced and remain in the system until the end of the service cycle. Requirements, for which there are no free channels in the zone under consideration, leave the system unserved (are denied service). One channel can service only one requirement at a time.

At any given moment in time, the system can be in one of the possible states $S_{i_1 \dots i_k}$ with specific numbers i_1, \dots, i_k and the number k of channels occupied by servicing. The set of all possible states with k occupied channels forms a group state S_k of the system, is called a tier of the model graph and contains C_n^k possible internal states.

The problem is to determine analytical expressions for the final probabilities P_k , $k=0, \dots, n$ of the service system's group states S_k .

Solution. A fragment of the model's graph, under consideration, with the number of service channels $n=4$ is shown in Figure 3, where only representative arcs of typical states $S_{i_1 \dots i_k}$ for each tier of the model graph are shown. The intensities of transitions along these arcs are marked in the breaks of the arcs. The return from the state $S_{i_1 \dots i_k}$ of the k -th tier to the $(k-1)$ -th tier occurs in the same states from which it was possible to get to the states $S_{i_1 \dots i_k}$ of the k -th tier.

Further, for the convenience of analysis, the hypothesis is adopted that $L \geq n$, which does not reduce the generality of the analysis, since it is always possible to include in the flow of requirements the missing probabilities b_r , setting them equal to zero.

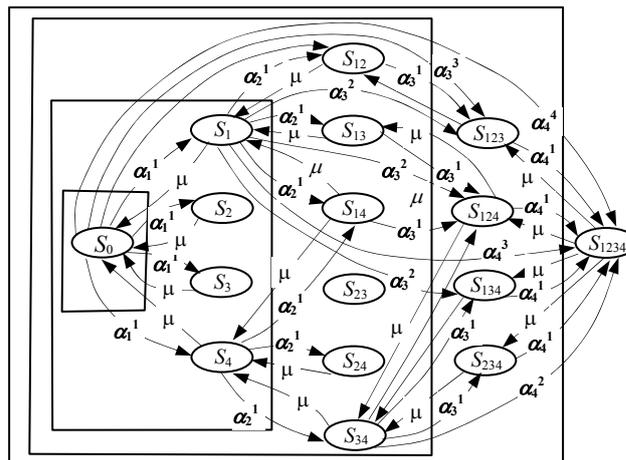


Figure 3 – Fragment of the service model graph of a non ordinary flow of requirements by a non-fully accessible QS with refusals

For the intensities of transitions to states of tiers with a large value of k , the notation α_k^q is used, in which the lower index denotes the absolute number of the tier to which the transition arc leads, and the upper index indicates the number of tiers traversed by the arc, including the tier of the final state of the model. For example, the designation α_3^2 shows the intensity with which transitions to a specific state of the 3rd tier follow from a specific state of the 1st tier (Fig. 3, an example of a transition from state S_1 to state S_{134}).

In order to understand the order of transition intensities formation, we will use the designation $Q_{i_1 \dots i_k}$ of a requirements group probability of falling into the service area of specific group of channels and compose expressions for transition intensities. Let us temporarily introduce notations, indicating the numbers of occupied channels in the initial state of the model graph and in the final state.

$$\begin{aligned} \alpha_{0 \rightarrow 1}^1 &= \lambda_1(Q_1 + \frac{1}{2} \cdot (Q_{12} + Q_{13} + Q_{14}) + \\ &+ \frac{1}{3} \cdot (Q_{123} + Q_{124} + Q_{134}) + \frac{1}{4} \cdot Q_{1234}) + Q_1 \cdot \sum_{r=2}^L \lambda_r; \\ \alpha_{1 \rightarrow 12}^1 &= \lambda_1(Q_2 + Q_{12} + \frac{1}{2} \cdot (Q_{23} + Q_{24}) + \\ &+ \frac{1}{2} \cdot (Q_{123} + Q_{124}) + \frac{1}{3} \cdot Q_{234} + \frac{1}{3} \cdot Q_{1234}) + \\ &+ (Q_2 + Q_{12}) \cdot \sum_{r=2}^L \lambda_r; \end{aligned} \quad (13)$$

$$\begin{aligned} \alpha_{12 \rightarrow 123}^1 &= \lambda_1(Q_3 + Q_{23} + Q_{13} + \frac{1}{2}Q_{34} + \\ &+ Q_{123} + \frac{1}{2} \cdot (Q_{134} + Q_{234}) + \frac{1}{2} \cdot Q_{1234}) + \\ &+ (Q_3 + Q_{13} + Q_{23} + Q_{123}) \cdot \sum_{r=2}^L \lambda_r; \\ \alpha_{123 \rightarrow 1234}^1 &= \lambda_1(Q_4 + Q_{14} + Q_{24} + Q_{34} + \\ &+ Q_{124} + Q_{134} + Q_{234} + Q_{1234}) + (Q_4 + Q_{14} + \\ &Q_{24} + Q_{34} + Q_{124} + Q_{134} + Q_{234} + Q_{1234}) \cdot \sum_{r=2}^L \lambda_r. \end{aligned}$$

Taking into account (4), that is, the hypothesis about the equal probability of requests (groups of requirements) getting into any service zone with the same number j of accessible service channels, we find that the intensities of transitions between the states of two different tiers are the same for all arcs of one direction, connecting the states of these tiers and are equal:

$$\begin{aligned} \alpha_{0 \rightarrow 1}^1 &= \alpha_{0 \rightarrow 2}^1 = \dots = \alpha_1^1 = \\ &= \lambda_1 \cdot (l_1 + \frac{3}{2}l_2 + l_3 + \frac{1}{4}l_4) + l_1 \cdot \sum_{r=2}^L \lambda_r; \\ \alpha_{1 \rightarrow 12}^1 &= \alpha_{1 \rightarrow 13}^1 = \dots = \alpha_2^1 = \\ &= \lambda_1 \cdot (l_1 + 2l_2 + \frac{4}{3}l_3 + \frac{1}{3}l_4) + (l_1 + l_2) \cdot \sum_{r=2}^L \lambda_r; \\ \alpha_{12 \rightarrow 123}^1 &= \alpha_{12 \rightarrow 124}^1 = \dots = \alpha_3^1 = \lambda_1 \cdot (l_1 + \\ &+ \frac{2}{5}l_2 + 2l_3 + \frac{1}{2}l_4) + (l_1 + 2l_2 + l_3) \cdot \sum_{r=2}^L \lambda_r; \\ \alpha_{123 \rightarrow 1234}^1 &= \alpha_{124 \rightarrow 1234}^1 = \dots = \alpha_4^1 = \lambda_1 \cdot (l_1 + \\ &+ 3l_2 + 3l_3 + l_4) + (l_1 + 3l_2 + 3l_3 + l_4) \cdot \sum_{r=2}^L \lambda_r. \end{aligned} \quad (14)$$

The obtained expressions (14) in the general case take the following form:

$$\begin{aligned} \alpha_k^1 &= \lambda_1 \cdot \sum_{j=1}^n l_j \cdot \sum_{i=0}^{j-1} \frac{1}{C_{i+1}^1} \cdot C_{k-1}^{j-1-i} \cdot C_{n-k}^i + \\ &+ (\sum_{j=0}^{k-1} l_{j+1} \cdot C_{k-1}^j) \cdot \sum_{r=2}^L \lambda_r. \end{aligned} \quad (15)$$

Reasoning similarly for the graph arcs that transfer the service process through one state, we find:

$$\begin{aligned} \alpha_{0 \rightarrow 12}^2 &= \lambda_2(Q_{12} + \frac{1}{3} \cdot Q_{123} + \frac{1}{3}Q_{124} + \\ &+ \frac{1}{6}Q_{1234}) + Q_{12} \cdot \sum_{r=3}^L \lambda_r; \\ \alpha_{1 \rightarrow 123}^2 &= \lambda_2(Q_{23} + Q_{123} + \frac{1}{3}Q_{234} + \\ &+ \frac{1}{3}Q_{1234}) + (Q_{12} + Q_{123}) \cdot \sum_{r=3}^L \lambda_r; \\ \alpha_{12 \rightarrow 1234}^2 &= \lambda_2(Q_{34} + Q_{134} + Q_{234} + \\ &+ Q_{1234}) + (Q_{34} + Q_{134} + Q_{234} + Q_{1234}) \cdot \sum_{r=3}^L \lambda_r. \end{aligned} \quad (16)$$

Taking into account equality (4), expressions (16) will take the following form:

$$\begin{aligned} \alpha_{0 \rightarrow 12}^2 &= \alpha_{0 \rightarrow 13}^2 = \dots = \alpha_2^2 = \\ &= \lambda_2 \cdot (l_2 + \frac{2}{3}l_3 + \frac{1}{6}l_4) + l_2 \cdot \sum_{r=3}^L \lambda_r; \\ \alpha_{1 \rightarrow 123}^2 &= \alpha_{1 \rightarrow 124}^2 = \dots = \alpha_3^2 = \\ &= \lambda_2 \cdot (l_2 + \frac{4}{3}l_3 + \frac{1}{3}l_4) + (l_2 + l_3) \cdot \sum_{r=3}^L \lambda_r; \\ \alpha_{12 \rightarrow 1234}^2 &= \alpha_{13 \rightarrow 1234}^2 = \dots = \alpha_4^2 = \\ &= \lambda_2 \cdot (l_2 + 2l_3 + l_4) + (l_2 + 2l_3 + l_4) \cdot \sum_{r=3}^L \lambda_r. \end{aligned} \quad (17)$$

However, in this case, the general expression for the transition intensity α_k^2 is not obvious. To derive such an expression, we introduce a model of one state at the k -th level of the graph (Fig. 4).

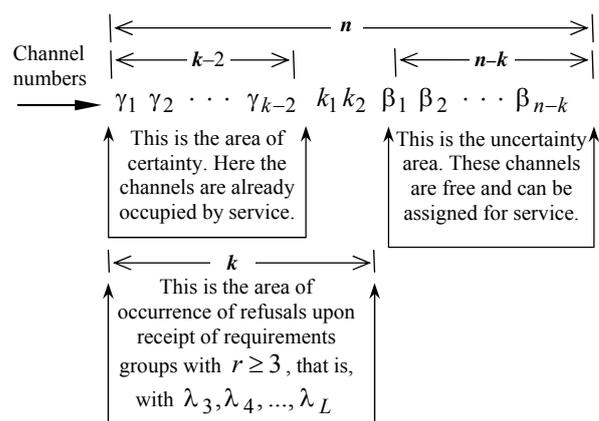


Figure 4 – Model of a single state with k busy channels on the k -th tier of none-full-accessible QS's graph

In this state $\gamma_1 \gamma_2 \dots \gamma_{k-2} k_1 k_2$, the service channels are busy and the $\beta_1 \beta_2 \dots \beta_{n-k}$ channels are free. The

edge with the desired transition intensity α_k^2 leads from the state with $\gamma_1 \gamma_2 \dots \gamma_{k-2}$ busy channels (the symbol γ_i can correspond to any number of the real service channel, provided that it occurs once in the set $\gamma_1 \gamma_2 \dots \gamma_{k-2}$) to the marked state $\gamma_1 \gamma_2 \dots \gamma_{k-2} k_1 k_2$.

Thus, in the analyzed state there are three types of channels:

γ_i – channels occupied in the previous and current states;

$k_1 k_2$ – channels occupied during the transition from previous states to the state under consideration;

β_i – unoccupied channels.

The transition to the state $\gamma_1 \gamma_2 \dots \gamma_{k-2} k_1 k_2$ can be caused by a partial flow λ_2 of requests consisting of 2 requirements, as well as by flows λ_r of requests with a large number of requirements ($r \geq 3$) in the request.

In the last case, these requests should not fall into areas with channel accessibility $\beta_1 \beta_2 \dots \beta_{n-k}$ (Fig. 4), since this will lead to the inclusion of β -channels and bypassing state $\gamma_1 \gamma_2 \dots \gamma_{k-2} k_1 k_2$ in consideration.

From the composition of each mentioned request λ_r of flows ($r \geq 3$) only 2 requirements will get into the service channels $k_1 k_2$, the remaining $(r-2)$ requirements of each such request will be lost. Let us denote the intensity of transitions caused by this part of the flows with “refusals” by the symbol $\alpha_{k^{**}}^2$.

At the same time, the flow requests λ_2 can also fall into the areas of β -channels. In this case, the probability of including the channels $k_1 k_2$ with a simultaneous transition to the analyzed state $\gamma_1 \gamma_2 \dots \gamma_{k-2} k_1 k_2$, although it decreases, but exists and should be considered.

We will first conduct the analysis for the flow of requests with the parameter λ_2 , finding the intensity of transitions $\alpha_{k^*}^2$, and present the result of this analysis in a compact notation:

$$\begin{aligned} \alpha_{k^*}^2 = & \lambda_2 \cdot [l_2 \cdot 1 + l_3 \cdot \sum_{i=0}^1 \frac{1}{C_{i+2}^2} \cdot C_{k-2}^{1-i} \cdot C_{n-k}^i + \\ & + l_4 \cdot \sum_{i=0}^2 \frac{1}{C_{i+2}^2} \cdot C_{k-2}^{2-i} \cdot C_{n-k}^i + \dots + \\ & + l_k \cdot \sum_{i=0}^{k-2} \frac{1}{C_{i+2}^2} \cdot C_{k-2}^{k-2-i} \cdot C_{n-k}^i + \dots + \\ & + l_n \cdot \sum_{i=0}^{n-2} \frac{1}{C_{i+2}^2} \cdot C_{k-2}^{n-2-i} \cdot C_{n-k}^i]. \end{aligned} \quad (18)$$

It can be noted that the last term is nonzero only at $i = n - k$. The final expression for the transition intensity $\alpha_{k^*}^2$ will take the following form:

$$\alpha_{k^*}^2 = \lambda_2 \cdot \sum_{j=1}^n l_j \cdot \sum_{i=0}^{j-2} \frac{1}{C_{i+2}^2} \cdot C_{k-2}^{j-2-i} \cdot C_{n-k}^i. \quad (19)$$

Let us consider the procedure for determining the second, “refusal” component $\alpha_{k^{**}}^2$ of the desired intensity.

When requests for streams with parameters $\lambda_r, r \geq 3$ are received, the transition to the state $\gamma_1 \gamma_2 \dots \gamma_{k-2} k_1 k_2$ is possible only when such a request enters an area accessible for servicing by channels $k_1 k_2$ and inaccessible for β -channels.

The number of such accessible areas for each value l_j ($j = 1, \dots, n$) of the overlap coefficients of the service devices accessibility zones is equal to:

$$\alpha_{k^{**}}^2 = \left[\sum_{j=0}^{k-2} l_{j+2} \cdot C_{k-2}^j \right] \cdot \sum_{r=3}^L \lambda_r. \quad (20)$$

Combining (19) and (20) for $k=2$ we find the desired transition intensity:

$$\begin{aligned} \alpha_k^2 = & \lambda_2 \cdot \sum_{j=1}^n l_j \cdot \sum_{i=0}^{j-2} \frac{1}{C_{i+2}^2} \cdot C_{k-2}^{j-2-i} \cdot C_{n-k}^i + \\ & + \left[\sum_{j=0}^{k-2} l_{j+2} \cdot C_{k-2}^j \right] \cdot \sum_{r=3}^L \lambda_r. \end{aligned} \quad (21)$$

Reasoning similarly and comparing expressions (15) and (21), we find the desired expression for the transition intensity in the general case:

$$\begin{aligned} \alpha_k^q = & \lambda_q \cdot \sum_{j=1}^n l_j \cdot \sum_{i=0}^{j-q} \frac{1}{C_{i+q}^q} \cdot C_{k-q}^{j-q-i} \cdot C_{n-k}^i + \\ & + \left[\sum_{j=0}^{k-q} l_{j+q} \cdot C_{k-q}^j \right] \cdot \sum_{r=q+1}^L \lambda_r, \quad q = \overline{1, n}; \quad k = \overline{1, n}. \end{aligned} \quad (22)$$

It should be noted that the arc with intensity α_k^q leaves the state $\gamma_1 \gamma_2 \dots \gamma_{k-q}$ belonging to the tier of the model graph with number $k-q$, and enters the state $\gamma_1 \gamma_2 \dots \gamma_{k-q} k_1 k_2 \dots k_k$ of the graph k -th tier.

To determine the final probabilities in the researched non-fully accessible QS with a non ordinary input flow of requirements groups, we will use the well-known [12]

property of Markov graphs. According to this property, in steady state, the total flow of movements along the ribs entering any closed contour on the graph is equal to the total flow of transitions along the edges leaving this contour. In this case, the flow of transitions along a separate edge of the graph is equal to the product of the intensity of transitions along this edge, for example μ for the state S_{34} in Fig. 3, by the probability of the state from which this edge comes out. Thus, for a state S_{34} , the flow of transitions along the outgoing edge can be obtained in the form: $\mu \cdot p_{34}$.

Next, we will use the noted property and, for the system of inserted contours in Fig. 3, we will compose equations for the balance of transition flows. In this case, for the probabilities of group (tiered) states, we use the probability symbol P , and for the probabilities of states within each tier, we use the probability symbol p and take into account the ratio of group probabilities and probabilities of states within each tier (Fig. 3):

$$\begin{aligned} P_0 &= p_0; & P_1 &= p_1 + p_2 + p_3 + p_4; \\ P_2 &= p_{12} + p_{13} + p_{14} + p_{23} + p_{24} + p_{34}; \\ P_3 &= p_{123} + p_{124} + p_{134} + p_{234}; \\ P_4 &= p_{1234}. \end{aligned} \quad (23)$$

For the inner contour enclosing the vertex S_0 , we obtain:

$$\begin{aligned} \mu p_1 + \mu p_2 + \mu p_3 + \mu p_4 &= \\ &= (4\alpha_1^1 + 6\alpha_2^2 + 4\alpha_3^3 + 1\alpha_4^4) \cdot p_0. \end{aligned} \quad (24)$$

Taking into account (23) and the number of channels $n = 4$, equality (24) can be represented in the following form

$$\mu P_1 = P_0 \cdot (C_n^1 \alpha_1^1 + C_n^2 \alpha_2^2 + C_n^3 \alpha_3^3 + C_n^4 \alpha_4^4). \quad (25)$$

For the remaining contours in Fig. 3, we omit the procedure for the transition from the probabilities of internal states to the probability of the group state, noting that such a transition did not require any hypotheses about the values of the internal states probabilities, we obtain:

$$\begin{aligned} 2\mu P_2 &= P_1(3\alpha_2^1 + 3\alpha_3^2 + 1\alpha_4^3) + \\ &+ P_0(6\alpha_2^2 + 4\alpha_3^3 + 1\alpha_4^4) = P_1(C_{n-1}^1 \alpha_2^1 + C_{n-1}^2 \alpha_3^2 + \\ &+ C_{n-1}^3 \alpha_4^3) + P_0(C_n^2 \alpha_2^2 + C_n^3 \alpha_3^3 + C_n^4 \alpha_4^4); \end{aligned} \quad (26)$$

$$\begin{aligned} 3\mu P_3 &= P_2(2\alpha_3^1 + 1\alpha_4^2) + P_1(3\alpha_3^2 + 1\alpha_4^3) + \\ &+ P_0(4\alpha_3^3 + 1\alpha_4^4) = P_2(C_{n-2}^1 \alpha_3^1 + C_{n-2}^2 \alpha_4^2) + \\ &+ P_1(C_{n-1}^2 \alpha_3^2 + C_{n-1}^3 \alpha_4^3) + P_0(C_n^3 \alpha_3^3 + C_n^4 \alpha_4^4); \end{aligned} \quad (27)$$

$$\begin{aligned} 4\mu P_4 &= P_3 \cdot 1\alpha_4^1 + P_2 \cdot 1\alpha_4^2 + P_1 \cdot 1\alpha_4^3 + P_0 \cdot 1\alpha_4^4 = \\ &= P_3 \cdot C_{n-3}^1 \alpha_4^1 + P_2 \cdot C_{n-2}^1 \alpha_4^2 + P_1 \cdot C_{n-1}^1 \alpha_4^3 + P_0 \cdot C_n^1 \alpha_4^4 \end{aligned} \quad (28)$$

In the general case, the equation for the balance of transition flows will take the following form:

$$k \cdot \mu \cdot P_k = \sum_{j=1}^k P_{k-j} \cdot \sum_{q=j}^{n-k+j} C_{n-k+j}^q \cdot \alpha_{k-j+q}^q. \quad (29)$$

Next, we will take into account the need for subsequent verification of correctness of the resulting analytical description of an incompletely accessible QS with input flow of non ordinary groups requirements, by degenerating them into already known and verified models, we will look for expressions for the final probabilities of group states in a form convenient for such verification:

$$\begin{aligned} P_k &= P_0 \cdot \frac{(\rho_1)^k}{k!} \cdot F_k; & k &= \overline{1, n}; \\ P_0 &= \left(\sum_{k=0}^n \frac{(\rho_1)^k}{k!} \cdot F_k \right)^{-1}. \end{aligned} \quad (30)$$

where for the functions F_k we'll need formulas:

$$\rho_q = \frac{\lambda_q}{\mu}, \quad q = \overline{1, n}. \quad (31)$$

To find an analytical calculation formula for the probability P_k deformation function F_k , we present the expression for the transition intensity (22) taking into account the transition probability d_k^q , and then we obtain:

$$\begin{aligned} \alpha_k^q &= \lambda_q \cdot \left[\sum_{j=1}^n l_j \cdot \sum_{i=0}^{j-q} \frac{1}{C_{i+q}^q} \cdot C_{k-q}^{j-q-i} \cdot C_{n-k}^i + \right. \\ &+ \left. \left(\sum_{j=0}^{k-q} l_{j+q} \cdot C_{k-q}^j \right) \cdot \frac{1}{\lambda_q} \cdot \sum_{r=q+1}^L \lambda_r \right] = \lambda_q \cdot d_k^q, \\ q &= \overline{1, n}; & k &= \overline{1, n}. \end{aligned} \quad (32)$$

Next, we substitute the expression for the final probabilities (30) into equality (29), and we obtain:

$$\begin{aligned} k \cdot \mu \cdot P_0 \cdot \frac{(\rho_1)^k}{k!} \cdot F_k &= \sum_{j=1}^k P_0 \cdot \frac{(\rho_1)^{k-j}}{(k-j)!} \cdot F_{k-j} \cdot \\ &\cdot \sum_{q=j}^{n-k+j} C_{n-k+j}^q \cdot \lambda_q \cdot d_{k-j+q}^q. \end{aligned} \quad (33)$$

After elementary transformations we obtain:

$$F_k = \sum_{j=1}^k F_{k-j} \cdot \frac{(k-1)!}{(\rho_1)^j (k-j)!} \cdot \sum_{q=j}^{n-k+j} C_{n-k+j}^q \cdot \rho_q \cdot d_{k-j+q}^q, \quad k = \overline{1, n}. \quad (34)$$

In order to find the value of a function F_0 , we substitute the value $k = 1$ into expression (29), and then we get:

$$\mu \cdot P_1 = P_0 \cdot \sum_{q=1}^n C_n^q \cdot \lambda_q \cdot d_q^q. \quad (35)$$

Then we substitute the probability value P_k from equality (30) into equality (35) at $k = 1$, and we obtain:

$$\mu \cdot P_0 \cdot \rho_1 \cdot F_1 = P_0 \cdot \sum_{q=1}^n C_n^q \cdot \lambda_q \cdot d_q^q. \quad (36)$$

From equality (36) and (31) we find:

$$F_1 = \frac{1}{\rho_1} \cdot \sum_{q=1}^n C_n^q \cdot \rho_q \cdot d_q^q. \quad (37)$$

On the other hand, from equality (36) with the value $k = 1$ we obtain:

$$F_1 = F_0 \cdot \frac{1}{\rho_1} \cdot \sum_{q=1}^n C_n^q \cdot \rho_q \cdot d_q^q. \quad (38)$$

Equating the right-hand sides of equalities (37) and (38)

$$\frac{1}{\rho_1} \cdot \sum_{q=1}^n C_n^q \cdot \rho_q \cdot d_q^q = F_0 \cdot \frac{1}{\rho_1} \cdot \sum_{q=1}^n C_n^q \cdot \rho_q \cdot d_q^q, \quad (39)$$

we find the value of the function F_0 :

$$F_0 = 1. \quad (40)$$

To check the correctness of expressions (22), (30) and (34) of the obtained analytical description of the non-fully accessible $M_L/M/n$ QS with a non ordinary input flow of requirements groups, we'll research the asymptotic transition of its description into the specification of a similar fully accessible $M_L/M/n$ QS with a non ordinary input flow of requirements groups, and then into the description of a similar non-fully accessible QS but with a simplest requirements flow.

When the analytical description of a non-fully accessible QS with a non ordinary input flow of requirements groups degenerates into a fully accessible QS, all values of accessibility probabilities for individual groups of service channels become equal to zero, except for the probability l_n :

$$l_k = \frac{\pi_k}{C_n^k} = 0, \quad k < n; \quad l_n = \frac{\pi_n}{C_n^n} = \pi_n = 1. \quad (41)$$

Then expression (22) degenerates into equality (42):

$$\alpha_k^q = \lambda_q \cdot l_n \cdot \sum_{i=0}^{n-q} \frac{1}{C_{i+q}^q} \cdot C_{k-q}^{n-q-i} \cdot C_{n-k}^i + [l_k] \cdot \sum_{r=q+1}^L \lambda_r, \quad q = \overline{1, n}; \quad k = \overline{1, n}. \quad (42)$$

The product of the combinations under the sum sign in equality (42) is different from zero and equal to one only when the value $i = n - k$. Then expression (42) takes the following form:

$$\alpha_i^q = \lambda_q \cdot l_n \cdot \frac{1}{C_{i+q}^q} + l_i \cdot \sum_{r=q+1}^L \lambda_r. \quad (43)$$

When substituting the value $i = k - j + q$ into equality (43), we obtain:

$$\alpha_{k-j+q}^q = \lambda_q \cdot l_n \cdot \frac{1}{C_{n-k+j}^q} + l_k \cdot \sum_{r=q+1}^L \lambda_r. \quad (44)$$

Next, we substitute the value α_{k-j+q}^q from equality (44) into equality (29) and, taking into account conditions (41), we find:

$$k \cdot \mu \cdot P_k = \sum_{j=1}^k P_{k-j} \cdot \sum_{q=j}^{n-k+j} C_{n-k+j}^q \cdot \alpha_{k-j+q}^q = \sum_{j=1}^k P_{k-j} \cdot \sum_{q=j}^{n-k+j} C_{n-k+j}^q \cdot [\lambda_q \cdot \frac{1}{C_{n-k+j}^q} + l_k \cdot \sum_{r=q+1}^L \lambda_r]. \quad (45)$$

In the internal sum in equality (44), we select the last term with the value $q = n - k + j$ and take into account the conditions (41). Then expression (45) will take the following form:

$$k \cdot \mu \cdot P_k = \sum_{j=1}^k P_{k-j} \cdot \left(\sum_{q=j}^{n-k+j-1} [\lambda_q \cdot \frac{C_{n-k+j}^q}{C_{n-k+j}^q} + 0] + C_{n-k+j}^{n-k+j} \cdot 0 \cdot \sum_{r=q+1}^L \lambda_r \right). \quad (46)$$

Finally, taking into account the accepted hypothesis $L = n$, we find:

$$k \cdot \mu \cdot P_k = \sum_{j=1}^k P_{k-j} \cdot \sum_{q=j}^L \lambda_q \cdot \quad (47)$$

Expression (47) completely coincides with the basic equality for finding the final probabilities P_k of group states in the $M_L/M/n$ QS with a non ordinary input flow of requirements groups (8).

Thus, when a found description of $M_L/M/n$ QS with incomplete approachability of service devices degenerates into the same system, but fully accessible, the found description is automatically transformed into a known description of QS $M_L/M/n$ with complete approachability of service devices. The noted phenomenon testifies in favor of correctness of the obtained analytical description in relation to simpler $M_L/M/n$ QS, which is a simplified version in relation to considered non-fully accessible QS with a non ordinary input flow.

The $M/M/n$ QS, with incomplete approachability of service devices and with an ordinary (simplest) input flow of requirements [4, formula (29)], taking into account the designations of variables, is described by the equation (8) of transitions flows balance (48):

$$k\mu P_k = P_{k-1} \cdot C_{n-k+1}^1 \cdot \alpha_k^1, \quad k = \overline{1, n} \quad (48)$$

In equalities (1) the probability $b_1 = 1$ and all other probabilities become equal to zero $b_r = 0, r = 2, \dots, L$.

Then the parameters of the partial flows of requirements will take the form $\lambda_1 = I; \lambda_r = 0, \text{ for } r \geq 2$, which will lead to changes in the found analytical description (22), where for all values $q \geq 2 \alpha_k^q = 0$.

On the right-hand side of (29), the inner sum turns out to be nonzero only when $j=1$, and expression (29) is transformed to the form:

$$k \cdot \mu \cdot P_k = \sum_{j=1}^k P_{k-j} \cdot \sum_{q=j}^{n-k+j} C_{n-k+j}^q \cdot \alpha_{k-j+q}^q = \left| \right|_{j=1} = P_{k-1} \cdot C_{n-k+1}^1 \cdot \alpha_k^1 \quad (49)$$

which coincides with the description (48) of a non-fully accessible $M/M/n$ QS, with refusals and with an ordinary (simplest) input flow of requirements. At the same time, expression (22) for α_k^q is transformed to the form:

$$\alpha_k^1 = I \cdot \sum_{j=1}^n I_j \cdot \sum_{i=0}^{j-1} \frac{1}{C_{i+1}^1} \cdot C_{k-1}^{j-1-i} \cdot C_{n-k}^i, \quad k = \overline{1, n}, \quad (50)$$

which completely coincides with the expression [4, formula (29)] (see here formula (8)) for a non-fully accessible $M/M/n$ QS, with refusals and with a simplest input flow.

As a result, when the found description of a non-fully accessible QS with a non ordinary input flow of requirements groups and with refusals degenerates into a non-fully accessible system with refusals, but with a simplest input flow, the found description is automatically transformed into a known description of a non-fully accessible $M/M/n$ QS with refusals and with an ordinary (simplest) input flow of requirements. The noted phenomenon testifies to the correctness of the obtained analytical description in relation to the simpler $M/M/n$ QS, which is a simplified version in relation to considered incompletely accessible QS with a non ordinary input flow. Thus, the application of the mathematical apparatus for analysis the groups of states of Markov graphs [12] made it possible to obtain analytical formulas for calculating values of states' final probabilities in process of servicing a non ordinary flow of requirements groups in a non-fully accessible system with refusals, which is a general case of previous types of QS.

4 EXPERIMENTS

In order to test the operability of analytical description of the incompletely accessible $M_L/M/n$ model, we use an example from the topical sphere – of important objects air defense (Fig. 5) by a grouping of four single-channel SAM systems (Table 1, No. 1, 2–3) – “service devices”, which should prevent a planned air blow of $N_{total.En} = 15$ enemy aircraft (Table 1, No. 2–8, 2–3) with a duration of 7,5 minutes and at intensity of 2 aircraft per minute (Table 1, No. 2–8, 2–3). Let's assume that shelling an aircraft in a fire zone ends with its destruction.

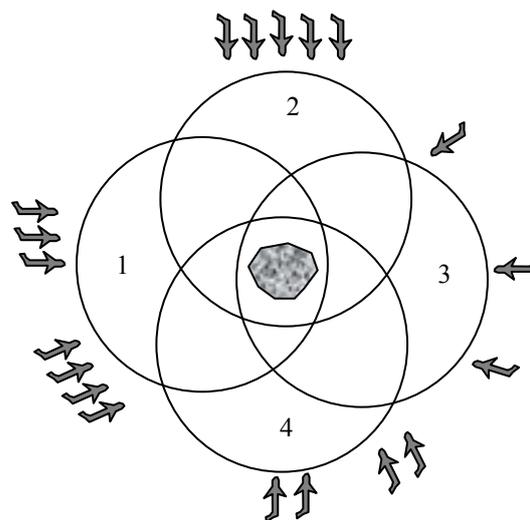


Figure 5 – An example of setting the task of assessing the objects' air defense effectiveness by the SAM systems' using the $M_5/M/4$ model of an incompletely accessible QS with input flow of requirement groups

To destroy an object covered, at least five aircraft are required. Therefore, the task of SAM systems grouping is considered fulfilled in the case when less than five aircraft can break through to the object ($N_{missed} < 5$).

In QS with incomplete approachability of service devices, refusal to service the next requirement is possible with any number of free devices. Therefore, to estimate the service probability, it is necessary to use information about the mathematical expectation of the occupied devices number $M_{b,d}$ and the performance μ of QS' one channel:

$$M_{b,d} = \sum_{k=0}^n k \cdot P_k; P_{service} = \frac{\mu \cdot M_{b,d}}{I}. \quad (51)$$

We also note that the mathematical expectation of the enemy aircraft number that broke through to the target (N_{missed}) can be found (52):

$$N_{missed} = N_{total} \cdot En \cdot (1 - P_{service}). \quad (52)$$

To perform the calculations, we use the initial data (Table 1 No. 1–24, 2–9), and the set of formulas (1), (41), (14), (17), (31). To calculate the missing expressions for the transition intensities, we use equality (22), and obtain:

$$\begin{aligned} \alpha_3^3 &= \lambda_3(l_3 + \frac{1}{4}l_4) + l_3 \cdot \sum_{r=4}^L \lambda_r; \\ \alpha_4^3 &= \lambda_3(l_3 + l_4) + l_4 \cdot \sum_{r=4}^L \lambda_r; \\ \alpha_4^4 &= \lambda_4(l_4) + l_4 \cdot \sum_{r=5}^L \lambda_r. \end{aligned} \quad (53)$$

Next, we use equalities (40) and (34). We find the missing expressions for the deformation functions F_2-F_4 of the final probabilities P_k of QS states group using the general recurrent expression (34), we obtain:

$$F_2 = F_1 \cdot \frac{1}{\lambda_1} \cdot (3\alpha_2^1 + 3\alpha_3^2 + 1\alpha_4^3) + \frac{(\mu)}{(\lambda_1)^2} \cdot (6\alpha_2^2 + 4\alpha_3^3 + 1\alpha_4^4); \quad (54)$$

$$F_3 = F_2 \cdot \frac{1}{\lambda_1} \cdot (2 \cdot \alpha_3^1 + 1 \cdot \alpha_4^2) + F_1 \cdot \frac{2\mu}{(\lambda_1)^2} \cdot (3 \cdot \alpha_3^2 + 1 \cdot \alpha_4^3) + F_0 \cdot \frac{2\mu^2}{(\lambda_1)^3} \cdot (4 \cdot \alpha_3^3 + 1 \cdot \alpha_4^4); \quad (55)$$

$$F_4 = F_3 \cdot \frac{1}{(\lambda_1)^1} \cdot \alpha_4^1 + F_2 \cdot \frac{3 \cdot (\mu)^1}{(\lambda_1)^2} \cdot \alpha_4^2 + F_1 \cdot \frac{6 \cdot (\mu)^2}{(\lambda_1)^3} \cdot \alpha_4^3 + F_0 \cdot \frac{6 \cdot (\mu)^3}{(\lambda_1)^4} \cdot \alpha_4^4. \quad (56)$$

We will calculate the probabilities of QS states group according to (30). To control the correctness of the obtained model $M_L/M/n$ with incomplete approachability of service devices, we'll use the well-known formulas of Erlang model [5]:

$$P_k = P_0 \cdot \frac{\rho^k}{k!}; k = \overline{1, n}; P_0 = \left(\sum_{k=0}^n \frac{\rho^k}{k!} \right)^{-1}. \quad (57)$$

Table 1 and Figures (Fig. 6–Fig. 9) present the calculation results..

For ease of analysis, Table 1 and Figures 6–9 introduce an abbreviated designation for the compared QS models, which allows us to establish a significant difference in the probability distribution of group states of the model developed in this research (Devel.) and known models (Req.gr., Inacces. and Erlang).

The known models do not provide for simultaneous consideration of the essential features of real QS – incomplete accessibility of service devices and the group composition of the input flow of requirements with an unknown composition of groups in advance, which leads to a significant distortion of the predicted results for QS operation.

Thus, (Fig. 9 and Table 1 No. 37) the excess in the assessment of the expected value of the efficiency indicator $P_{service}$, relative to the value obtained using the developed model, turns out to be at least one and a half times greater than the more realistic assessment using the developed model.

As a result, all known models allow us to assume that the SAM group will let through less than 5 enemy aircraft and thus reliably perform its task (Table 1 No. 38).

However, simultaneous consideration of the incomplete accessibility of service devices and the group composition of requirements in the input flow more objectively shows the significant inability of the SAM group to perform its task, since more than eight enemy aircraft can pass "without service" to the protected object (Table 1 No. 38, 12).

At the same time, the result of the experiment demonstrates the automatic degeneration of the deformation functions (F_0-F_4) into the functions of incomplete accessibility (Table 1 No. 27–31, 13) and in the function of non ordinary (Table 1 No. 27–31, 14), as well as in a single value (Table 1 No. 27–31, 15) with an automatic transition to the description of the final probabilities QS (Table 1 No. 32–36) of group states in the model with incomplete accessibility of service channels, in the model with group arrival of requirements and in the Erlang model.

The calculation expressions of the deformation functions F_0-F_4 are recurrent, which allows automating the calculation of their values for specific operating conditions and the specific configuration of QS.

Table 1 – Comparative assessment of service probability and task performance efficiency by the SAMS group (Fig. 5) using the developed and previous types of QS models with refusals

Names and values of models' parameters						Types of models								
#	Name	Value	#	Name	Value	#	Name	Value	#	Name	Devel.*	Inaccess.*	(Req.gr.)*	Erlang*
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	n	4	13	μ	1	15	α_1^1	0.087	27	F_0	1	1	1	1
2	I	2	14	ρ_1	0.133	16	α_2^1	0.120	28	F_1	5	1	5	1
3	λ	0.667	15	L	5	17	α_3^1	0.164	29	F_2	40.1	0.9	55	1
4	a_1	0.200	16	λ_1	0.133	18	α_4^1	0.340	30	F_3	372.204	0.486	912.5	1
5	a_2	0.200	17	λ_2	0.133	19	α_2^2	0.033	31	F_4	3408.55	0.247	19637.5	1
6	a_3	0.200	18	λ_3	0.133	20	α_3^2	0.056	32	P_0	0.451	0.178	0.360	0.143
7	a_4	0.200	19	λ_4	0.133	21	α_4^2	0.144	33	P_1	0.301	0.356	0.240	0.286
8	a_5	0.200	20	λ_5	0.133	22	α_3^3	0.021	34	P_2	0.161	0.321	0.176	0.286
9	π_1	0.4	21	l_1	0.1	23	α_4^3	0.061	35	P_3	0.066	0.115	0.130	0.190
10	π_2	0.3	22	l_2	0.05	24	α_4^4	0.037	36	P_4	0.020	0.029	0.093	0.095
11	π_3	0.16	23	l_3	0.04	25	$M_{b,d}$	0.9	37	$P_{service}$	0.452	0.731	0.678	0.905
12	π_4	0.14	24	l_4	0.14	26	M_{in}	3	38	N_{missed}	8.23	4.04	4.84	1.43

*Devel. – The developed QS $M_L/M/n$ with refusals with not fully accessible service channels and with input flow of requirements' groups.
 *Req.gr. – QS $M_L/M/n$ with input flow of requirements' groups and with refusals. The functions F_j automatically degenerate into known non-ordinary functions [3, formula (18)].
 *Inaccess. – QS $M/M/n$ with not fully accessible service channels and with refusals. The functions F_j automatically degenerate into known functions of incomplete accessibility [4, formula (28)].
 *Erlang – QS $M/M/n$ with refusals. The functions F_j automatically become equal to one [5]

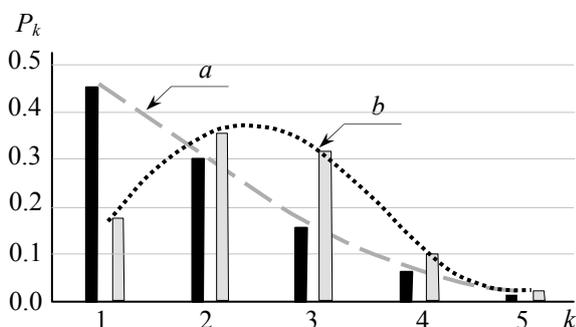


Figure 6 – Final probabilities P_k of states' groups in queuing systems with refusals and under identical conditions:
 a) the developed $M_L/M/n$ model;
 b) the well-known $M/M/n$ model. *Inaccess. – see (*) in Table 1,
 ($b_1 = 1; b_r = 0; \lambda_1 = I; \lambda_r = 0, \text{ for } r \geq 2$)

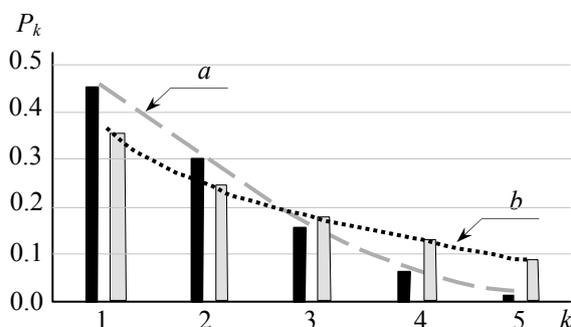


Figure 7 – Final probabilities P_k of states' groups in queuing systems with refusals and under identical conditions:
 a) the developed $M_L/M/n$ model;
 b) the well-known $M_L/M/n$ model. *Req.gr. – see (*) in Table 1, ($l_n = \pi_n = 1; l_k = 0, k < n, (41)$)

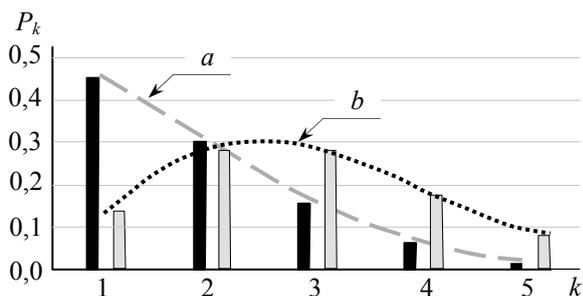


Figure 8 – Final probabilities P_k of states' groups in queuing systems with refusals and under identical conditions:
 a) the developed $M_L/M/n$ model;
 b) the well-known $M/M/n$ model. *Erlang – see (*) in table 1,
 ($l_n = \pi_n = 1; l_k = 0, k < n$;
 $b_1 = 1; b_r = 0; \lambda_1 = I; \lambda_r = 0, \text{ for } r \geq 2$)

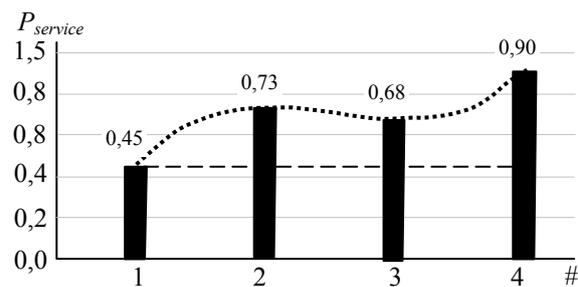


Figure 9 – Predicted value of performance indicator for a group of SAM systems (Fig. 5), using QS models:
 1) the developed $M_L/M/n$ model;
 2) $M/M/n$ model *Inaccess. – see (*) in table 1;
 3) $M_L/M/n$ model. *Req.gr. – see (*) in table 1;
 4) $M/M/n$ model. *Erlang – see (*) in table 1.

5 RESULTS

In the course of this research, analytical formulas for calculating the values of final probabilities in Markov QS © Gorodnov V. P., Druzhynin V. S., 2025
 DOI 10.15588/1607-3274-2025-1-4

$M_L/M/n$ with incomplete approachability of service devices were obtained for the first time, which makes it possible to evaluate and predict the values of all known



indicators of system efficiency. In this case, the deformation functions (F_k) of the states groups' probability distribution in QS have a recurrent form (34), which is convenient both for finding their analytical expressions and for performing numerical calculations. When the parameters of the QS operating conditions degenerate, the resulting description automatically goes into a description of one of a known QS: with refusals and with incomplete accessibility of service devices ($M/M/n$); into a description of a known QS with refusals, with full accessibility of service devices and with non-ordinary input flow of requirements groups ($M_L/M/n$), into a description of a known Erlang QS with refusals. During the numerical experiment, were obtained results (Table 1, Fig. 6–Fig. 9) that testify in favor of this statement. Thus, the noted types of QS turn out to be special cases of the analytical description of the final probabilities of QS $M_L/M/n$ with refusals, with partially accessible service devices and with an input flow of requirements groups with random composition obtained in this research.

CONCLUSIONS

In the course of the research, the analytical formulas for calculating numerical values of final probabilities of states in the $M_L/M/n$ QS with incomplete approachability of service devices for the input flow of requirements groups were received. The results of numerical experiment testify in favor of correctness the analytical formulas for calculating numerical values of final probabilities and in favor of possibility of their practical application in real QSs when solving problems of forecasting efficiency, as well as analyzing and synthesizing the parameters of real queuing systems.

The scientific novelty of the results obtained in research lies in the creation of possibilities for forecasting the effectiveness of known type of Markov queuing systems with refusals, with incomplete approachability of service devices and with an input flow of requirements groups with random composition. The obtained description (24), (27)–(29), (33) of a queuing system is a general for known QS with refusals, with not full approachability of service devices and with the simplest input flow of requirements ($M/M/n$), for known QS of $M_L/M/n$ type and for known Erlang model $M/M/n$.

The practical significance of the research results consists in obtaining analytical calculation formulas for performing rapid quantitative assessments of final probabilities, as well as all indicators of the efficiency of QS $M_L/M/n$ with incomplete approachability of service devices and its less complex QS variants in the course of solving practical problems of analysis, synthesis and management of real objects, for which such models can serve as a formal representation. The recurrent formulas obtained for calculating the values of the deformation functions of probabilities for states groups in QS are convenient for performing quick practical calculations.

Prospects for further research may be a research and development of an analytical description for final probabilities in Markov models with incomplete accessibility of service devices, with an input flow of groups of requirements and with waiting.

REFERENCES

1. Khinchin A. Ya. Raboty po matematicheskoy teorii massovogo obsluzhivaniya. Pod red. B. V. Gnedenko. Moscow, Fizmatgiz, 1963, 236 p.
2. Lidskiy E. A. Zadachi trafika v setyakh svyazi: Uchebnoye posobiye. Yekaterinburg, UGTU – UPI GOU VPO, 2006, 202 p.
3. Gorodnov V. The analytical description of final probabilities for states of queuing systems with input flow of groups of requirements, *Radio Electronics, Computer Science, Control*, 2019, №4 (51), pp. 25–37 DOI 10.15588/1607-3274-2019-4-3
4. Gorodnov V. P., Ovcharenko V. V. The states' final probabilities analytical description in an incompletely accessible queuing system with refusal, *Radio Electronics, Computer Science, Control*, 2022, №2 (61), pp. 32–42 DOI 10.15588/1607-3274-2022-2-4
5. Erlang A.K. The Theory of Probabilities and Telephone Conversations, *Nyt Tidsskrift for Mathematic Ser.*, B 20, 1909.
6. Plotkin B. K., Delyukin L. A.. Ekonomiko-matematicheskiye metody i modeli v kommercheskoy deyatelnosti i logistike: Uchebnik. SPb, Izd-vo, 2015, 345 p.
7. Albey E., Bilge U., Uzsoy R.: Multi-dimensional clearing functions for aggregate capacity modeling in multi-stage production systems, *International Journal of Production Research*, 2017, Vol. 55, Issue 14, pp. 4164–4179, Publisher Taylor & Francis. <https://doi.org/10.1080/00207543.2016.1257169>.
8. Karasev N. A., Kislukhina Ye. V., Vasil'yev V. A. Ispol'zovaniye teorii massovogo obsluzhivaniya v organizatsii lechebnogo protsessa na gospital'nom etape skoroy meditsinskoy pomoshchi, *Vestnik mezhdunarodnoy akademii nauk (Russkaya sektiya)*, 2006, pp. 65–68. <http://simulation.su/uploads/files/default/2006-karasev-kislukhina-vasiliev.pdf>
9. Afanasyeva L. G., Bulinskaya E. V. Mathematical models of transport systems based on queue theory, *Proceedings of MIPT*, 2010, V. 2, No. 4, pp. 6–10.
10. Zhao Z. Q., Hao J. X., Li L. J. Study on combat effectiveness of air defense missile weapon system based on queuing theory, *Proc. SPIE 10322, Seventh International Conference on Electronics and Information Engineering, 1032205 (23 January 2017)*. Nanjing, China; <https://doi.org/10.1117/12.2266077>
11. Ebadi M., Ahmadi-Javid A. Socio-economic design of control charts for monitoring service processes: a case study of a restaurant system, *Journal Quality Technology & Quantitative Management*, 2018. DOI: <https://doi.org/10.1080/16843703.2018.1519880>
12. Gorodnov V.P. Theoretical foundations of modeling microeconomic and other processes and systems: Monograph. Kh., Publishing house of the ABB Ministry of Internal Affairs of Ukraine, 2008, 484 p.

Received 10.12.2024.
Accepted 16.01.2025.

АНАЛІТИЧНИЙ ОПИС ФІНАЛЬНИХ ЙМОВІРНОСТЕЙ СТАНІВ У НЕПОВНО ДОСТУПНІЙ СИСТЕМІ ОБСЛУГОВУВАННЯ З ВІДМОВИМИ І З ВХІДНИМ ПОТОКОМ ГРУП ВИМОГ

Городнов В. П. – д-р військ. наук, професор, професор Національної академії Національної гвардії України, м. Харків, Україна.

Дружинін В.С. – д-р філософії, доцент кафедри паливного забезпечення Одеської військової академії, м. Одеса, Україна.

АНОТАЦІЯ

Актуальність. Основою створення та управління реальними системами масового обслуговування є можливість прогнозу їхньої ефективності. Для загального випадку таких систем з відмовами, з неповною доступністю приладів обслуговування та з випадковим складом груп вимог у вхідному потоці прогноз ефективності їх роботи залишається не вирішеною проблемою.

Метод. Для вищевказаних типів СМО ймовірності станів можна описати системою диференціальних рівнянь Колмогорова, яка для стаціонарного випадку перетворюється в однорідну систему лінійно залежних алгебраїчних рівнянь. Для реальних СМО в системах зв'язку кількість рівнянь може бути оцінена множиною-ступенем і досягати кількох тисяч, що породжує проблему їх запису та чисельного розв'язання для конкретного набору значень параметрів умов роботи. Прогностична цінність такого рішення не перевищує ймовірність вгадування числових значень параметрів умов роботи СМО, а для параметрів з безперервними значеннями, наприклад, для випадкових інтервалів часу між вимогами, дорівнює нулю.

Використаний метод заснований на аналітичному переході до опису груп станів СМО з однаковою кількістю зайнятих пристроїв. При цьому прагнення отримати кінцеві ймовірності станів у формі, наближеній до формул Ерланга, залишається. Вплив згаданих вище властивостей СМО можна локалізувати в окремих рекурентних функціях, які мультиплікативно спотворюють формули Ерланга.

Результати. Для вищевказаних типів СМО вперше знайдено аналітичні розрахункові формули для оцінки фінальних ймовірностей станів СМО, що дає змогу прогнозувати значення всіх відомих показників ефективності системи. У цьому випадку функції деформації розподілу ймовірностей груп станів у СМО мають рекурентний вигляд, що зручно як для знаходження їх аналітичних виразів, так і для чисельних розрахунків.

Коли параметри умов роботи СМО вироджуються, результуючий опис автоматично перетворюється на опис однієї з відомих СМО з відмовами, аж до СМО Ерланга.

Висновки. Знайдені аналітичні розрахункові вирази для фінальних ймовірностей вищевказаної СМО виявилися застосовними до всіх типів Марківської СМО з відмовами, що підтверджено результатами чисельного експерименту. У результаті стало можливим практично застосовувати отриманий аналітичний опис розглянутої СМО для оперативних оцінок ефективності розробленої та існуючої СМО в можливому діапазоні умов їх функціонування.

КЛЮЧОВІ СЛОВА: Марківські системи масового обслуговування, групи вимог.

ЛІТЕРАТУРА

1. Хинчин А. Я. Работы по математической теории массового обслуживания; под ред. Б.В. Гнеденко / А. Я. Хинчин. – М. Физматгиз. – 1963. – 236 с.
2. Лидский Э.А. Задачи трафика в сетях связи: Учебное пособие. – Екатеринбург: УГТУ – УПИ ГОУ ВПО, 2006. – 202с.
3. Gorodnov V. The analytical description of final probabilities for states of queuing systems with input flow of groups of requirements / V. Gorodnov // Radio Electronics, Computer Science, Control. – 2019. – № 4 (51). – P. 25–37 DOI 10.15588/1607-3274-2019-4-3
4. Gorodnov V. P. The states' final probabilities analytical description in an incompletely accessible queuing system with refusal / V. P. Gorodnov, V. V. Ovcharenko // Radio Electronics, Computer Science, Control. – 2022. – №2 (61). – P. 32–42. DOI 10.15588/1607-3274-2022-2-
5. Erlang A. K. The Theory of Probabilities and Telephone Conversations. / A. K. Erlang // Nyt Tidsskrift for Mathematic Ser. B 20. – 1909.
6. Плоткин Б. К. Экономико-математические методы и модели в коммерческой деятельности и логистике: учебник / Б. К. Плоткин, Л. А. Делюкин. – СПб. : Изд-во, 2015. – 345 с.
7. Albey E. Multi-dimensional clearing functions for aggregate capacity modeling in multi-stage production systems / E. Albey, U. Bilge, R. Uzsoy // International Journal of Production Research. – 2017. – Vol. 55 Issue 14. – P. 4164–4179, Publisher Taylor & Francis. <https://doi.org/10.1080/00207543.2016.1257169>.
8. Карасев Н. А. Использование теории массового обслуживания в организации лечебного процесса на госпитальном этапе скорой медицинской помощи / Н. А. Карасев, Е. В. Кислухина, В. А. Васильев // Вестник международной академии наук (Русская секция). – 2006. – С. 65–68. <http://simulation.su/uploads/files/default/2006-karasev-kisluhina-vasiliev.pdf>
9. Афанасьева Л. Г. Математические модели транспортных систем, основанные на теории очередей. / Л. Г. Афанасьева, Е. В. Булинская // Труды МФТИ. – 2010. – Т. 2, № 4. – С. 6–10.
10. Zhao Z. Q. Study on combat effectiveness of air defense missile weapon system based on queuing theory / Z. Q. Zhao, J. X. Hao, L. J. Li // Proc. SPIE 10322, Seventh International Conference on Electronics and Information Engineering, 1032205 (23 January 2017). – Nanjing, China; <https://doi.org/10.1117/12.2266077>
11. Ebadi M. Socio-economic design of control charts for monitoring service processes: a case study of a restaurant system. / M. Ebadi, A. Ahmadi-Javid // Journal Quality Technology & Quantitative Management. – 2018. DOI: <https://doi.org/10.1080/16843703.2018.1519880>
12. Городнов В. П. Теоретические основы моделирования микроэкономических и других процессов и систем: Монография / В. П. Городнов. – Х. : Изд-во АБВ МВД Украины, 2008. – 484 с.

METHOD FOR DETERMINING THE STRUCTURE OF NONLINEAR MODELS FOR TIME SERIES PROCESSING

Pysarchuk O. O. – Dr. Sc., Professor, Professor of the Department of Computer Engineering, Faculty of Informatics and Computing, National Technical University of Ukraine “Ihor Sikorski Kyiv Polytechnic Institute”, Kyiv, Ukraine.

Tuhanskykh O. A. – Assistant of the Department of Computer Engineering, Faculty of Informatics and Computing, National Technical University of Ukraine “Ihor Sikorski Kyiv Polytechnic Institute”, Kyiv, Ukraine.

Baran D. R. – Assistant of the Department of Computer Engineering, Faculty of Informatics and Computing, National Technical University of Ukraine “Ihor Sikorski Kyiv Polytechnic Institute”, Kyiv, Ukraine.

ABSTRACT

Context. The practice of today’s problems actualizes the increase in requirements for the accuracy, reliability and completeness of the results of time series processing in many applied areas. One of the methods that provides high-precision processing of time series with the introduction of a stochastic model of measured parameters is statistical learning methods. However, modern approaches to statistical learning are limited, for the most part, to simplified polynomial models. Practice proves that real data most often have a complex form of a trend component, which cannot be reproduced by polynomials of even a high degree. Smoothing of nonlinear models can be implemented by various approaches, for example, by the method of determining the parameters of nonlinear models using the differential spectra balance (DSB) in the scheme of differential-non-Taylor transformations (DNT). The studies proved the need for its modification in the direction of developing a conditional approach to determining the structure of nonlinear mathematical models for processing time series with complex trend dynamics.

Objective. The development of a method for determining the structure of nonlinear by mathematical models for processing time series using DSB in DNT transformations.

Method. The paper develops a method for constructing nonlinear mathematical models in the DNT transformation scheme. The modification of the method consists in controlling the conditions for the formation of a certain system of equations in the DSB scheme to search for the parameters of a nonlinear model with its analytical solutions. If the system is indeterminate, the nonlinear model is supplemented by linear components. In the case of an overdetermined system, its solution is carried out using the least squares norm. A defined system is solved by classical approaches. These processes are implemented with the control of stochastic and dynamic accuracy of models in the areas of observation and extrapolation. If the results of statistical learning are unsatisfactory in accuracy, the obtained values of the nonlinear model are used as initial approximations of numerical methods.

Result. Based on carried-out research, a method for determining the structure of nonlinear models for processing time series using BDS in the scheme of DNT transformations is proposed. Its application provides a conditional approach to determining the structure of models for processing time series and increasing the accuracy of estimation at the interval of observation and extrapolation.

Conclusions. The application of the proposed method for determining the structure of nonlinear models for processing time series allows obtaining models with the best predictive properties in terms of accuracy.

KEYWORDS: data science, statistical learning, time series, nonlinear models, numerical methods, least square method.

ABBREVIATIONS

ARMA – autoregressive moving average;
ARIMA – autoregressive integrated moving average;
DNT – differential-non-Taylor transformations;
DS – differential spectrum;
DSB – is a balance of differential spectra;
EMA – exponential moving average;
MA – moving average;
OLS – ordinary least squares.

NOMENCLATURE

Δ – linear deviation of the model;
 Φ – OLS matrix algorithm;
 $\varphi_i(t)$ – basis model functions;
 a_i – parameters of the nonlinear model;
 c_i – free polynomial coefficients e.g. in the format of a step basis;
 d – distance control;
 $f(t, a)$ – nonlinear model;

H – segment of the argument on which the function is considered;

k – ordinal number of the discrete spectrum;

n – time series size (sample size);

$P\{\dots\}_t^*$ – direct differential transformation;

R^2 – determination coefficient of the model;

t – function argument;

t^* – specific value of the argument at which the conversion is performed;

y_n – n -th dimension (element of the time series – dimension);

$z(t)$ – polynomial model;

$Z(t)$ – discrete argument function $k = 0, 1, 2, \dots$;

$\hat{z}(t)$ – trained polynomial model;

$\hat{Z}(k), F(k)$ – images of models.

INTRODUCTION

One of the fastest growing areas in the field of modern information technology is undoubtedly Data Science. Currently, methods and technologies of data research play a key role in e-commerce software systems (trading, retailing, aggregation), computer systems of automated and automatic control (unmanned systems, traffic control systems), etc. [1, 4]. These applied fields often operate with data in the time series format. Their processing is quite illustratively presented: approximation methods (Moving Average (MA) algorithms, Exponential Moving Average (EMA), autocorrelation algorithms such as ARIMA); statistical learning methods (smoothing) – Statistical Learning (Ordinary Least Square Method (OLS), Kalman filtering, etc.); methods of deep learning using artificial neural networks [1–7].

Real data in the form of time series, as a rule, depends on many factors that are difficult to describe. This introduces errors into discrete measurements that are attributed to the model randomness. Therefore, despite the successful application of approximation and deep learning methods in practice, statistical learning methods are quite effective from the point of view of accuracy, reliability of the result and productivity of calculations.

Traditionally, Statistical Learning methods use linear models in the form of power polynomials. But the reality is that most of the studied processes are non-linear in nature, thus a priori limiting the possibility of using this approach. That is why approximation and deep learning methods are sometimes preferred, which, in fact, reproduce the stochastic process. There are quite a few approaches that allow us to partially solve the problem of nonlinear smoothing (linearization, numerical methods) [1–3]. Most of them consist in simplifying a nonlinear model to a set of linear components or require initial conditions for starting iterative processes of “fitting” the parameters of nonlinear models to the data. However, in practice, such models lose both most of their nonlinear information and their own predictive value.

Therefore, the task of developing an approach to determining the parameters of a nonlinear model based on the data of the time series format using statistical learning methods is relevant.

The object of study is the process of determining the structure of nonlinear mathematical models for processing time series.

The subject of study is methods of processing time series with nonlinear models in terms of parameters

The purpose of the work is to develop a methodology for determining the structure of nonlinear mathematical models for processing time series using DSB in DNT transformations.

1 PROBLEM STATEMENT

Suppose while observing a certain process, a time series is y obtained that is described as:

$$y = \{y_0, y_1, y_2, \dots, y_n\}. \quad (1)$$

This data may be heterogeneous, contaminated, and have anomalous measurements, but all this must be addressed at the data preparation stage. Therefore, we further assume that the time series (1) is homogeneous, and its values are normally distributed with standard deviation σ (base random error).

The general form of the nonlinear approximating function is known:

$$f(t, a), a = a_0, a_1, \dots, a_i, \dots, a_m, \quad (2)$$

which with a high degree of adequacy describes the process under study, represented by a set of discrete dimensions (1). The task of processing the time series (1) by statistical learning methods is to determine the parameters a of the nonlinear model (2) with the requirement to reduce the random measurement error (1). The result is a nonlinear mathematical model, consistent – “trained” from measurements (1), which describes the law of trend change of the process under study with reduction – smoothing of the random error of the input data [12].

However, classical methods of statistical learning are based on the use of linear models – polynomials. The nonlinearity of such approaches is reproduced by adding high-stage components to the structure of the model. This approach shows good results within the measurement sample (1), but the predictive properties in terms of accuracy (point and interval) and the forecast interval are not the best. Nonlinear in terms of parameters models (2) overcome these disadvantages of polynomial forms [7, 10, 11].

To determine the parameters of a nonlinear model, it is proposed, first, to use OLS to train a polynomial model $\hat{z}(t)$:

$$\hat{z}(t) = \Phi(y), \quad (3)$$

$$z(t) = \sum_{i=0}^m c_i \Phi_i(t), \quad (4)$$

So, we have two models $f(t, a)$ and $\hat{z}(t)$ are, respectively, theoretical and experimental, and the parameters of the second are known. To determine the parameters of the theoretical nonlinear model, it is necessary to perform its approximation to the experimental one. That is, to transfer the properties of one to the other: the certainty of the parameters $\hat{z}(t)$, which are obtained from sample (1), on a nonlinear model that is uncertain in terms of parameters $f(t, a)$. One method to accomplish this is DSB.

The balance of differential spectra is based on the differential transformations described in [7]. In general, they can be presented as:

$$Z(k) = P\{z(t)\}_t^* = \frac{H^k}{k!} \left[\frac{d^k z(t)}{dt^k} \right]_t^*, \quad (5)$$

$$z(t) = \sum_{k=0}^{\infty} \left(\frac{t}{H}\right)^k Z(k). \quad (6)$$

Expression (5) provides the ability to get an image $Z(k)$ from its original $z(t)$ (direct transformation). Inverse conversion (6) provides the ability to restore the original $z(t)$. Differential image $Z(k)$ is called the differential spectrum (DS) or a P-spectrum, and the value of the $Z(k)$ for specific values of the argument – DS discretizes (P-discretizes). The reconstruction of the original is reduced to the summation of the discrete P-spectrum in the form of a segment of the Taylor series. ex. (6) (basic or differential-Taylor (DT) transformations). If the restoration of DS is carried out on an arbitrary (non-Taylor) basis, such transformations are called differential-non-Taylor (DNT). It is possible to implement DNT by equating the discrete of the same name to the original function $Z(k)$ and functions from the selected basis, e.g. $f(t, a)$. This is the essence of the DSB method.

Combining the properties of an experimental model $\hat{z}(t)$ and a nonlinear model $f(t, a)$ according to the DSB in the DNT scheme is carried out according to the model:

$$\begin{aligned} [P\{\hat{z}(t)\}_t^* \Rightarrow \hat{Z}(k)] = \\ = [P\{f(t, a)\}_t^* \Rightarrow F(k, a)] \\ \rightarrow \hat{Z}(k) = F(k, a). \end{aligned} \quad (7)$$

Determination of the parameters a of a nonlinear model $f(t, a)$ is implemented by forming and solving a system of equations:

$$\begin{cases} \hat{Z}(0) - F(0, a_1) = 0, \\ \hat{Z}(1) - F(1, a_2) = 0, \\ \hat{Z}(2) - F(2, a_3) = 0, \\ \dots \\ \hat{Z}(k) - F(k, a_m) = 0. \end{cases} \quad (8)$$

The processes described represent the essence of the method of constructing nonlinear mathematical models in the DNT transformation scheme [7].

The problem of the practical application of the DSB is the need to comply with a number of requirements: DS $\hat{z}(t)$ and $f(t, a)$ must give a definite system of equations (8); model $\hat{z}(t)$ must be adequate to the dynamics of data changes (1); model $f(t, a)$ should not have zero discrete for at the interval of existence of the DS model $\hat{z}(t)$. Such requirements can only be met in partial cases.

Therefore, the paper proposes the development of a method for constructing nonlinear mathematical models in the DNT transformation scheme by developing a conditional approach to determining the structure of nonlinear models for processing time series.

2 REVIEW OF THE LITERATURE

The focus of the analysis of existing approaches to the processing of time series with nonlinear models is placed on the methods of statistical learning. It should be noted that we will consider nonlinear models where the reproduction of the nonlinearity (seasonality, fluctuations, etc.) of the trend is ensured by introducing a nonlinear function into the structure of the model – exponential, logarithmic, trigonometric, differential components, etc. In this case, nonlinear operations are implemented on the parameters of the model, and the models are nonlinear in terms of parameters. Otherwise, the models will be considered linear in parameters, although capable of reproducing nonlinear time series. The advantage of nonlinear models over linear ones is the smaller number of structural components of the former. That is, the reproduction of the nonlinearity of data is assumed by a nonlinear function, and not a significant number, for example, of high-degree polynomial elements.

Currently, there are quite a few approaches to reproducing nonlinear models, but they can be divided into three classes: linearization; calculation of parameters by iterative numerical methods; operator methods. The best efficiency was proved by methods based on operator transformations. In this direction, [4, 6, 10, 11] proposes an approach to determining the parameters of nonlinear models using the method of differential transformations [7]. However, its shortcomings lead to the possibility of effective application to a rather narrow range of partial cases.

This determines the formed goal of research on the development of an approach that will unify the use of operator methods to determine the parameters of nonlinear models consistent with the time series of the format (1).

3 MATERIALS AND METHODS

The use of DSB to determine the parameters of a nonlinear model implements the transfer of properties of a simplified polynomial model to a complex nonlinear model. This means a priori that in the observation area, or in the provided time series, this polynomial must adequately describe the process itself sufficiently, otherwise the transfer of low-quality characteristics will be performed. That is, within the time series (1), the characteristics of the curvature of the trend must be comparable with the radius of convergence of the polynomial of the selected order, otherwise it simply cannot cope with the nonlinearity of the process under study. Or vice versa, for a time series with significant nonlinearity of trend and volume, one should have a polynomial of comparable radius of convergence, which is not always possible. Because in this case, the number of calculation operations for statistical training will increase and random errors of polynomial coefficients will accumulate. It should be noted that the indicators of the radius of convergence will

have a wide range of variation depending on the shape and physical essence of the process under study. For asymptotic processes, the convergence of the model will be supported by small degrees of the polynomial (up to the third and fourth orders) trained on relatively small samples. For processes with seasonal repetitions, periodic polynomials and polynomials with significant nonlinearity that are acceptable due to computational complexity will not give adequate processing results.

Such a limitation is critical for the DSB (7) and a priori makes it impossible to use some models that do not have enough non-zero components in the differential spectrum to form a certain system of equations (8). An example of this is nonlinear models:

$$f(t,a) = a_0 \sin(a_1 t) + a_2 \cos(a_1 t), \quad (9)$$

$$f(t,a) = a_0 + a_1 t + a_2 e^{a_3 t}. \quad (10)$$

For models (9), (10) we have only three unknown parameters $\{a_0, a_1, a_2\}$, i.e., for a defined system of type (8), three discrete DSs are sufficient and, accordingly, it is sufficient to have a third-order polynomial. But such a polynomial is effective only for a limited time interval and trend curvature of a time series characterized by models (8), (7).

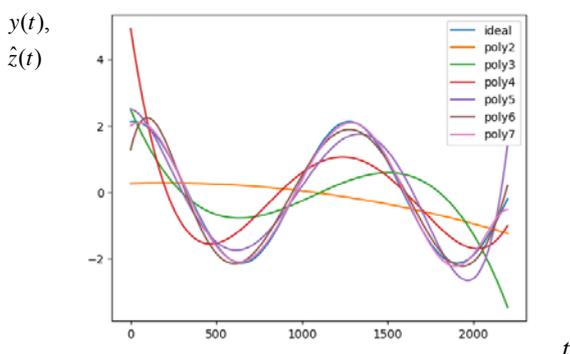


Figure 1 – Periodic function charts: ideal $y(t)$ and with polynomial models $\hat{z}(t)$, trained according to the OLS algorithm

Thus, the DSB in the DNT scheme (7) limits the requirements for structural optimization of the approximation model with control of the concordance coefficient, linear deviation, and high-order derivatives [1]. It is generally characteristic of periodic processes/functions to have many cycles, and they can be represented within an empirically assembled time series (1) as a superposition of many processes (factors).

Fig. 1 shows graphs of reproduction of periodic time series [8] by polynomial models trained by the OLS algorithm. The periodic series was modeled as a perfect trend.

The presented results clearly demonstrate the obvious consequences. Increasing the order of the polynomial model to a certain value provides an increase in its ade-

quacy (according to absolute deviation Δ and concordance coefficient R^2 – see Table 1).

Table 1 – Results of model adequacy assessment

n	2	3	4	5	6	7
Δ	1.439	1.169	0.835	0.394	0.198	0.064
R^2	9.050	0.500	0.559	0.927	0.983	0.998

At the same time, a significant increase in the structure of the polynomial model leads to a decrease in the rate of improvement in the results of the use of OLS. This is due to a decrease in the absolute value of the high-degree coefficients of the polynomial, which gives “sensitive” solutions from the standpoint of the accumulation of calculation errors, the number of calculation operations and the influence of random errors of input data. Therefore, for a time series with any trend complexity, including a periodic structure, it is potentially possible to determine the optimal (with a minimum Δ and maximum R^2 and with acceptable estimates of accuracy) structure of a polynomial model. For example, the optimal order of the model $m = 6$.

However, such a decision on the order of a polynomial on the example of a nonlinear model with transcendental functions (9) that reproduce the periodic properties of the time series contradicts with the requirements of the DSB. That is, the number of significant (all non-zero – seven) discrete DSs of a polynomial model is greater than the number of necessary and sufficient (three) discrete models (9). The loss of discrete DSs of a polynomial optimal in terms of structure will lead to the negligence of the positive properties of the nonlinear model in terms of parameters. Thus, we have a redefined system of equations of the form (7).

These statements are explained by the following calculations.

$$P\{\hat{z}(t)\}_t^* \Rightarrow \hat{Z}(k) = \{\hat{Z}(0), \hat{Z}(1), \dots, \hat{Z}(7)\}, \quad (11)$$

$$P\{f(t,a)\}_t^* \Rightarrow F(k,a) = \{F(0,a), F(1,a), \dots, F(7,a)\}^T, \quad (12)$$

$$\{\hat{Z}(0) = F(0,a)\}. \quad (13)$$

To solve the redefined system of equations (11), it is proposed to use an OLS with a numerical approximation algorithm, which in general has a norm for minimizing the quadratic residual:

$$\sum_{i=1}^m [\hat{Z}_i(i) - F_i(i,a)]^2 \rightarrow \min. \quad (14)$$

Minimization of the quadratic form (14) with respect to the parameters of the nonlinear model (see general form of (2) and specific examples (9), (10) considering the known parameters of the polynomial model (see (3),

(4) proposed to be carried out using the Leverberg-Marhardt algorithm [2]. This algorithm is presented in most software libraries with an interface for a wide range of modern high-level programming languages. Importantly, the Leverberg-Marhardt algorithm gives stable solutions of minimization forms (14) for approximation models of any complexity. To establish initial approximations, it is proposed to choose the solution of the system of equations (12), (13) with the constraints of the DSB (7), (8). That is, as a solution for models (9), (10) of a definite system of form (7) from three equations.

The use of numerical methods gives an approximate and partial solution – the accuracy of which depends on the computational costs of iterative calculation processes. Therefore, such a solution is not productive, especially for high-stage models and for large information arrays.

Consequently, the determination of the optimal polynomial structure by the properties of the time series with the control of the indicators in Table 1 is not always acceptable from the standpoint of the computational complexity of the solution of the system of equations (12), (13). Moreover, the goal of processing the time series in the presented studies is a high-precision nonlinear model (2), where the polynomial (3) is an intermediate result and a means to achieve the goal [11]. At the same time, it is possible to assume that the shortcomings of the polynomial model (stochastic and dynamic errors) in transferring its properties to a nonlinear model in the DSB scheme (7), (8) can be partially compensated by the minimizing form of OLS (14). Therefore, it is important to control not so much the adequacy indicators of the polynomial model Δ and R^2 , how many indicators of adequacy of a nonlinear model in terms of parameters [12]. However, to implement such an operation, it is necessary to investigate, identify and prove the dependence of the indicators of adequacy of models $f(t, a)$, $\hat{z}(t) - \Delta$ and R^2 :

$$\Delta = \frac{1}{n-1} \sum_{i=0}^n (y_i - \hat{y}_i)^2, \quad (15)$$

$$R^2 = 1 - \frac{\sum_{i=0}^n (y_i - \hat{y}_i)^2}{\sum_{i=0}^n (y_i - \bar{y}_i)^2}. \quad (16)$$

Research conditions. To obtain the time series, an additive mixture of an ideal trend with random noise is implemented [9]. To model an ideal trend, nonlinear models with the following parameters were used:

– Trigonometric:

- model: $f(t) = 2.2 \sin(0.01t)$;
- sample size: 1000 times;
- random noise parameters: normal law of distribution, standard deviation $\sigma = 0.2c.u.$

– Exponential:

- model: $f(t) = 0.001t + e^{0.0001t}$;
- sample size: 10000 times;
- random noise parameters: normal law of distribution, standard deviation $\sigma = 5c.u.$

The processing of time series was carried out by constructing nonlinear mathematical models in the DNT scheme of transformations according to the DSB [6].

The dependencies of indicators (15), (16) for polynomial models of form (4) and nonlinear models (9), (10) were studied. The latter have an infinite DC, which makes it possible to test polynomial models of any degree.

The research results are presented in Figs. 2–4 [8].

Fig. 2. The curves and realizations of dependencies between indicators (15) and (16) for the polynomial model (4) and the nonlinear transcendental model (9) are depicted. On the axis y metrics are given for the nonlinear, and on the axis x for a polynomial model. For research by the Monte Carlo method, 100 implementations of the method for determining nonlinear by mathematical models for using DSB in the DNT scheme were carried out. The result is the dispersion of the dependence of indicators (15), (16) with a clear correlation.

Analogous dependencies for the nonlinear exponential model (10) are shown in Fig. 3.

The above graphs clearly demonstrate the presence of a correlation between the structural properties of the polynomial and nonlinear models. The variation that is present in the results is due to random errors in the data. The graphs also show: the horizontal line is the median of the scattering, and the arc corresponds to the quadratic approximation of the resulting correlation dependence. The latter visualizes the nature of the dependence of indicators (15), (16), polynomial (4) and nonlinear models (9), (10).

Analysis of the graphs of Figs. 2, 3 proves that the structure of the nonlinear model under study affects the stochastic properties of scattering, i.e. the size of the confidence domain.

Compared to the transcendental models (9), the exponential model (10) demonstrates a high correlation between parameters (15), (16) (see Fig. 3 for a linear trend with a slope) and a relatively small confidence interval of scattering (concentration of realizations relative to the linear trend from the comparison of Figs. 2, 3). In practice, this gives and explains the stability of the solution of the DSB system in the DNT, including in the presence of random measurement errors.

At the same time, the transcendental model (10) is relatively unstable in terms of R^2 (see Fig. 2b).

As a result, we have the final confirmation of the informativeness and the conditioned properties of indicators (15), (16), which are suitable for the formation of the structure of models for DBS in the DNT, as a requirement for minimization Δ and maximization of R^2 .

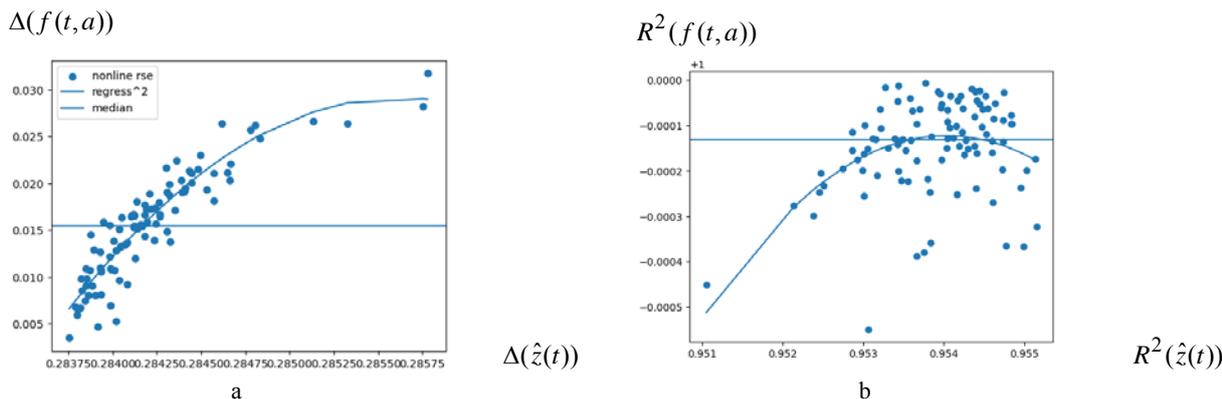


Figure 2 – Depiction of the dependence of model parameters from DSB in DNT: a – indicator (15) for models (4) and (9), b – indicator (16) for models (4) and (9)

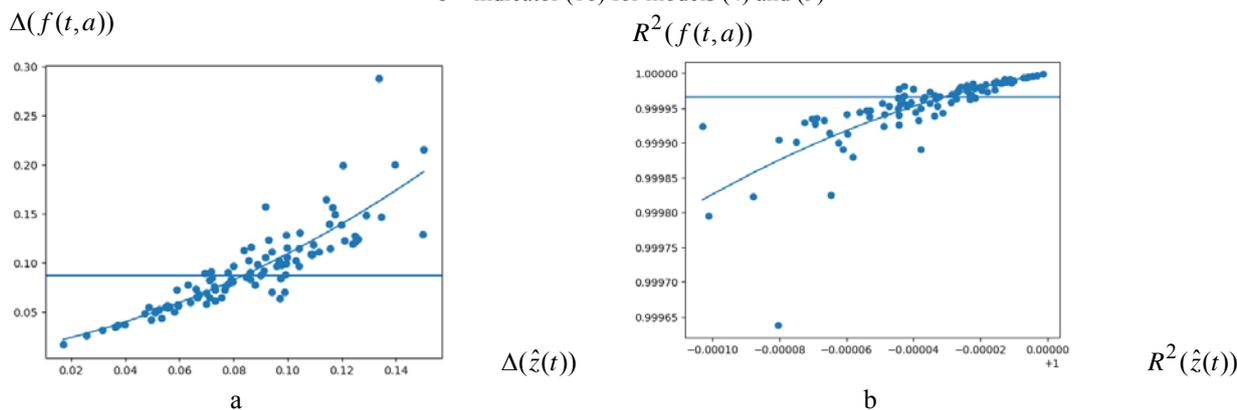


Figure 3 – Depiction of the dependence of model parameters from DSB in DNT: a – indicator (15) for models (4) and (10), b – indicator (16) for models (4) and (10)

Thus, optimization of the structure of the polynomial model with the control of indicators (15), (16) improves the properties of the nonlinear model determined from the DSB in the DNT scheme. However, optimization of a polynomial model is possible, albeit with big data, but on a limited time series (1). That is, the volume of the time series through the composition of nonlinear oscillations (seasonality) directly determines the order of the polynomial. Therefore, it is worth investigating the influence of the sample size (1) – the time series interval on the properties of the polynomial model.

Fig. 4 shows the results of studies of the dependence of the indicator (15) – control of the dynamic error of the model on the time series interval n for a polynomial of

the 6th (Fig. 4a) and 10th (Fig. 4b) orders. A perfect (without random errors) sample generated by the following model was subject to modeling: $f(t) = 2.2 \sin(0.00t)$. The choice of indicator (15) is due to its sensitivity to the relationship between the structure of the model and the sample size, while indicator (16) is responsible for the probabilistic ratio – relative to the stochastic component of accuracy.

We have the following local results. For a 6th-order polynomial, the minimum $\Delta(\hat{z}(t))$ is achieved for $n = 780$, and for a polynomial of degree 10 – for $n = 2100$. Therefore, the values of indicators (15) and (14) should be correlated with the size of the time series.

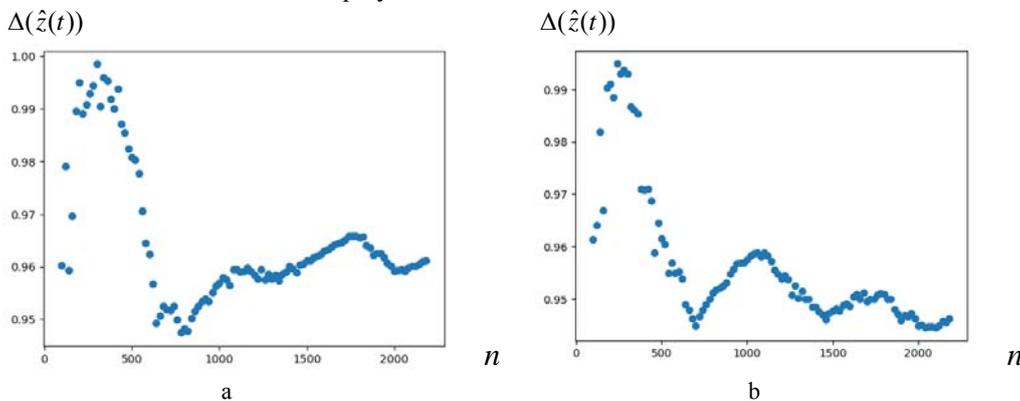


Figure 4 – Graph of the indicator Δ dependence on the size of the model the polynomial rank: a – 6th, b – 10th

Based on the results of the research, it is possible to formalize *the methodology for determining the structure of nonlinear models for processing time series using DSB in DNT transformations.*

1. Form a polynomial model (4), the order of which is equal to the number of unknown parameters of the nonlinear model (2).

2. Based on the results of step 1, using the classical OLS, determine the parameters of the polynomial model (3) from the input time series (1) and evaluate its indicators (15), (16).

3. Repeat step 2 of the methodology for the polynomial model above in the order specified in step 1, with the control of indicators (15), (16). If the distance d (see (17)) between its indicators and the previous ones (obtained in step 2 of the methodology) decreases – continue to increase the order of the model until the distance begins to increase.

$$d = |\Delta(\hat{z}_{i-1}(t) - \hat{z}_i(t))| \times |R^2_{i-1}(t) - R^2_i(t)|. \quad (17)$$

Distance control (17) – indirectly allows you to consider the volume of the time series.

4. If the polynomial model obtained under the conditions of step 3 has an order equal to the number of parameters of the nonlinear model (3), then a system of equations of the form (7) should be formed, which is solved analytically with respect to the desired parameters of the nonlinear model.

5. If the polynomial model obtained under the conditions of step 3 is of an order of magnitude higher than the number of parameters of the nonlinear model (3), then a system of equations of the form (12), (13) is formed, which is solved by the chosen numerical method. To establish initial approximations, a system of equations (7) with DSB constraints is formed.

6. To implement the formation of a nonlinear model (3) from the parameters specified in step 4 or 5.

7. To apply the nonlinear model (3) (because of step 6 of the methodology) to estimate the parameters of the stochastic time series (1) in the observation area and in the forecasting interval – the formation of the current and forecast trend.

8. Evaluate indicators (15), (16) of the nonlinear model.

9. If the results of step 8 do not meet the established efficiency requirements – modify the structure of nonlinear models (3) by adding polynomial and / or nonlinear components with control of the fulfillment of one of the conditions of steps.4 and 5. Repeat steps 1–8 of the methodology for determining the parameters of the nonlinear model.

10. If the results of step 9 are unsatisfactory, use the results of steps 4 and 5 as initial approximations for any numerical methods for solving nonlinear problems of time series approximation.

4 EXPERIMENTS

Evaluation of the effectiveness of the proposed approach was carried out by methods of mathematical modeling. The time series was generated according to model (9). Samples of different sizes were studied, which ensures the variability of the composition of seasonal changes (fluctuations) [9]. Simulation configuration:

– model: $f(t) = 2.2 \sin(0.01t) + \cos(0.03t)$;

– random noise parameters: normal law, $\sigma = 0.2$;

– training sample sizes: 500, 1000, 1500;

– the time series is generated from the training (observation area) and test (forecasting area) parts with a volume of $[0, n]$ and $[n, 2n]$ – respectively.

– nonlinear model for processing the form (9), polynomial structure was used in accordance with the results of the proposed technique.

5 RESULTS

The results of the research are presented in Fig. 5 and Table 2.

The graphs of Fig. 5 contain: training data (blue); ideal trend, relative to which the indicators given in Table 2 are calculated (orange); Trend restored using the proposed technique (green color).

Table 2 contains data on model quality indicators (15), (16) and standard deviations of the results of processing the input time series with random noise. Characteristics of the degree of the polynomial – record the result of the proposed technique.

The results of the studies demonstrate the obvious properties of statistical learning results – an increase in seasonality in the data gives an increase in dynamic and stochastic error. However, the discrepancy between the estimation results, compared to the ideal trend, is commensurate with the confidence interval of the forecasting error. This confirms the compliance of the proposed solutions with the basic provisions of the methods of processing time series.

In favor of the proposed solutions should be attributed the useful prognostic properties of nonlinear models, which were obtained entirely by the methods of analytical synthesis of the structure of nonlinear models in terms of parameters and determination of their parameters. This is ensured using positive properties of differential transformations in the proposed technique [4, 5].

6 DISCUSSION

In this paper the method of constructing nonlinear mathematical models in the DNT transformation scheme was further developed. The modification of the method consists in controlling the conditions for the formation of a certain system of equations in the DSB scheme to find the parameters of a nonlinear model with its analytical solutions. If the system is uncertain, the nonlinear model is supplemented with components that are linear in parameters. In the case of an overdetermined system, its solution is carried out using the least Squares. A defined system is solved by classical approaches. These processes

are implemented with the control of stochastic and dynamic accuracy of models in the areas of observation and extrapolation. If the results of statistical learning are unsatisfactory in terms of accuracy, the obtained parameters of the nonlinear model are used as initial approximations of numerical methods.

All these modifications are based on empirically proven interdependencies of quality indicators of polynomial models at the level of linear deviation and prob-

ability of approximation with the quality of the results of synthesis and application of nonlinear models in terms of parameters.

The proposed solutions are based on the analytical synthesis of the structure of nonlinear models and the determination of their parameters due to the positive properties of the method of differential transformations.

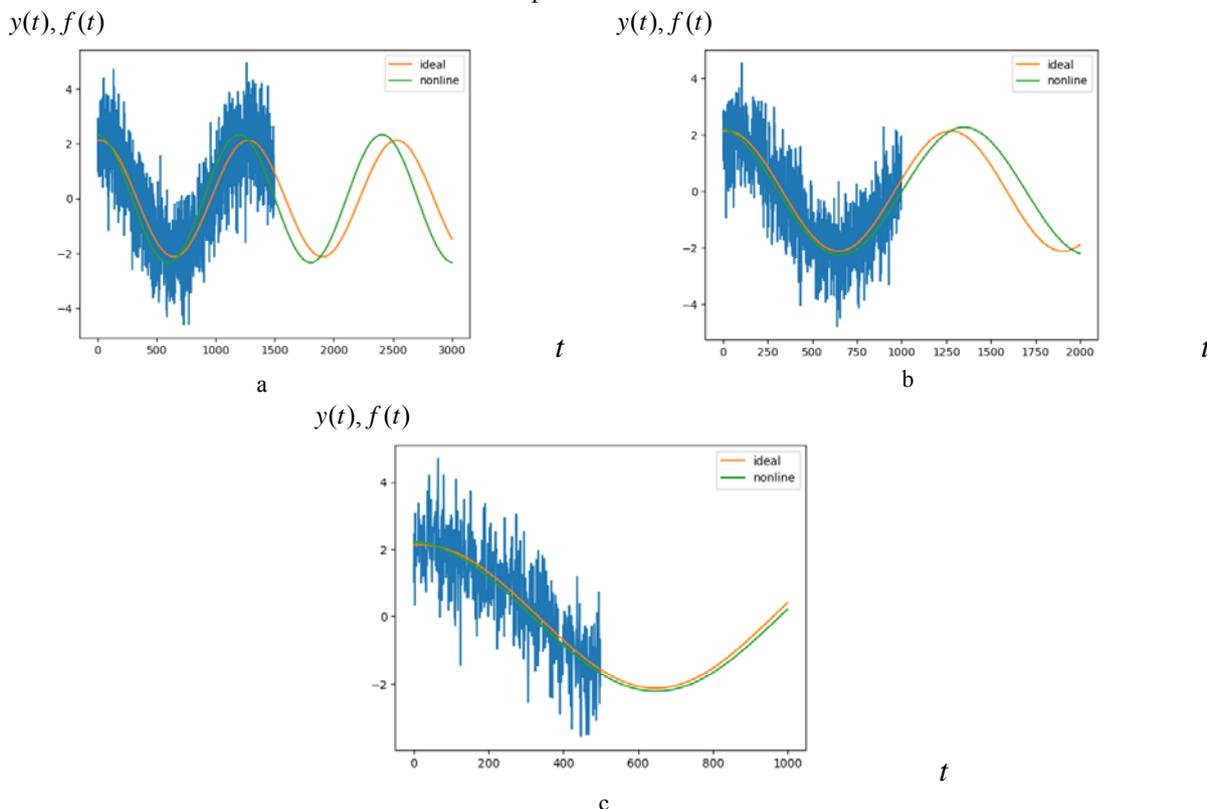


Figure 5 – Research results of the proposed methodology for samples: a – 1500, b – 1000, c – 500

Table 2 – Results of the model experiment

	Sample size	Polynomial degree	Δ	R^2	σ
train	500	4	0.12747	0.98981	1.26074
test	500	4	0.13555	0.9643	0.71604
train	1000	5	0.23514	0.97291	1.42798
test	1000	5	0.71622	0.73303	1.3848
train	1500	7	0.46893	0.91643	1.62107
test	1500	7	0.9894	0.62647	1.61779

Practical processing of the synthesized data proved the high predictive properties of the proposed approaches to the construction of nonlinear models.

The practice of applying the proposed approaches is possible in e-commerce software systems (trading, retailing, aggregation), computer systems of automated and automatic control (unmanned systems, traffic control systems) in the medical field and in other industries where there are large amounts of data in the time series format.

CONCLUSIONS

In this paper a conditional approach to analytical synthesis and determination of parameters of nonlinear models is proposed. The obtained solutions are an alternative and complement to the methods of numerical solution of nonlinear problems for processing time series and belong to the class of statistical machine learning methods.

The scientific novelty. The method of constructing nonlinear mathematical models in the DNT transforma-

tion scheme was further developed. The modification of the method consists in controlling the conditions for the formation of a certain system of equations in the DSB scheme to search for the parameters of a nonlinear model with its analytical solutions.

The practical significance of the proposed methodology is to obtain an analytical solution to the problem of determining the parameters of nonlinear models, which increases the predictive accuracy and productivity of statistical training methods on time series.

Prospects for further research. Extension of the proposed solutions to approximate methods into an alternative to DBS, creation of a program script in the format of a specialized library of nonlinear statistical learning.

ACKNOWLEDGEMENTS

The results of these studies were used in scientific research project “Nonlinear and multicriteria mathematical models for Data Science and Embedded Systems technology” (state registration number 0124U003323).

REFERENCES

1. James G., Witten D., Hastie T., Tibshirani R., Taylor J. An introduction to statistical learning: With applications in python. Springer Nature, 2023, 613 p.
2. Nguyen M. A. Guide on Data Analysis. Bookdown, 2020. https://bookdown.org/mike/data_analysis/
3. Chapra S. C., Canale R. P. Numerical Methods for Engineers, Seventh Edition. New York, McGraw-Hill Education, 2015, 987 p.
4. Hastie T., Tibshirani R., Friedman J. The elements of statistical learning: data mining, inference, and prediction, *Springer Series in Statistics*. Springer, USA, 2009, 745 p.
5. Provost F., Fawcett T. Data Science for Business. O’Reilly Media, Inc, USA, 2013, 409 p.
6. Guang-yong C., Gan M., Chen G. Generalized exponential autoregressive models for nonlinear time series: stationarity, estimation and applications, *Information Sciences*, 2018, Vol. 438, pp. 46–57. DOI: <https://doi.org/10.1016/j.ins.2018.01.029>
7. Tuhanskyh O., Baran D., Pysarchuk O.; Yang X. S., Sherratt S., Dey N., Joshi A. (eds) Method for Statistical Evaluation of Nonlinear Model Parameters in Statistical Learning Algorithms, *Proceedings of Ninth International Congress on Information and Communication Technology. ICICT 2024. Lecture Notes in Networks and Systems*. Springer, Singapore, 2024, Vol. 1013, pp. 265–274. DOI: https://doi.org/10.1007/978-981-97-3559-4_21
8. Dietrich D., Heller B., Yang B. Data Science & Big Data Analytics: Discovering, Analyzing, Visualizing and Presenting Data, John Wiley & Sons, Inc., Indianapolis, Indiana, 2015, 420 p.
9. Pronzato L., Pázman A. Design of experiments in nonlinear models, *Lecture notes in statistics*, Vol. 212.1. Springer New York, 2013, 399 p.
10. Konishi S. Introduction to multivariate analysis: linear and nonlinear modeling, *CRC Press*. Taylor & Francis Group, Tokyo, 2014, 312 p.
11. Grafarend E. W., Awange J. L., Zwanzig S. Applications of linear and nonlinear models, *Springer Geophysics, Fixed Effects*. Springer Nature, Switzerland, 2012, 1113 p.
12. Raymond J. C., Ruppert D., Stefanski L. A., Crainiceanu C. M. Measurement error in nonlinear models: a modern perspective. Chapman and Hall, CRC, 2006, 242 p.

Received 26.12.2024.
Accepted 09.02.2025.

УДК 519.2+004.8

МЕТОДИКА ВИЗНАЧЕННЯ СТРУКТУРИ НЕЛІНІЙНИХ ЗА ПАРАМЕТРАМИ МОДЕЛЕЙ ДЛЯ ОБРОБКИ ЧАСОВИХ РЯДІВ

Писарчук О. О. – д-р техн. наук, професор, професор кафедри обчислювальної техніки, факультету інформатики та обчислювальної техніки, КПІ ім. Ігоря Сікорського, Київ, Україна.

Туғанських О. А. – аспірант кафедри обчислювальної техніки, факультету інформатики та обчислювальної техніки КПІ ім. Ігоря Сікорського, Київ, Україна.

Баран Д. Р. – аспірант кафедри обчислювальної техніки, факультету інформатики та обчислювальної техніки КПІ ім. Ігоря Сікорського, Київ, Україна.

АНОТАЦІЯ

Актуальність. Практика задач сьогодення актуалізує підвищення вимог до точності, достовірності і повноти результатів обробки часових рядів в багатьох прикладних сферах. Одним із методів, що забезпечує високоточну обробку часових рядів із впровадженням стохастичної моделі вимірних параметрів є методи статистичного навчання. Однак, сучасні підходи до статистичного навчання обмежуються, здебільшого, спрощеними – лінійними за параметрами поліноміальними моделями. Практика доводить, що реальні дані найчастіше мають складну форму трендової складової, яка не може бути відтворена поліномами навіть високого ступеня. Згладжування нелінійних за параметрами моделей можливо реалізувати різними підходами, наприклад методом визначення параметрів нелінійних моделей з використанням балансу диференціальних спектрів (БДС) в схемі диференціально-нетейлорівських перетворень (ДНТ). Дослідження довели необхідність його модифікації в напрямку розробки обумовленого підходу до визначення структури нелінійних за параметрами математичних моделей для обробки часових рядів із складною динамікою тренду.

Метою роботи є розробка методики визначення структури нелінійних за математичних моделей для обробки часових рядів з використанням БДС в ДНТ перетвореннях.

Метод. В статті отримав розвиток метод побудови нелінійних за параметрами математичних моделей в схемі ДНТ перетворень. Модифікація методу полягає у контролі умов формування визначеної системи рівнянь в схемі БДС для пошуку параметрів нелінійної моделі з її аналітичним розв’язком. Якщо система невизначена – нелінійна модель доповнюється лінійними за параметрами компонентами. У випадку перевизначеної системи – її розв’язок здійснюється з використанням

норми найменших квадратів. Визначена система – розв’язується класичними підходами. Зазначені процеси реалізуються із контролем стохастичної та динамічної точності моделей на ділянках спостереження та екстраполяції. Якщо результати статистичного навчання є незадовільними за точністю – отримані значення нелінійної моделі використовуються як початкові наближення чисельних методів.

Результат. На підставі проведених досліджень запропоновано методику визначення структури нелінійних за параметрами моделей для обробки часових рядів з використанням БДС в схемі ДНТ перетворень. Її застосування забезпечує обумовлений підхід до визначення структури моделей для обробки часових рядів та підвищення точності оцінювання на інтервалі спостереження та екстраполяції.

Висновки. Застосування запропонованої в статті методики визначення структури нелінійних за параметрами моделей для обробки часових рядів дозволяє отримати моделі із кращими, за показником точності, прогностичними властивостями.

КЛЮЧОВІ СЛОВА: наука про дані, статистичне навчання, часові послідовності, нелінійні моделі, чисельні методи, метод найменших квадратів.

ЛІТЕРАТУРА

1. An introduction to statistical learning: With applications in python / [G. James, D. Witten, T. Hastie et al.]. – Springer Nature, 2023. – 613 p.
2. Nguyen M. A. Guide on Data Analysis / M. A. Nguyen. – Bookdown, 2020. https://bookdown.org/mike/data_analysis/
3. Chapra S. C. Numerical Methods for Engineers, Seventh Edition. / S. C. Chapra, R. P. Canale. – New York : McGraw-Hill Education, 2015. – 987 p.
4. Hastie T. The elements of statistical learning: data mining, inference, and prediction. / T. Hastie, R. Tibshirani, J. Friedman // Springer Series in Statistics. – Springer, USA, 2009 – 745 p.
5. Provost F. Data Science for Business / F. Provost, T. Fawcett. – O’Reilly Media, Inc, USA, 2013. – 409 p.
6. Guang-yong C. Generalized exponential autoregressive models for nonlinear time series: stationarity, estimation and applications. / C. Guang-yong, M. Gan, G. Chen // Information Sciences. – 2018. – Vol. 438. – P. 46–57. DOI: <https://doi.org/10.1016/j.ins.2018.01.029>
7. Method for Statistical Evaluation of Nonlinear Model Parameters in Statistical Learning Algorithms. / [O. Tuhanskyh, D. Baran, O. Pysarchuk], Yang XS., Sherratt S., Dey N., Joshi A. (eds) // Proceedings of Ninth International Congress on Information and Communication Technology. ICICT 2024. Lecture Notes in Networks and Systems. – Springer, Singapore. – 2024. – Vol. 1013 – P. 265–274. DOI: https://doi.org/10.1007/978-981-97-3559-4_21
8. Dietrich D. Data Science & Big Data Analytics: Discovering, Analyzing, Visualizing and Presenting Data / D. Dietrich, B. Heller, B. Yang. – John Wiley & Sons, Inc., Indianapolis, Indiana, 2015. – 420 p.
9. Pronzato L. Design of experiments in nonlinear models. L. Pronzato, A. Pázman // Lecture notes in statistics. – Vol. 212.1. – Springer New York, 2013. – 399 p.
10. Konishi S. Introduction to multivariate analysis: linear and nonlinear modeling / S. Konishi // CRC Press. – Taylor & Francis Group, Tokyo, 2014. – 312 p.
11. Grafarend. E.W. Applications of linear and nonlinear models / E. W. Grafarend, J. L. Awange. S. Zwanzig // Springer Geophysics, Fixed Effects. – Springer Nature, Switzerland, 2012. – 1113 p.
12. Measurement error in nonlinear models: a modern perspective / [C. J. Raymond, D. Ruppert, L. A. Stefanski, C. M. Crainiceanu]. – Chapman and Hall, CRC, 2006. – 242 p.

МАТЕМАТИЧНЕ МОДЕЛЮВАННЯ БОЙОВИХ ДІЙ З МОЖЛИВІСТЮ ПЕРЕРОЗПОДІЛУ БОЙОВИХ РЕСУРСІВ МІЖ ДІЛЯНКАМИ ЗІТКНЕННЯ ТА РОЗПОДІЛУ РЕЗЕРВІВ

Фурсенко О. К. – канд. техн. наук, доц., завідувач кафедри вищої математики, Харківський національний університет Повітряних Сил імені Івана Кожедуба, Харків, Україна.

Черновол Н. М. – старший викладач кафедри вищої математики, Харківський національний університет Повітряних Сил імені Івана Кожедуба, Харків, Україна.

АНОТАЦІЯ

Актуальність. Математичні і комп'ютерні моделі динаміки бойових дій є важливим інструментом для прогнозування їх результату. Відомі моделі ланчестерівського типу були імітаційними і не враховували кінцеву мету і перерозподіл ресурсів в ході бойових зіткнень. В даній роботі пропонується оптимізаційна модель динаміки бойових дій між сторонами А і В на двох ділянках зіткнення, в основу якої покладено метод динамічного програмування з максимізацією цільової функції як функції втрат супротивника. Розроблена математична і комп'ютерна модель типової для сучасної війни ситуації ведення бойових дій між сторонами А і В на двох ділянках зіткнення з метою нанесення противнику максимальних втрат бойових ресурсів. Досягається ця мета шляхом перерозподілу ресурсів між ділянками зіткнення і введення на ці ділянки відповідних резервів.

Мета роботи. Побудувати математичну і комп'ютерну моделі динаміки бойових дій між сторонами А і В на двох ділянках зіткнення, в яких метою сторони А є максимізація втрат сторони В шляхом використання трьох ресурсів (першим параметром є кількість бойових одиниць, яку має можливість сторона А розподілити по ділянках зіткнення в початковий момент часу; другим параметром є кількість бойових одиниць, яку має перекинути сторона А з однієї ділянки на іншу в деякий наступний момент часу; третій параметр – це кількість бойових одиниць, яку має розподілити сторона А, використовуючи резерв) і шляхом моделювання знайти оптимальні значення цих параметрів.

Метод. В основу математичної моделі покладено метод динамічного програмування з функцією цілі як функцією втрат супротивника, а параметрами є одиниці бойових ресурсів на різних ділянках зіткнення. Їх число змінюється шляхом перерозподілу між цими ділянками і введенням резервних бойових одиниць. Втрати противника визначаються за допомогою систем диференціальних рівнянь Ланчестера. Враховуючи складність цільової функції для пошуку її максимуму використовується мова програмування Python.

Результати. Побудована математична модель і реалізована у вигляді алгоритму комп'ютерна модель поставленої задачі, яка базується на поєднанні метода динамічного програмування з розв'язанням систем диференціальних рівнянь динаміки бою Ланчестера з певними початковими умовами на кожному з трьох етапів бою. За допомогою чисельного експерименту проаналізовано допустимість параметрів задачі (кількостей бойових одиниць сторони А, які відповідно розподіляються, перекидаються з ділянки на ділянку або із числа резерву на кожному етапі бою). В роботі запропоновано алгоритм у вигляді блок-схеми, який дозволяє для будь-яких початкових даних давати відповідь щодо оптимального розподілу ресурсів сторони А, в тому числі із числа резерву, на трьох етапах бою і підраховувати відповідні найбільші втрати противника в заданий момент часу або давати відповідь, що немає допустимих значень параметрів задачі, тобто задача при певних початкових даних не має розв'язання.

Висновки. Наукова новизна полягає в тому, що розроблено математичну і комп'ютерну моделі динаміки бою на двох ділянках зіткнення, в якій враховується перерозподіл бойових ресурсів і резерву з метою нанесення противнику максимальних втрат. Чисельне моделювання за допомогою розробленого алгоритму дало можливість проаналізувати допустимість параметрів перерозподілу і резерву. На основі розглянутих прикладів зроблено висновок, що якщо задача не має розв'язання при певних даних, то це означає, що потрібно зменшити час перерозподілу бойових одиниць на одному або декількох етапах бою, тобто скоротити тривалість бою на певному етапі, тим самим можна прогнозувати час перерозподілу бойових ресурсів і резерву.

КЛЮЧОВІ СЛОВА: функція цілі як функція втрат, диференціальні рівняння динаміки «поганоорганізованого» бою, ділянки зіткнення, перерозподіл бойових одиниць, розподіл резервних бойових одиниць, ефективна скорострільність, максимізація втрат, допустимість параметрів перерозподілу і резервів.

АБРЕВІАТУРИ

БО – бойові одиниці;

ДЗ – ділянка зіткнення;

ДР – диференціальні рівняння;

А – сторона А;

В – сторона В.

НОМЕНКЛАТУРА

t_0 – час початку бою;

t_i – час закінчення i -го етапу бою ($i = 1, 2, 3$);

K – кількість бойових одиниць А на початку бою;

L – кількість бойових одиниць В на початку бою;

M – резерв А;

K_1 – кількість БО А, що розподіляється в момент часу t_0 на ДЗ Π_1 ;

L_1 – кількість БО В, що розподіляється в момент часу t_0 на ДЗ Π_1 ;

α_1 – середня ефективна скорострільність БО А на першому етапі бою на ДЗ Π_1 і ДЗ Π_2 ;

α_2 – середня ефективна скорострільність БО В на першому етапі бою на ДЗ П₁;

α_3 – середня ефективна скорострільність БО В на першому етапі бою на ДЗ П₂;

K_2 – кількість неушкоджених БО А, що перекидається з ДЗ П₁ на ДЗ П₂ в момент часу t_1 ;

K_3 – кількість БО А із числа резерву, що розподіляється на ДЗ П₁ в момент часу t_2 ;

α_{11} – середня ефективна скорострільність БО А на другому етапі бою на ДЗ П₁;

α_{12} – середня ефективна скорострільність БО А на другому етапі бою на ДЗ П₂;

α_{13} – середня ефективна скорострільність БО А на третьому етапі бою на ДЗ П₁;

α_{14} – середня ефективна скорострільність БО А на третьому етапі бою на ДЗ П₂;

$x_1(t, K_1)$ – середня чисельність неушкоджених БО А в момент часу t на ДЗ П₁ на першому етапі бою;

$x_2(t, K_1)$ – середня чисельність неушкоджених БО В в момент часу t на ДЗ П₁ на першому етапі бою;

$y_1(t, K_1)$ – середня чисельність неушкоджених БО А в момент часу t на ДЗ П₂ на першому етапі бою;

$y_2(t, K_1)$ – середня чисельність неушкоджених БО В в момент часу t на ДЗ П₂ на першому етапі бою;

$\hat{x}_1(t, K_1, K_2)$ – середня чисельність неушкоджених БО А в момент часу t на ДЗ П₁ на другому етапі бою;

$\hat{x}_2(t, K_1, K_2)$ – середня чисельність неушкоджених БО В в момент часу t на ДЗ П₁ на другому етапі бою;

$\hat{y}_1(t, K_1, K_2)$ – середня чисельність неушкоджених БО А в момент часу t на ДЗ П₂ на другому етапі бою;

$\hat{y}_2(t, K_1, K_2)$ – середня чисельність неушкоджених БО В в момент часу t на ДЗ П₂ на другому етапі бою;

$\tilde{x}_1(t, K_1, K_2, K_3)$ – середня чисельність неушкоджених БО А в момент часу t на ДЗ П₁ на третьому етапі бою;

$\tilde{x}_2(t, K_1, K_2, K_3)$ – середня чисельність неушкоджених БО В в момент часу t на ДЗ П₁ на третьому етапі бою;

$\tilde{y}_1(t, K_1, K_2, K_3)$ – середня чисельність неушкоджених БО А в момент часу t на ДЗ П₂ на третьому етапі бою;

$\tilde{y}_2(t, K_1, K_2, K_3)$ – середня чисельність неушкоджених БО В в момент часу t на ДЗ П₂ на третьому етапі бою;

$w_3(K_1, K_2, K_3)$ – втрати В за третій етап бою;

$W_3(K_1, K_2)$ – найбільші втрати В за третій етап бою;

$w_2(K_1, K_2)$ – сумарні втрати В за другий і третій етапи бою;

$W_2(K_1)$ – найбільші втрати В за другий і третій етапи бою;

$w_1(K_1)$ – втрати В за весь бій;

W – найбільші втрати В за весь бій;

K_{1opt} – оптимальне значення параметру K_1 ;

K_{2opt} – оптимальне значення параметру K_2 ;

K_{3opt} – оптимальне значення параметру K_3 ;

b_1 – проміжна змінна;

c_1 – проміжна змінна;

b_2 – проміжна змінна;

c_2 – проміжна змінна;

b_3 – проміжна змінна;

c_3 – проміжна змінна.

ВСТУП

Одним з важливих факторів, які впливають на результат бойових дій, є обґрунтованість прийняття управлінських рішень. Зокрема, рішень командирів про оптимальний розподіл бойових ресурсів між різними ділянками зіткнення з метою нанесення супротивнику максимальних втрат. Математичне моделювання динаміки бойових дій дає можливість загалом прогнозувати їх результат і тим самим може бути одним з інструментів в руках командирів на різних рівнях. Розробка математичних моделей динаміки бойових дій на кількох ділянках зіткнення, в яких враховується їх мета і перерозподіл ресурсів для досягнення її, є актуальною на сьогоднішній день проблемою.

Об'єкт дослідження: динаміки бою двох сторін.

Предмет дослідження: математичні і комп'ютерні моделі динаміки бойових дій на двох ділянках зіткнення з можливістю перерозподілу бойових ресурсів і введенні резерву.

Мета роботи: розробити математичну і комп'ютерну моделі динаміки бою на двох ділянках зіткнення, яка передбачає перерозподіл бойових оди-

ниць на певному етапі і введенні в бій резерву з метою максимізації втрат противника.

1 ПОСТАНОВКА ЗАДАЧІ

Відомі імітаційні моделі ланчестерівського типу, в яких не враховується мета бойових дій і перерозподіл ресурсів для її досягнення. В розроблених авторами оптимізаційних моделях [1, 2] не враховується можливість введення на ділянки зіткнення резервів. Своєчасне введення резервів є важливим фактором, що часто визначає результат бою. Тому важливо розробити математичну і комп'ютерну модель, яка б враховувала цю можливість. Опишемо спочатку процес бойових дій, який підлягає моделюванню.

1. Розглядається бій між двома угрупованнями. Задано K, L, M . Сторони воюють на двох ДЗ Π_1 і Π_2 . Вони мають однорідні БО, не обов'язково однорідні між собою.

2. В момент часу $t_0 = 0$ розподіляється K_1 (відома α_1) і L_1 (відома α_2). На другу ділянку розподіляється, відповідно, $K - K_1$ БО А (зберігається α_1) і $L - L_1$ БО В (відома α_3). Ефективною скорострільністю бойової одиниці називають кількість її успішних (влучних) пострілів за одиницю часу.

3. В момент часу $t_1 > t_0$ розподіляється K_2 . Після перерозподілу БО А з першої ДЗ на другу відбувається зміна середньої ефективної скорострільності БО А, середня ефективна скорострільність БО В залишається незмінною.

4. В момент часу $t_2 > t_1$ А розподіляє свій резерв по ДЗ: на першу ДЗ – K_3 БО, а на другу ДЗ $M - K_3$ БО. Відбувається зміна середньої ефективної скорострільності БО А на обох ДЗ, середня ефективна скорострільність БО В залишається незмінною.

5. Бій триває до моменту часу $t_3 > t_2$.

6. Для моделювання динаміки бою використовуються системи ДР Ланчестера. Вважається, що між моментами часу t_0 і t_1 (1-ий етап бою), t_1 і t_2 (2-ий етап бою), t_2 і t_3 (3-ій етап бою) бій відбувається за умов «поганоорганізованого бою».

Потрібно знайти оптимальні значення параметрів K_1, K_2 і K_3 , тобто такі, щоб В на момент часу t_3 мала найбільші втрати, і знайти ці втрати.

2 ОГЛЯД ЛІТЕРАТУРИ

В основу математичної моделі покладено метод динамічного програмування в комбінації з розв'язанням задач Коші для систем ДР Ланчестера, за допомогою розв'язків яких підраховуються кількості неушкоджених БО обох сторін в будь-який допустимий момент часу і виявленням допустимості пара-

метрів задач. Для комп'ютерного моделювання побудовано алгоритм та створено програму на мові програмування Python [3].

Метод динамічного програмування широко застосовувався перш за все в задачах розподілу ресурсів [4–6], але не бойових. Стосовно бойових ресурсів він вперше був запропонований авторами для розв'язання більш простої задачі [1, 2]. Моделі Ланчестера, які виникли на початку 20-го сторіччя [7], досліджуються і по цей час, наприклад, в роботах [8–17]. В цих роботах проводяться свої дослідження, але не з точки зору максимізації втрат супротивника в залежності від розподілу бойових ресурсів, в тому числі і резерву, в певні моменти часу, що пропонують автори даної роботи. Зауважимо, що в даній роботі між моментами перерозподілу ресурсів і резерву розглянуто одну з найпростіших моделей бойових дій (модель Ланчестера «поганоорганізованого бою» [18]). Бій називається «поганоорганізованим боєм» завдяки умові: інформація про враження цілі доходить до кожної сторони не миттєво, тобто обстрілюються і вже вражені БО.

3 МАТЕРІАЛИ І МЕТОДИ

Оскільки на трьох етапах бою відбувається перерозподіл бойових ресурсів, то для моделювання застосовуємо метод динамічного програмування. На кожному етапі (починаючи з останнього згідно методу динамічного програмування) потрібно виписати функцію цілі, що є сумою втрат В на двох ДЗ в моменті часу t_1, t_2, t_3 відповідно і максимізувати функцію цілі по відповідному параметру. При цьому кількість неушкоджених БО обох сторін як функції часу знаходяться за допомогою відповідних розв'язків систем ДР Ланчестера з заданими початковими умовами, що залежать від параметрів K_1, K_2 і K_3 .

Розглянемо детальніше постановку і розв'язання задачі. Спочатку розглянемо 1-ий етап бою.

Згідно умов «поганоорганізованого бою» [18] система ДР Ланчестера відносно середньої чисельності неушкоджених БО угруповань, враховуючи початкові умови, має вигляд:

$$\begin{array}{l} \text{на ДЗ } \Pi_1 : \\ \left\{ \begin{array}{l} x_1' = -\frac{\alpha_2}{K_1} x_1 x_2; \\ x_2' = -\frac{\alpha_1}{L_1} x_1 x_2; \end{array} \right. \\ x_1(0) = K_1; \\ x_2(0) = L_1; \end{array} \quad \begin{array}{l} \text{на ДЗ } \Pi_2 : \\ \left\{ \begin{array}{l} y_1' = -\frac{\alpha_3}{K - K_1} y_1 y_2; \\ y_2' = -\frac{\alpha_1}{L - L_1} y_1 y_2; \end{array} \right. \\ y_1(0) = K - K_1; \\ y_2(0) = L - L_1. \end{array}$$

Тут параметр K_1 задовольняє нерівність

$$2 \leq K_1 \leq K - 1,$$

тобто на першу ділянку має бути розподілено принаймні дві БО А (щоб принаймні одну БО можна було перекинути на другу ДЗ і принаймні одна одиниця лишилась на першій ДЗ) і на другу ДЗ має бути розподілена принаймні одна БО.

Позначимо

$$b_1 = \frac{L_1}{K_1} \sqrt{\frac{\alpha_2}{\alpha_1}}, \quad c_1 = \frac{L - L_1}{K - K_1} \sqrt{\frac{\alpha_3}{\alpha_1}}.$$

В останніх формулах $b_1 = b_1(K_1)$, $c_1 = c_1(K_1)$, але залежність від параметру K_1 нижче будемо опускати, щоб не ускладнювати формули.

Розв'язавши вказані задачі Коші, отримаємо, що в момент часу t_1 залишилась неушкодженими наступна кількість БО А, В:

на ДЗ Π_1 :

$$x_1(t_1, K_1) = \begin{cases} K_1 \frac{b_1^2 - 1}{b_1^2 e^{\sqrt{\alpha_1 \alpha_2} (b_1 - b_1^{-1}) t_1} - 1}, & b_1 \neq 1 \\ K_1 \frac{1}{\sqrt{\alpha_1 \alpha_2} t_1 + 1}, & b_1 = 1 \end{cases}, \quad (1)$$

$$x_2(t_1, K_1) = \begin{cases} L_1 \frac{1 - b_1^2}{e^{\sqrt{\alpha_1 \alpha_2} (b_1^{-1} - b_1) t_1} - b_1^2}, & b_1 \neq 1 \\ L_1 \frac{1}{\sqrt{\alpha_1 \alpha_2} t_1 + 1}, & b_1 = 1 \end{cases}, \quad (2)$$

на ДЗ Π_2 :

$$y_1(t_1, K_1) = (K - K_1) \times \begin{cases} \frac{c_1^2 - 1}{c_1^2 e^{\sqrt{\alpha_1 \alpha_3} (c_1 - c_1^{-1}) t_1} - 1}, & c_1 \neq 1 \\ \frac{1}{\sqrt{\alpha_1 \alpha_3} t_1 + 1}, & c_1 = 1 \end{cases}, \quad (3)$$

$$y_2(t_1, K_1) = (L - L_1) \times \begin{cases} \frac{1 - c_1^2}{e^{\sqrt{\alpha_1 \alpha_3} (c_1^{-1} - c_1) t_1} - c_1^2}, & c_1 \neq 1 \\ \frac{1}{\sqrt{\alpha_1 \alpha_3} t_1 + 1}, & c_1 = 1 \end{cases} \quad (4)$$

Але не всі натуральні $K_1 \in [2; K - 1]$ будуть допустимими, тобто

$$K_1 \in [K_{11}; K_{12}] \subset [2; K - 1],$$

виходячи з умов

$$x_1(t_1, K_1) \geq 2 \quad \text{і} \quad y_1(t_1, K_1) \geq 1,$$

які потрібні як необхідна умова, щоб продовжувати бій далі для А на обох ділянках, і умов

$$x_2(t_1, K_1) \geq 0,1 \quad \text{і} \quad y_2(t_1, K_1) \geq 0,1$$

(оскільки, якщо у противника – В у момент часу t_1 БО менше вже 0,1 (наприклад), то потрібно вже далі не вести бій з В на цьому етапі). Зауважимо, що розв'язки розглянутих в роботі систем диференціальних рівнянь є спадними функціями по змінній t і при $t \rightarrow +\infty$ можуть прямувати до нуля. Тому вести бій до тих пір, поки значення цих функцій, що виражають кількість неушкоджених БО В, буде менше ніж, наприклад, 10^{-10} не оптимально і недоцільно. Звідки розглядаються для супротивника умови атаки з боку А, а саме

$$x_2(t_1, K_1) \geq 0,1 \quad \text{і} \quad y_2(t_1, K_1) \geq 0,1.$$

Далі в момент часу t_1 з першої ДЗ на другу перекидається $K_2 : K_2 \in [0; x_1(t_1, K_1) - 1]$ неушкоджених БО А (щоб принаймні одна БО залишилась на першій ДЗ в момент часу t_1). Середня ефективна скорострільність БО А на ДЗ Π_1 зменшиться (це може відбутися, наприклад, за рахунок відведення БО на передові позиції), а на ДЗ Π_2 зросте (за рахунок, наприклад, виведення БО на більш передові позиції). При чому, зміна середньої ефективної скорострільності БО буде прямо пропорційна кількості перекинутих одиниць K_2 і обернено пропорційна кількості одиниць $x_1(t_1, K_1)$ А, що залишились неушкодженими на ДЗ Π_1 . А саме, нехай

$$\alpha_{11} = \alpha_1 \left(1 - \frac{K_2}{x_1(t_1, K_1)} \right); \quad \alpha_{12} = \alpha_1 \left(1 + \frac{K_2}{x_1(t_1, K_1)} \right). \quad (5)$$

В формулах (5) $\alpha_{11} = \alpha_{11}(K_1, K_2)$, $\alpha_{12} = \alpha_{12}(K_1, K_2)$, але залежність від параметрів K_1, K_2 нижче будемо опускати, щоб не ускладнювати формули.

Далі розглянемо бій на другому етапі. Система ДР Ланчестера відносно середньої чисельності неушкоджених БО угруповань, враховуючи початкові умови, має вигляд:

на ДЗ П₁ :

$$\begin{cases} \hat{x}_1' = -\frac{\alpha_2}{x_1(t_1, K_1) - K_2} \hat{x}_1 \hat{x}_2; \\ \hat{x}_2' = -\frac{\alpha_{11}}{x_2(t_1, K_1)} \hat{x}_1 \hat{x}_2; \\ \hat{x}_1(t_1) = x_1(t_1, K_1) - K_2; \\ \hat{x}_2(t_1) = x_2(t_1, K_1); \end{cases}$$

на ДЗ П₂ :

$$\begin{cases} \hat{y}_1' = -\frac{\alpha_3}{y_1(t_1, K_1) + K_2} \hat{y}_1 \hat{y}_2; \\ \hat{y}_2' = -\frac{\alpha_{12}}{y_2(t_1, K_1)} \hat{y}_1 \hat{y}_2; \\ \hat{y}_1(t_1) = y_1(t_1, K_1) + K_2; \\ \hat{y}_2(t_1) = y_2(t_1, K_1). \end{cases}$$

Позначимо

$$b_2 = \frac{x_2(t_1, K_1)}{x_1(t_1, K_1) - K_2} \sqrt{\frac{\alpha_2}{\alpha_{11}}}, \quad c_2 = \frac{y_2(t_1, K_1)}{y_1(t_1, K_1) + K_2} \sqrt{\frac{\alpha_3}{\alpha_{12}}}.$$

В останніх формулах $b_2 = b_2(K_1, K_2)$, $c_2 = c_2(K_1, K_2)$.

Розв'язавши вказані задачі Коші, отримаємо, що в момент часу t_2 залишилась неушкодженими наступна кількість БО А і В:

на ДЗ П₁ :

$$\hat{x}_1(t_2, K_1, K_2) = (x_1(t_1, K_1) - K_2) \times \begin{cases} \frac{b_2^2 - 1}{b_2^2 e^{\sqrt{\alpha_{11}\alpha_2}(b_2 - b_2^{-1})(t_2 - t_1)} - 1}, & b_2 \neq 1 \\ \frac{1}{\sqrt{\alpha_{11}\alpha_2}(t_2 - t_1) + 1}, & b_2 = 1 \end{cases}, \quad (6)$$

$$\hat{x}_2(t_2, K_1, K_2) = x_2(t_1, K_1) \times \begin{cases} \frac{1 - b_2^2}{e^{\sqrt{\alpha_{11}\alpha_2}(b_2^{-1} - b_2)(t_2 - t_1)} - b_2^2}, & b_2 \neq 1 \\ \frac{1}{\sqrt{\alpha_{11}\alpha_2}(t_2 - t_1) + 1}, & b_2 = 1 \end{cases}, \quad (7)$$

на ДЗ П₂ :

$$\hat{y}_1(t_2, K_1, K_2) = (y_1(t_1, K_1) + K_2) \times$$

$$\times \begin{cases} \frac{c_2^2 - 1}{c_2^2 e^{\sqrt{\alpha_{12}\alpha_3}(c_2 - c_2^{-1})(t_2 - t_1)} - 1}, & c_2 \neq 1 \\ \frac{1}{\sqrt{\alpha_{12}\alpha_3}(t_2 - t_1) + 1}, & c_2 = 1 \end{cases}, \quad (8)$$

$$\hat{y}_2(t_2, K_1, K_2) = y_2(t_1, K_1) \times \begin{cases} \frac{1 - c_2^2}{e^{\sqrt{\alpha_{12}\alpha_3}(c_2^{-1} - c_2)(t_2 - t_1)} - c_2^2}, & c_2 \neq 1 \\ \frac{1}{\sqrt{\alpha_{12}\alpha_3}(t_2 - t_1) + 1}, & c_2 = 1 \end{cases}. \quad (9)$$

Не всі невід'ємні цілі $K_2 \in [0; x_1(t_1, K_1) - 1]$ будуть допустимими, тобто

$$K_2 \in [K_{21}; K_{22}] \subset [0; x_1(t_1, K_1) - 1],$$

виходячи з умов:

$$\begin{aligned} \hat{x}_1(t_2, K_1, K_2) &\geq 1, \quad \hat{y}_1(t_2, K_1, K_2) \geq 1, \\ \hat{x}_2(t_2, K_1, K_2) &\geq 0, 1, \quad \hat{y}_2(t_2, K_1, K_2) \geq 0, 1. \end{aligned}$$

Пояснення щодо необхідності чотирьох останніх нерівностей такі самі, як і при поясненні аналогічних нерівностей при з'ясуванні допустимості параметру K_1 .

Тепер розглянемо бій на 3-му етапі. Для досягнення успіху А в момент часу t_2 розподіляє свої резерви по ДЗ: на першу ДЗ – K_3 бойових одиниць, тоді на другу ДЗ – $M - K_3$ БО. Вважаємо, що середня ефективна скорострільність БО А на обох ділянках зростає (зростає на величину, прямо пропорційну кількості розподілених на відповідну ДЗ одиниць із числа резерву і обернено пропорційну кількості всього резерву з урахуванням тієї кількості БО А, що залишились неушкодженими на відповідній ДЗ). Нехай,

$$\begin{aligned} \alpha_{13} &= \alpha_{11} \left(1 + \frac{K_3}{M + \hat{x}_1(t_2, K_1, K_2)} \right), \\ \alpha_{14} &= \alpha_{12} \left(1 + \frac{M - K_3}{M + \hat{y}_1(t_2, K_1, K_2)} \right). \end{aligned} \quad (10)$$

В формулах (10) $\alpha_{13} = \alpha_{13}(K_1, K_2, K_3)$, $\alpha_{14} = \alpha_{14}(K_1, K_2, K_3)$.

Система ДР Ланчестера відносно середнього числа неушкоджених БО угруповань, враховуючи початкові умови, має вигляд:

на ДЗ Π_1 :

$$\begin{cases} \tilde{x}_1' = -\frac{\alpha_2}{\hat{x}_1(t_2, K_1, K_2) + K_3} \tilde{x}_1 \tilde{x}_2; \\ \tilde{x}_2' = -\frac{\alpha_{13}}{\hat{x}_2(t_2, K_1, K_2)} \tilde{x}_1 \tilde{x}_2; \\ \tilde{x}_1(t_2) = \hat{x}_1(t_2, K_1, K_2) + K_3; \\ \tilde{x}_2(t_2) = \hat{x}_2(t_2, K_1, K_2); \end{cases}$$

на ДЗ Π_2 :

$$\begin{cases} \tilde{y}_1' = -\frac{\alpha_3}{\hat{y}_1(t_2, K_1, K_2) + M - K_3} \tilde{y}_1 \tilde{y}_2; \\ \tilde{y}_2' = -\frac{\alpha_{14}}{\hat{y}_2(t_2, K_1, K_2)} \tilde{y}_1 \tilde{y}_2; \\ \tilde{y}_1(t_2) = \hat{y}_1(t_2, K_1, K_2) + M - K_3; \\ \tilde{y}_2(t_2) = \hat{y}_2(t_2, K_1, K_2). \end{cases}$$

Позначимо

$$b_3 = \frac{\hat{x}_2(t_2, K_1, K_2)}{\hat{x}_1(t_2, K_1, K_2) + K_3} \sqrt{\frac{\alpha_2}{\alpha_{13}}},$$

$$c_3 = \frac{\hat{y}_2(t_2, K_1, K_2)}{\hat{y}_1(t_2, K_1, K_2) + M - K_3} \sqrt{\frac{\alpha_3}{\alpha_{14}}}.$$

В останніх формулах $b_3 = b_3(K_1, K_2, K_3)$,
 $c_3 = c_3(K_1, K_2, K_3)$.

Розв'язавши вказані задачі Коші, отримаємо, що в момент часу t_3 залишилась неушкодженими наступна кількість одиниць А,В:

на ДЗ Π_1 :

$$\tilde{x}_1(t_3, K_1, K_2, K_3) = (\hat{x}_1(t_2, K_1, K_2) + K_3) \times$$

$$\begin{cases} \frac{b_3^2 - 1}{b_3^2 e^{\sqrt{\alpha_{13}\alpha_2}(b_3 - b_3^{-1})(t_3 - t_2)} - 1}, & b_3 \neq 1 \\ \frac{1}{\sqrt{\alpha_{13}\alpha_2}(t_3 - t_2) + 1}, & b_3 = 1 \end{cases}, \quad (11)$$

$$\tilde{x}_2(t_3, K_1, K_2, K_3) = \hat{x}_2(t_2, K_1, K_2) \times$$

$$\begin{cases} \frac{1 - b_3^2}{e^{\sqrt{\alpha_{13}\alpha_2}(b_3^{-1} - b_3)(t_3 - t_2)} - b_3^2}, & b_3 \neq 1 \\ \frac{1}{\sqrt{\alpha_{13}\alpha_2}(t_3 - t_2) + 1}, & b_3 = 1 \end{cases}, \quad (12)$$

на ДЗ Π_2 :

$$\tilde{y}_1(t_3, K_1, K_2, K_3) = (\hat{y}_1(t_2, K_1, K_2) + M - K_3) \times$$

$$\begin{cases} \frac{c_3^2 - 1}{c_3^2 e^{\sqrt{\alpha_{14}\alpha_3}(c_3 - c_3^{-1})(t_3 - t_2)} - 1}, & c_3 \neq 1 \\ \frac{1}{\sqrt{\alpha_{14}\alpha_3}(t_3 - t_2) + 1}, & c_3 = 1 \end{cases}, \quad (13)$$

$$\tilde{y}_2(t_3, K_1, K_2, K_3) = \hat{y}_2(t_2, K_1, K_2) \times$$

$$\begin{cases} \frac{1 - c_3^2}{e^{\sqrt{\alpha_{14}\alpha_3}(c_3^{-1} - c_3)(t_3 - t_2)} - c_3^2}, & c_3 \neq 1 \\ \frac{1}{\sqrt{\alpha_{14}\alpha_3}(t_3 - t_2) + 1}, & c_3 = 1 \end{cases}. \quad (14)$$

Параметр K_3 : $K_3 \in [K_{31}; K_{32}] \subset [0; M]$, виходячи з умов

$$\tilde{x}_1(t_3, K_1, K_2, K_3) \geq 1, \quad \tilde{y}_1(t_3, K_1, K_2, K_3) \geq 1,$$

$$\tilde{x}_2(t_3, K_1, K_2, K_3) \geq 0, 1, \quad \tilde{y}_2(t_3, K_1, K_2, K_3) \geq 0, 1.$$

Для розв'язання задачі потрібно знайти такі значення параметрів K_1 , K_2 і K_3 , щоб втрати В до моменту часу t_3 були найбільшими. Почнемо з третього етапу бою. Маємо

$$w_3(K_1, K_2, K_3) = \hat{x}_2(t_2, K_1, K_2) - \tilde{x}_2(t_3, K_1, K_2, K_3) + \hat{y}_2(t_2, K_1, K_2) - \tilde{y}_2(t_3, K_1, K_2, K_3)$$

Відповідно, цільова функція

$$W_3(K_1, K_2) = \max_{K_3 \in [K_{31}; K_{32}] \subset [0; M]} w_3(K_1, K_2, K_3). \quad (15)$$

Втрати В на другому етапі бою становлять величину $x_2(t_1, K_1) - \hat{x}_2(t_2, K_1, K_2) + y_2(t_1, K_1) - \hat{y}_2(t_2, K_1, K_2)$. Тоді враховуючи, що на третьому етапі бою вже знайдено максимальні втрати $W_3(K_1, K_2)$ В, то сумарні втрати за другий і третій етапи становлять величину

$$w_2(K_1, K_2) = x_2(t_1, K_1) - \hat{x}_2(t_2, K_1, K_2) + y_2(t_1, K_1) - \hat{y}_2(t_2, K_1, K_2) + W_3(K_1, K_2).$$

Відповідно, цільова функція має наступний вигляд

$$W_2(K_1) = \max_{K_2 \in [K_{21}; K_{22}] \subset [0; x_1(t_1, K_1) - 1]} w_2(K_1, K_2). \quad (16)$$

Втрати В на першому етапі бою дорівнюють

$$L - x_2(t_1, K_1) - y_2(t_1, K_1).$$

Додаючи максимальні втрати за другий і третій етап бою, отримаємо величину

$$w_1(K_1) = L - x_2(t_1, K_1) - y_2(t_1, K_1) + W_2(K_1).$$

Отже, кінцева цільова функція по підрахунку сумарних максимальних втрат сторони 2 за весь бій має вигляд:

$$W_1 = \max_{K_1 \in [K_{11}; K_{12}] \subset [2; K-1]} w_1(K_1). \quad (17)$$

Зауважимо, що при $t \geq 0$ всі розв'язки задач Коші для систем ДР, що розглядаються в роботі, є додатними. Тому вирази в (1)–(4), (6)–(9), (11)–(14) є додатними.

4 ЕКСПЕРИМЕНТИ

Комп'ютерна модель поставленої задачі полягає в наступному алгоритмі, який наведемо у вигляді блок-схеми на рис. 1–3. Даний алгоритм є новим з точки зору демонстрації реалізації методу динамічного програмування в комбінації з підрахованими кількостями неушкоджених бойових одиниць кожної сторони в кінці трьох етапів бою за допомогою систем ДР Ланчестера з певними початковими умовами. Крім того, вказаний алгоритм відображає перевірку допустимості параметрів задачі на кожному етапі бою. Більше

того, запропонований алгоритм в разі відсутності допустимих параметрів задачі дає відповідь: 1) або для певного допустимого першого параметру немає допустимого або другого або третього параметрів; 2) або для певних допустимих перших двох параметрів задачі немає третього допустимого параметру; 3) або, взагалі, немає жодної трійки допустимих параметрів, тобто задача розв'язання не має.

На рис. 1 наведено частину блок-схеми, пов'язаної з першим параметром K_1 . На початку вводяться дані задачі. Особливістю алгоритму є «індикатор» $K_{1opt} = -1$, який контролює: чи будуть знайдено далі трійки допустимих параметрів K_1, K_2, K_3 , якщо ні, то виводиться, що «задача розв'язання не має», в протилежному разі виводяться K_{1opt} та W_1 . Крім цього, в цій частині блок-схеми в циклі по K_1 знаходяться кількості неушкоджених бойових одиниць обох сторін в кінці першого етапу бою $x_1(t_1, K_1), x_2(t_1, K_1), y_1(t_1, K_1), y_2(t_1, K_1)$ і знаходяться допустимі значення параметру K_1 . Якщо K_1 є допустимим, то переходимо до наступної частини блок-схеми рис. 2.

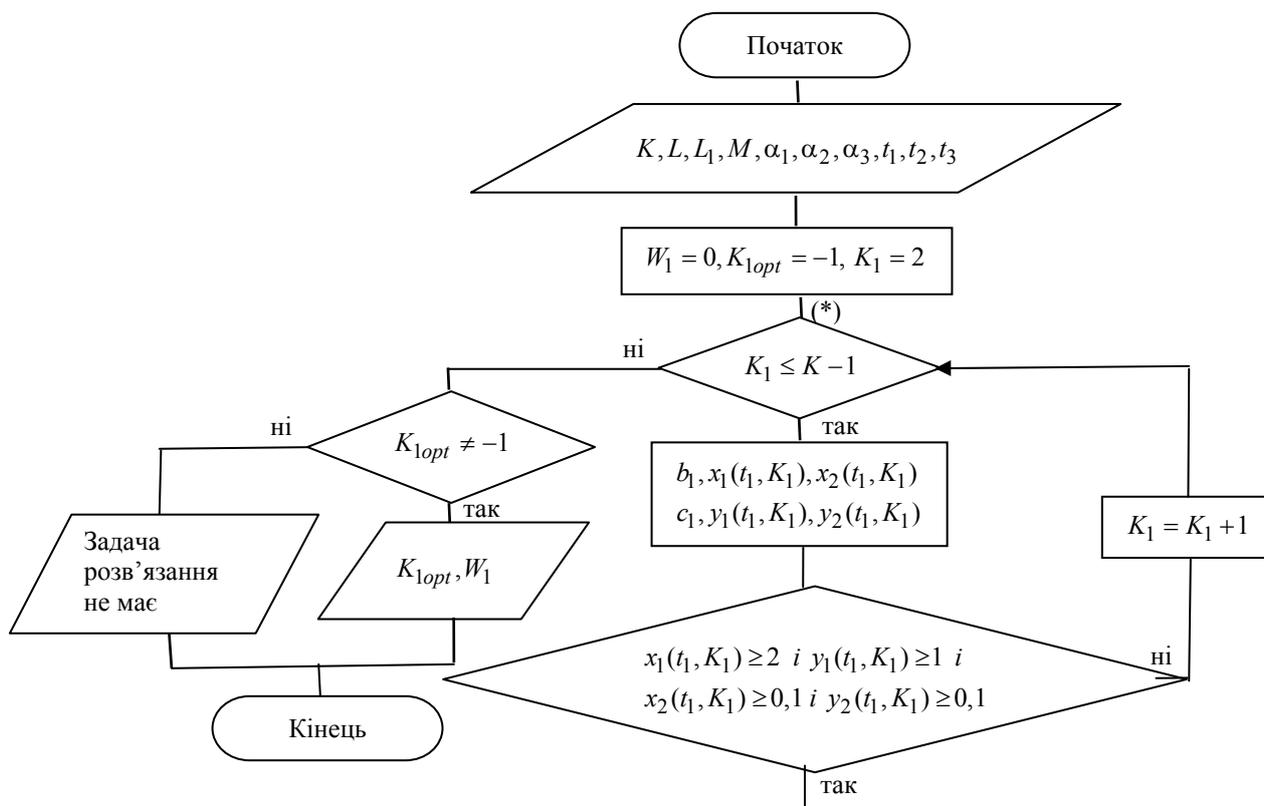


Рисунок 1 – Перша частина блок-схеми алгоритму реалізації методу динамічного програмування

На рис. 2 наведено частину блок-схеми, пов'язаної з дослідженням по другому параметру K_2 при фіксованому K_1 . Особливістю цієї частини алгоритму є «індикатор» $K_{2opt} = -1$, який контролює: чи буде знайдено при фіксованому допустимому K_1 допустимий параметр K_2 і K_3 , якщо ні, то виводиться, що при цьому K_1 «допустимих значень параметру K_2 або K_3 не має», в противному разі виводяться K_{2opt} та $W_2(K_1)$. Далі в частині алгоритму (рис. 2) підраховуються втрати B за другий і третій етапи бою $w_1(K_1)$ і найбільші втрати B за два етапи бою W_1 . Крім цього, в цій частині блок-схеми в циклі по K_2 знаходяться кількості неушкоджених бойових одиниць обох сторін в кінці другого етапу бою $x_1(t_2, K_1, K_2)$, $x_2(t_2, K_1, K_2)$, $y_1(t_2, K_1, K_2)$, $y_2(t_2, K_1, K_2)$ і знаходяться допустимі значення па-

раметру K_2 . Якщо K_2 є допустимим, то переходимо до наступної частини блок-схеми рис. 3.

Нарешті, наведемо закінчення блок-схеми (рис. 3), тобто дослідження по третьому параметру K_3 при фіксованих значеннях параметрів K_1 , K_2 . Особливістю третьої частини алгоритму є «індикатор» $K_{3opt} = -1$, який контролює: чи буде знайдено при фіксованих допустимих K_1 , K_2 допустимий параметр K_3 , якщо ні, то виводиться, що при цих K_1 , K_2 «допустимих значень параметру K_3 не має», в противному разі виводяться K_{3opt} та $W_3(K_1, K_2)$. Далі в частині алгоритму (рис. 3) підраховуються втрати B за другий і третій етапи бою $w_2(K_1, K_2)$ і найбільші втрати B за два етапи бою $W_2(K_1)$. Крім цього, в цій частині блок-схеми в циклі по K_3 знаходяться кількості неушкоджених бойових одиниць обох сторін в

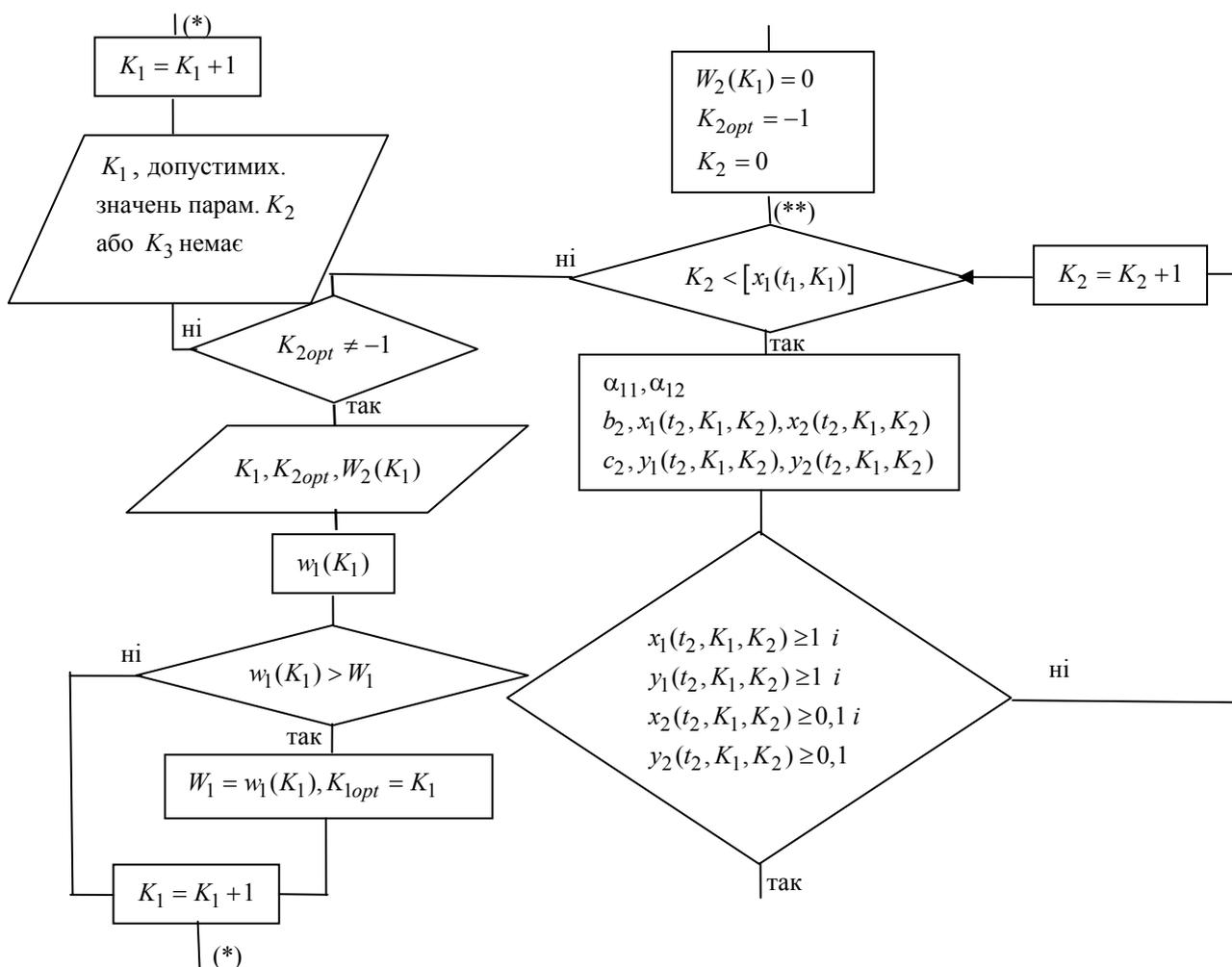


Рисунок 2 – Продовження блок-схеми алгоритму реалізації методу динамічного програмування

кінці третього етапу бою $x_1(t_3, K_1, K_2, K_3)$, $x_2(t_3, K_1, K_2, K_3)$, $y_1(t_3, K_1, K_2, K_3)$, $y_2(t_3, K_1, K_2, K_3)$ і знаходяться допустимі значення параметру K_3 ; виводиться допустима трійка параме-

трів K_1, K_2, K_3 і кількості неушкоджених бойових одиниць обох сторін за три етапи бою; підраховуються втрати $w_3(K_1, K_2, K_3)$ і найбільші втрати W за третій етап бою $W_3(K_1, K_2)$.

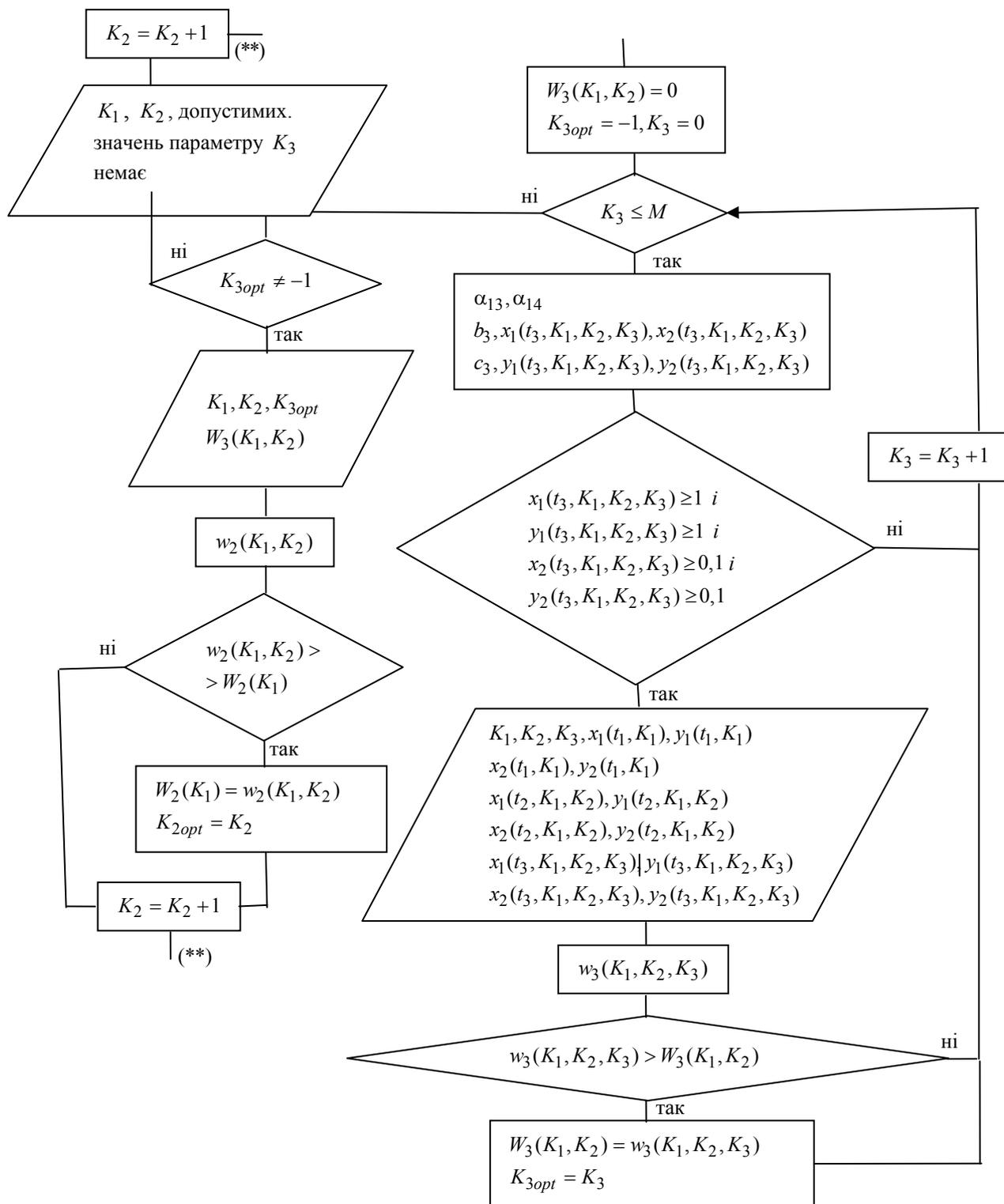


Рисунок 3 – Закінчення блок-схеми алгоритму реалізації методу динамічного програмування

Наведемо результати чисельних експериментів для різних значень початкових даних у вигляді табл. 1. Найбільші втрати B в цій таблиці округлено до цілого числа.

Обчислення зроблено для довільних даних, але їх можна зробити і використовуючи, при можливості публікації у відкритих засобах інформації, реальні характеристики певних видів зброї.

Таблиця 1 – Чисельні обчислення програми для різних значень початкових даних

№	K	L	L_1	M	α_1	α_2	α_3	t_1	t_2	t_3	K_{1opt}	K_{2opt}	K_{3opt}	W_1
1.	20	25	15	10	2,5	2	2	0,1	0,2	0,3	9	0	10	13
2.	25	20	8	10	2,5	2	2	0,1	0,2	0,3	22	13	0	16
3.	15	30	10	5	2,5	2	2	0,3	0,4	0,6	8	2	2	10
4.	40	20	12	8	2	1,8	1,6	0,1	0,2	0,3	22	0	8	18
5.	40	30	15	10	1,8	2,5	2,2	0,2	0,4	0,5	35	19	0	23
6.	60	50	20	15	3	2,8	3,5	0,1	0,3	0,4	54	30	0	43
7.	50	60	30	10	2,5	4	3	0,2	0,3	0,5	41	21	6	34
8.	50	30	15	5	2	4	2,5	0,1	0,2	0,3	44	21	2	21
9.	50	30	15	5	2	5	3	0,2	0,4	0,5	44	26	0	27
10.	40	40	20	10	2	3	4	0,2	0,3	0,4	32	20	2	23
11.	40	20	12	8	2	1,8	1,6	0,1	0,3	0,5	35	10	5	20

Прокоментуємо, наприклад, сьомий приклад з табл. 1. Отже, отримали: якщо на початку бою A має 50 БО (середня ефективна скорострільність БО 2,5 постріли за годину), B має 60 БО і розподіляє на першу ДЗ 30 одиниць (середня ефективна скорострільність БО 4 постріли за годину), а на другу ДЗ теж 30 одиниць (середня ефективна скорострільність БО 3 постріли за годину), то для досягнення успіху A потрібно в початковий момент часу на першу ділянку розподілити 41 БО, а на другу тоді – 9 БО, потім в момент часу 0,2 години з першої ДЗ на другу перекинути 21 БО з тих БО, що залишились неушкодженими, а далі в момент часу 0,3 години розподілити 6 БО із числа резерву на першу ДЗ і 4 БО із числа резерву на другу ДЗ. В результаті в момент часу 0,5 години найбільші втрати B становлять 34 БО.

Наведемо приклад, коли задача не має розв'язання. Нехай $K = 15$, $L = 30$, $L_1 = 10$, $M = 5$, $\alpha_1 = 2,5$, $\alpha_2 = 2$, $\alpha_3 = 3$, $t_1 = 0,2$, $t_2 = 0,4$, $t_3 = 0,6$. Тоді допустимих значень параметрів K_1, K_2, K_3 не має і задача не має розв'язання, тобто бій на одній з ділянок закінчиться для A раніше часу 0,6 години. Якщо час перерозподілу ресурсів або резерву або час закінчення третього етапу бою зменшити, то тоді задача може мати розв'язання. Тобто, при даних кількостях БО обох сторін K, L, L_1 , резерву M і даних середніх ефективних скорострільностях БО $\alpha_1, \alpha_2, \alpha_3$ можна, змінюючи дані моменти часу t_1, t_2, t_3 , прогнозувати результат бою.

5 РЕЗУЛЬТАТИ

В даній роботі отримано наступне:

1. Розроблена математична модель динаміки бою на двох ділянках зіткнення як задачі динамічного програмування з функцією цілі як функцією втрат супротивника. В результаті розв'язання відповідних задач

Коші для систем ДР Ланчестера отримано кількості неушкоджених БО кожної сторони в кінці кожного з трьох етапів бою, а саме формули (1)–(4), (6)–(9), (11)–(14).

2. Виписано цільові функції (15)–(17), що підраховують відповідно найбільші втрати B на третьому етапі бою, сумарні найбільші втрати за третій і другий етапи бою і сумарні найбільші втрати за весь бій. З'ясовано допустимість параметрів задачі.

3. Для комп'ютерного моделювання розроблено алгоритм у вигляді блок-схеми (рис. 1–3), що реалізує метод динамічного програмування. Цей алгоритм дає змогу 1) виявляти допустимі значення параметрів задачі або давати відповідь, що задача розв'язання немає, тобто на одному або декількох етапах бою потрібно зменшити час закінчення відповідного етапу; 2) якщо існує хоча б одна допустима трійка параметрів, давати результат оптимального розподілу ресурсів і резерву A по ділянкам зіткнення для досягнення її успіху та надавати найбільші втрати противника в заданий момент часу; 3) прогнозувати бій для даних кількостей БО обох сторін, резерву A і відповідних ефективних скорострільностей, змінюючи часи закінчення трьох етапів бою.

4. На основі чисельних обчислень програми (табл. 1) зроблено висновок, що результати можуть бути непередбачуваними, тобто потрібно прорахувати за допомогою програми.

6 ОБГОВОРЕННЯ

Зауважимо, що між моментами перерозподілу бойових ресурсів і резерву можна розглянути будь-яку модель бою (це можливо дослідження наступних робіт), але головним є застосування до такого роду задач алгоритму, що реалізує метод динамічного програмування. Такі задачі потребують комп'ютерного моделювання, а саме внаслідок: 1) складності знаходження максимуму цільових функцій, причому мак-

симум знаходиться по одному параметру при відповідно фіксованих інших; 2) урахування допустимості параметрів задачі і знаходження оптимальної кількості бойових ресурсів, які має А розподіляти по ДЗ (в тому числі і резерв, що характерно для даної задачі); 3) складності самого метода динамічного програмування, враховуючи, що етапів бою може бути і більше трьох і параметрів перерозподілу разом з параметрами резерву може бути більше трьох.

ВИСНОВКИ

Наукова новизна. Розроблено математичну і комп'ютерну моделі динаміки бою двох ворогуючих сторін на двох ділянках зіткнення, в якій одна із сторін з метою нанесення противнику максимальних втрат може перерозподіляти бойові ресурси по ділянкам зіткнення і розподіляти з цією метою наявний в неї резерв. Для проведення чисельного експерименту розроблено відповідний алгоритм реалізації методу динамічного програмування.

Практична цінність. При використанні реальних характеристик бойових ресурсів розроблена модель може використовуватися для прогнозування результату бойових дій на двох ділянках зіткнення.

Перспективи подальших досліджень. В подальшому слід доопрацювати модель на випадок більше 2-х ділянок зіткнення, а також врахувати можливість збільшення моментів перерозподілу ресурсів і розподілу резерву. Крім того, доцільно також розглянути модель з ДР Ланчестера «високорганізованого», «мішаного» бою.

ПОДЯКИ

Дякуємо керівництво Харківського національного університету Повітряних Сил імені Івана Кожедуба за можливість проводити наукові дослідження.

ЛІТЕРАТУРА

1. Фурсенко О. К. Математичне моделювання бойових дій на двох ділянках зіткнення з можливістю перерозподілу бойових ресурсів / О. К. Фурсенко, Н. М. Черновол, Г. М. Антоненко // Системи обробки інформації. – 2022. – Випуск 4(171). – С. 76–81. DOI: <https://doi.org/10.30748/soi.2022.171.08>
2. Фурсенко О. К. Математичне моделювання бойових дій на двох ділянках зіткнення з використанням динамічного програмування і пакета символічної математики Wolfram Mathematica / О. К. Фурсенко, Н. М. Черновол, Г. М. Антоненко // Військово-

технічний збірник. – 2023. – Випуск 29. – С. 71–81. DOI: <https://doi.org/10.33577/2312-4458.29.2023.71-81>

3. Lubanovic B. Introducing Python: Modern Computing in Simple Packages / B. Lubanovic. – United States of America: O'Reilly Media, 2015. – 454 p.
4. Bellman R. Dynamic Programming. / R. Bellman. – United States of America : Princeton University Press, Princeton, New Jersey, 1972. – 342 p.
5. Dasgupta S. Algorithms / S. Dasgupta, C. H. Papadimitriou, U. Vazirani. – McGraw-Hill Science/Engineering/Math., 2006. – 336 p.
6. Bertele U. Nonserial dynamic programming / U. Bertele, F. Brioshi. – Y. : Academic Press, 1972. – 235 p.
7. Lanchester F. Aircraft in Warfare: the Dawn of the Fourth Arm / F. Lanchester. – London : Constable and Co, 1916. – 243 p.
8. Feng B. S. Study of air combat efficiency of carrier-borne fighter based on stochastic Lanchester battle theory / B. S. Feng, X. G. Zhou, Y. J. Lin // Computer Technology and Development. – 2013. – Vol. 23(5). – P. 199–201.
9. Mackay N. J. When Lanchester met Richardson, the outcome was Stalemate: a parable for mathematical models of insurgency / N. J. Mackay // J. of Operational Res. Soc. – 2015. – Vol. 66. – № 2. – P. 191–201. DOI: <https://doi.org/10.1057/jors.2013.178>
10. Vesa K. A Combat Equation Derived from Stochastic Modeling of Attrition Data. / K. Vesa // Military Operations Research. – 2015. – Vol. 20, Issue 3. – P. 49–69.
11. Lanchester model for three-way combat/ [M. Kress, J. P. Caulkins, G. Feichtinger et al.] // European Journal of Operational Research. – 2018. – Vol. 264, Issue 1. – P. 46–54. DOI: <https://doi.org/10.1016/j.ejor.2017.07.026>
12. Optimising structure in a networked Lanchester model for fires and manoeuvre in warfare / [A. C. Kalloniatis, K. Hoek, M. Zuparic et al.] // Journal of the Operational Research Society. – 2021. – Vol. 72(8). – P. 1863–1878. DOI: <https://doi.org/10.1080/01605682.2020.1745701>
13. Atkinson M. P. Targeting, deployment, and loss-tolerance in lanchester engagements / M. P. Atkinson, M. Kress., N. J. MacKay // Operations Research. – 2021. – Vol. 69(1). – P. 71–81. DOI: <https://doi.org/10.1287/opre.2020.2022>
14. Lanchester equation for cognitive domain using hesitant fuzzy linguistic terms sets / [Q. Han., W. M. Li, Q. L. Xu et al.] // Journal of Systems Engineering and Electronics. – 2022. – Vol. 33(3). – P. 674–682. DOI: <https://doi.org/10.23919/JSEE.2022.000062>
15. A swarm confrontation method based on Lanchester law and Nash equilibrium / [X. Ji, W. P. Zhang, F. T. Xiang et al.] // Electronics. – 2022. – Vol. 11(6). – P. 896–911. DOI: <https://doi.org/10.3390/electronics11060896>
16. Cangiotti N. A generalization of unaimed fire Lanchester's model in multi-battle warfare / N. Cangiotti, M. Capolli, S. Mattia // Operational Research. – 2023. – Vol. 23(2). – P. 38–57. DOI: <https://doi.org/10.1007/s12351-023-00776-8>
17. Kostić Mladen S. Modeling of combat operations / S. Mladen Kostić, D. AcaJovanović, V. Mitar Kovač // Vojnotehnicki glasnik/Military Technical Courier. – 2023. – Vol. 71, Issue. 3. – P. 529–558. DOI: <https://doi.org/10.5937/vojtehg71-43509>
18. Вентцель Е. С. Введение в исследование операций: монография / Е. С. Вентцель. – М. : Советское радио, 1964. – 388 с.

Стаття надійшла до редакції 22.10.2024.

Після доробки 16.02.2025.

UDC 519.8

MATHEMATICAL MODELLING OF COMBAT OPERATIONS WITH THE POSSIBILITY OF REDISTRIBUTING COMBAT RESOURCES BETWEEN THE AREAS OF CONTACT AND DISTRIBUTING RESERVES

Fursenko O. K. – PhD, Associate Professor, Head of the Department of Higher Mathematics, Ivan Kozhedub Kharkiv National Air Force University, Kharkiv, Ukraine.

Chernovol N. M. – Senior Lecturer of the Department of Higher Mathematics, Ivan Kozhedub Kharkiv National Air Force University, Kharkiv, Ukraine.

ABSTRACT

Context. Mathematical and computer models of the dynamics of combat operations are an important tool for predicting their outcome. The known Lanchester-type models were simulation models and did not take into account the ultimate goal and

© Фурсенко О. К., Черновол Н. М., 2025

DOI 10.15588/1607-3274-2025-1-6



redistribution of resources during combat operations. This paper proposes an optimisation model of the dynamics of combat operations between parties A and B in two areas of collision, based on the method of dynamic programming with maximisation of the objective function as a function of enemy losses. The article develops a mathematical and computer model of a typical situation in modern warfare of combat operations between parties A and B in two areas of collision with the aim of inflicting maximum losses of combat resources on the enemy. This goal is achieved by redistributing resources between the areas of collision and introducing appropriate reserves to these areas.

Objective. To build a mathematical and computer model of the dynamics of combat operations between parties A and B in two areas of collision, in which the goal of party A is to maximise the losses of party B by using three resources (the first is the number of combat units that party A can distribute across the areas of collision at the initial moment of time; the second is the number of combat units that party A must transfer from one area to another at some subsequent moment of time; the third is the number of combat units that party A must distribute using the reserve) and by modelling the

Method. The mathematical model is based on the method of dynamic programming with the objective function as a function of enemy losses, and the parameters are units of combat resources in different areas of the clash. Their number is changed by redistributing them between these areas and introducing reserve combat units. The enemy's losses are determined using Lanchester's systems of differential equations. Given the complexity of the objective function, the Python programming language is used to find its maximum.

Results. A mathematical model of the problem has been constructed and implemented, based on a combination of the dynamic programming method with the solution of Lanchester's systems of differential equations of battle dynamics with certain initial conditions at each of the three stages of the battle. With the help of a numerical experiment, the admissibility of the parameters of the optimisation problem (the number of combat units of side A, which are appropriately distributed, transferred from area to area or from the reserve at each stage of the battle) is analysed. The developed Python program allows, for any initial data, to give an answer to the optimal allocation of resources of party A, including from the reserve, at three stages of the battle and to calculate the corresponding largest enemy losses at a given time or to give an answer that there are no valid values of the problem parameters, i.e. the problem has no solution for certain initial data.

Conclusions. The scientific novelty lies in the development of mathematical and computer models of the dynamics of combat in two areas of collision, which takes into account the redistribution of combat resources and reserves in order to inflict maximum losses on the enemy. Numerical modelling made it possible to analyse the admissibility of redistribution and reserve parameters. Based on the examples considered, it is concluded that if the problem is unsolvable with certain data, it means that it is necessary to reduce the time of redeployment of combat units at one or more stages of the battle, i.e. to reduce the duration of the battle at a certain stage, thus allowing to predict the time of redeployment of combat resources.

KEYWORDS: objective function as a function of losses, differential equations of the dynamics of "poorly organised" combat, clash areas, redeployment of combat units, distribution of reserve combat units, effective rate of fire, maximisation of losses, admissibility of redeployment and reserve parameters.

REFERENCES

1. Fursenko O. K., Chernovol N. M., Antonenko H. M. Matematychnye modeljувannya bojovyh dij na dvoх diljankah zitkennja z mozhyvistju pererozpodilu bojovyh resursiv, *Systemy obrobky informacii*, 2022, Vypusk 4(171), pp. 76–81. DOI: <https://doi.org/10.30748/soi.2022.171.08>
2. Fursenko O. K., Chernovol N. M., Antonenko H. M. Matematychnye modeljувannya bojovyh dij na dvoх diljankah zitkennja z vykorystannjam dynamichnogo prohramuvannya ta paketa symvol'noyi matematyky Wolfram Mathematica, *Vijs'kovotekhnichnyj zbirnyk*, 2023, Vypusk 29, pp. 71–81. DOI: <https://doi.org/10.33577/2312-4458.29.2023.71-81>
3. Lubanovic B. Introducing Python: Modern Computing in Simple Packages. United States of America, O'Reilly Media, 2015, 454 p.
4. Bellman R. Dynamic Programming. United States of America, Princeton University Press, Princeton, New Jersey, 1972, 342 p.
5. Dasgupta S., Paradimitriou C. H., Vazirani U. Algorithms. McGraw-Hill Science/Engineering/Math., 2006, 336 p.
6. Bertele U., Brioshi F. Nonserial dynamic, Y., Academic Press, 1972, 235 p.
7. Lanchester F. Aircraft in Warfare: the Dawn of the Fourth Arm. London, Constable and Co, 1916, 243 p.
8. Feng B. S., Zhou X. G., Lin Y. J. Study of air combat efficiency of carrier-borne fighter based on stochastic Lanchester battle theory, *Computer Technology and Development*, 2013, Vol. 23(5), pp. 199–201.
9. Mackay N. J. When Lanchester met Richardson, the outcome was Stalemate: a parable for mathematical models of insurgency, *J. of Operational Res. Soc.*, 2015, Vol. 66, № 2, pp. 191–201. DOI: <https://doi.org/10.1057/jors.2013.178>
10. Vesa K. A Combat Equation Derived from Stochastic Modeling of Attrition Data, *Military Operations Research*, 2015, Vol. 20, Issue 3, pp. 49–69.
11. Kress M., Caulkins J.P., Feichtinger G., Grass D., Seidl A. Lanchester mode for three-way combat, *European Journal of Operational Research*, 2018, Vol. 264, Issue 1, pp. 46–54. DOI: <https://doi.org/10.1016/j.ejor.2017.07.026>
12. Kalloniatis A. C., Hoek K., Zuparic M., Brede M. Optimising structure in a networked Lanchester model for fires and manoeuvre in warfare, *Journal of the Operational Research Society*, 2021, Vol. 72(8), pp. 1863–1878. DOI: <https://doi.org/10.1080/01605682.2020.1745701>
13. Atkinson M. P., Kress M., MacKay N. J. Targeting, deployment, and loss-tolerance in lanchester engagements, *Operations Research*, 2021, Vol. 69(1), pp. 71–81. DOI: <https://doi.org/10.1287/opre.2020.2022>
14. Han Q., Li W.M., Xu Q.L. et al. Lanchester equation for cognitive domain using hesitant fuzzy linguistic terms sets, *Journal of Systems Engineering and Electronics*, 2022, Vol. 33(3), pp. 674–682. DOI: <https://doi.org/10.23919/JSEE.2022.000062>
15. Ji X., Zhang W. P., Xiang F. T., Yuan V., Chen Z. A swarm confrontation method based on Lanchester law and Nash equilibrium, *Electronics*, 2022, Vol. 11(6), pp. 896–911. DOI: <https://doi.org/10.3390/electronics11060896>
16. Cangioti N., Capolli M., Mattia S. A generalization of unaimed fire Lanchester's model in multi-battle warfare, *Operational Research*, 2023, Vol. 23(2), pp. 38–57. DOI: <https://doi.org/10.1007/s12351-023-00776-8>
17. Kostić Mladen S., AcaJovanović D., Mitar Kováč V. Modeling of combat operations, *Vojnotehnicki glasnik/Military Technical Courier*, 2023, Vol. 71, Issue. 3, pp. 529–558. DOI: <https://doi.org/10.5937/vojtehg71-43509>
18. Ventcel' E. S. Vvedenie v issledovanie operacij: monog-rafija. Moskow, Sovetskoe radio, 1964, 388 p.

НЕЙРОІНФОРМАТИКА ТА ІНТЕЛЕКТУАЛЬНІ СИСТЕМИ

NEUROINFORMATICS AND INTELLIGENT SYSTEMS

UDC 004.93

LIGHTWEIGHT MULTI-SCALE CONVOLUTIONAL TRANSFORMER FOR AIRCRAFT FAULT DIAGNOSIS USING VIBRATION ANALYSIS

Didenko Andrii Y. – Postgraduate student of the Department of Software Tools, National University “Zaporizhzhia Polytechnic”, Zaporizhzhia, Ukraine.

Didenko Artem Y. – Postgraduate student of the Department of Software Tools, National University “Zaporizhzhia Polytechnic”, Zaporizhzhia, Ukraine.

Subbotin S. A. – Dr. Sc., Professor, Head of the Department of Software Tools, National University “Zaporizhzhia Polytechnic”, Zaporizhzhia, Ukraine.

ABSTRACT

Context. Fault diagnosis in rotating machinery, especially in aircraft, plays an important role in health monitoring systems. Early and accurate fault detection can significantly reduce the cost of repair and increase the lifetime of the mechanism. To detect the fault efficiently, intelligent methods based on traditional machine learning and deep learning techniques are used. The object of the research is the process of detecting faults in aircraft based on vibration analysis.

Objective of the work is the development of a deep learning method for fault diagnosis in rotating machinery with a high accuracy rate.

Method. The proposed method employs Transformer architecture. The first stage of processing the vibration signal is the multi-scale feature extractor. This stage allows the model to examine input signals in different scales and reduce the impact of the noise. The second stage is the Convolutional Transformer neural network. The convolution was introduced to the Transformer to combine locality and long-range dependencies feature extraction. The Self-attention mechanism of the Transformer was changed to Channel Attention, which reduces the number of parameters but maintains the strength of the attention. To maintain this idea, similar changes were made in the position-wise feed-forward network.

Results. The proposed method is tested on the aircraft vibration dataset. Two conditions were chosen for testing: limited data and noisy environment. The limited data condition is simulated by selecting a small number of samples into the training set (a maximum of 10 per class). The noisy environment condition is simulated by adding Gaussian noise to the raw signal. According to the obtained results, the proposed method achieves a high average precision metric rate with a small number of parameters. The experiments also show the importance of the proposed modules and changes, confirming the assumptions about the process of feature extraction.

Conclusion. The results of the conducted experiments show that the proposed model can detect faults with almost perfect accuracy, even with a small number of parameters. The proposed lightweight model is robust in limited data conditions and noisy environment conditions. The prospects for further research are the development of fast and accurate neural networks for fault diagnosis and the development of limited data training techniques.

KEYWORDS: fault detection, deep learning, rotating machinery, signal processing, transformer, neural networks.

ABBREVIATIONS

AP is an Average Precision;
CA is a Channel Attention;
CNN is a Convolutional Neural Network;
DL is a Deep Learning;
FD is a Fault Detection;
FFN is a Feed-forward Network;
HUMS is a Health and Usage Monitoring System;
LMCT is a Lightweight Multi-scale Convolutional Transformer;
MHSA is a Multi-head Self-attention;
ML is a Machine Learning;
MLP is a Multi-layer Perceptron;

MSFE is a Multi-scale Feature Extractor;
NLP is a Natural Language Processing;
RNN is a Recurrent Neural Network;
SNR is a Signal-to-noise Ratio.

NOMENCLATURE

T is a time-series sensor output;
 t_i is a datapoint of sensor output;
 Y is a class label;
 D is a data set;
 M is a model;
 J is a classification metric;
 P_s is a signal power;

P_n is a noise power;
 N_p is a number of patches;
 f_{in} is a set of input features to the network module;
 f_{out} is a set of output features of the network module;
 f_{avg} is a set of features from the average pooling layer in CA module;
 f_{max} is a set of features from the max pooling layer in CA module;
 f^* is a set of features from intermediate layers in network modules.

INTRODUCTION

Rotating machinery is an integral part of mechanical systems. It finds its applications in many industrial sectors, including aircraft, wind turbines, pumps, car engines, etc. In aircraft and helicopters, rotating components like bearings, rotors, and gearboxes are critical for efficient operations. HUMS are sensor-based systems designed to monitor the health and performance of critical components as they are often subject to potential failure due to continuous mechanical stress (overload, overheating, lack of lubrication, etc). The early FD of these parts ensures the reliability and safety of the vehicle's exploitation, reducing the risk of critical failures and the repair cost.

Vibration-based analysis is a common FD technique in rotating machinery. Traditional ML methods often employ hand-crafted features combined with statistical-based algorithms like SVM [1], kNN [2], or decision trees [3]. While these methods can be used for fault prediction, they still struggle to capture complex features and require the complicated process of manual feature engineering. Recently, DL methods have gained popularity in the field of vibration analysis in FD [4], showing promising results in solving this task. The main advantage is the ability of DL-based models to capture complex data patterns and learn features of the input data, which eliminates the need for extensive feature engineering. (MLP [5] are commonly used in classification, CNNs [6–8] and RNNs [9, 10]. Also, a combination of CNN and RNN can be used [11, 12] to better analyze both local and temporal features of the sensor data. The recent popularity of the Transformer architecture [13] in solving sequence-based problems (including sensor data) stimulates the development of the Transformer-based FD methods.

Despite the exceptional performance of the DL-based methods for FD, these methods often need a large number of layers and parameters to extract comprehensive features, which can significantly increase the size of the model, training, and inference time and make the usage inefficient under limited resource conditions.

Moreover, one of the main issues with fault diagnosis is the limited amount of data. Usually, the number of positive samples (containing the fault) in the dataset is much smaller than the number of negative samples (healthy condition). This imbalance can significantly reduce the accuracy of the classification model, making predictions unreliable. To address this issue, techniques like zero-shot [14], few-shot [15], or transfer training can be used.

© Didenko Andrii Y., Didenko Artem Y., Subbotin S. O., 2025
DOI 10.15588/1607-3274-2025-1-7

The object of the research is the process of detecting faults in aircraft health monitoring systems.

The subject of the research is the deep learning method to detect faults using vibration-based analysis.

The purpose of the research is to develop and evaluate an efficient deep learning model to classify sensor data from aircraft health monitoring systems.

1 PROBLEM STATEMENT

Let $T = \{t_i | i = 1 \dots n\}$ be a vector of real values t_i of a time-series sensor data. A class label $Y \in \{0,1\}$ represents either an absence or a presence of a fault. Let $D = \{(T_i, Y_i) | i = 1 \dots N\}$ be a labeled dataset that contains ordered pairs (T_i, Y_i) . The objective is to choose such model architecture M that reaches the maximum classification accuracy, i.e., $J(Y_i, M(T_i)) \rightarrow \max$ for each $(T_i, Y_i) \in D$.

2 REVIEW OF THE LITERATURE

DL methods are widely used in rotating machinery FD tasks and have proven advantages over traditional ML methods [16]. Among them, CNNs play a significant role. Due to their strong feature extraction capabilities and local receptive field, CNNs can be successfully applied to solve vibration-based classification. As a vibration signal is one-dimensional data, 1D CNNs are the common choice. Vibration signals or other sensor data often contain localized patterns that indicate anomalies and faults, and by sliding 1D filters across the time-series data, 1D CNNs can effectively capture local, time-dependent changes that may signal an occurring fault. WDCNN method [17], for example, uses 1D CNN with wide kernels in convolutional layers, which helps to reduce noise and extract features from the input signal. DSNR method [18] employs deep residual architecture with soft thresholding (shrinkable function) to suppress redundant features and reduce the effect of the noise. Zhang et al. [19] also use a deep residual 1D CNN model to extract local data features to analyze faults. Here, residual learning helps to design deeper architectures to extract more complicated features. The authors of the AMMFN method [20] proposed the use of multi-sensor input data (vibration and current) processed by the 1D CNN model with the special attention-based fusion module that helps to extract features at different hierarchical levels and correlation information between sensors' signals.

Some methods combine 1D CNN with RNNs [11, 12]. This synthesis can improve the overall model performance by leveraging the strengths of each model type. This hybrid approach allows the network to capture both local patterns (via the CNN) and long-term dependencies (via the RNN) in time-series data. While these methods can achieve promising results, they require more training time due to the sequential nature of the RNNs.

Two-dimensional CNNs can also be used for sensor data analysis. To do this, the 1D signal is converted from the time domain to the time-frequency domain, and the obtained 2D spectrogram is used for further processing by the network. Verstraete et al. [21] proposed the use of 2D

CNN to analyze the Short-time Fourier Transform representation of the vibration signal. Pham et al. [22] employed the VGG16 model to analyze faults under inconsistent working conditions, showing excellent results. Although 2D CNN shows good performance, they require more samples to train in order to defeat overfitting. Moreover, 2D CNN usually requires more layers to extract deep features, which increases the model's size and inference time.

In recent years, Transformer-based models have become popular in many fields, including NLP [13], Computer Vision [23], Time-series analysis [24], FD [25], etc., outperforming common deep learning techniques. The main idea of Transformers lies behind the attention mechanism that allows focusing on the most relevant features and learning long-range dependencies more effectively. For example, the TST method [26] uses a 1D Transformer to directly analyze raw vibration signal data without pre-processing, showing high accuracy. The other studies [27] use 2D Transformers, called Vision Transformers [23], to analyze the spectrogram representation of the signal time-frequency domain. For example, ECTN [28] combines CNN and 2D Transformer to efficiently extract local and global information of the input signal. Integrated ViT model [29] decomposes the input signal with a Discrete Wavelet Transform and then applies soft voting to combine preliminary results.

Despite the advantages of the Transformers, they also have several notable disadvantages, including quadratic complexity of the attention mechanism and lack of intrinsic inductive bias [30], which makes them difficult to train. To address this issue in FD, the convolution can be introduced to Transformers [28, 31]. TWC method [32] is used in aircraft engine bearing FD and applies convolutional feature extraction layers to obtain features from raw input signal before passing it to the Transformer backbone. In TCN [33], a combined model of convolution and Transformer is trained with transfer learning technique.

Usually, the changes in vibration signal that refer to faults can vary in scale. Thus, using single-scale signal analysis, the fault can be easily overlooked. Multi-scale analysis can be used with different DL models. MCF-1DViT [25] uses a Multi-scale Convolution Fusion Layer to extract features at different time scales and pass them to the Transformer. AM-CNN [34] employs a Multi-scale convolutional block with CNN for the same purpose. Chen et al. [35] proposed MCNN-LSTM model with low-frequency and high-frequency feature extraction branches. Saghi et al. [11] developed CNN with three-scale branches followed by a bidirectional GRU model. These studies show that multi-scale feature extraction can effectively capture more information at different scales and enhance the overall performance of the method.

Recent studies demonstrate the potential of DL-based methods for FD in rotating machinery. However, further improvements are needed to reduce model complexity and the amount of noise in the input signal, achieve scale invariance, and improve local and global feature extraction, emphasizing the relevance of the proposed method.

3 MATERIALS AND METHODS

To solve the above challenges, the novel FD method called LMCT was developed. The proposed method mainly consists of two parts: MSFE and Transformer encoder. Fig 1. depicts the overall architecture of the neural network.

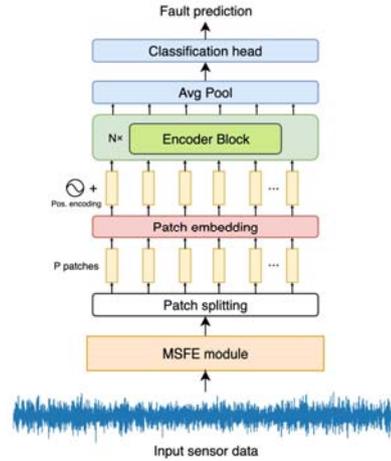


Figure 1 – Architecture of the proposed method

Both MSFE and Transformer encoder employ CA [36]. CA allows model to adjust the importance of each channel, by helping it to focus more on the informative ones while paying less attention to those that contribute less. The architecture of the CA module is show in Fig 2.

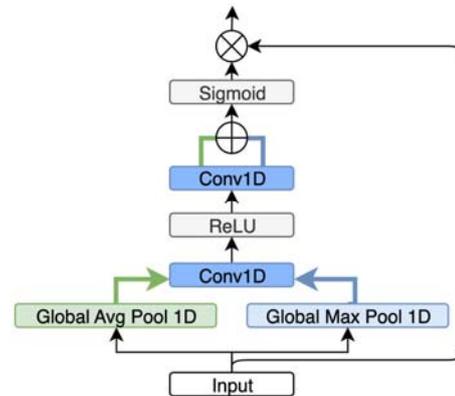


Figure 2 – Architecture of CA

Formally, the CA block used in this method is the following (1–3):

$$f_{avg} = Conv(RELU(Conv(AvgPool(f_{in}))), \quad (1)$$

$$f_{max} = Conv(RELU(Conv(MaxPool(f_{in}))), \quad (2)$$

$$f_{out} = Sigmoid(f_{avg} + f_{max}) \cdot f_{in}. \quad (3)$$

The MSFE (Fig. 3) module extracts preliminary features from the input signal at different scales, which makes the model more invariant to changes in the scale of

fault occurrence. Inspired by [25], it consists of 3 branches, each enhanced by CA. The kernel sizes of each branch B_n in MSFE are: 301, 401, 501. Wide kernels help to reduce the influence of the noise that might be present in the input signal. The outputs of the branches are then concatenated and processed by the convolutional layer to effectively merge important information.

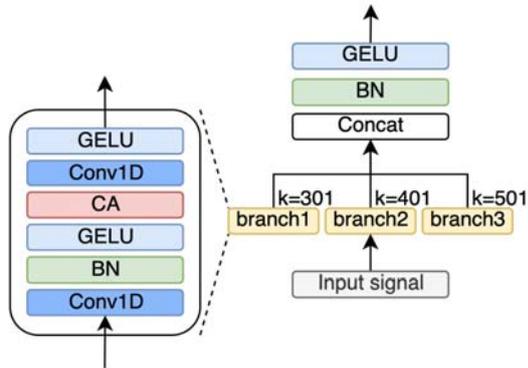


Figure 3 – Architecture of MSFE module

MSFE module is defined as follows (4):

$$f_{out} = GELU(BN(concat[B_1(f_{in}); B_2(f_{in}); B_3(f_{in})])). \quad (4)$$

After the processing by MSFE, the signal is split into N_p patches before passed to the Encoder. First, each patch is encoded with convolutional embedding module to enrich them with more information. Then, positional encodings are added to the patches to preserve the information about the relative position of each patch. After that, the information is passed to the Encoder blocks.

The Encoder block (Fig. 4) follows the standard architecture [13] but with several changes. MHSA was replaced with CA, which treats each patch as a channel. It allows to significantly reduce the number of parameters while keeping the attention mechanism in the model. Position-wise FFN (Fig. 5) was made fully convolutional containing depth-wise convolution. These changes also reduce the complexity of the model while fusing the information from each patch.

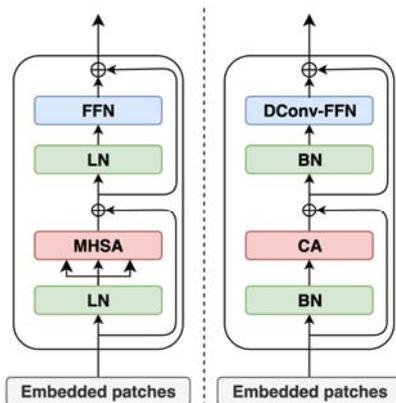


Figure 4 – A comparison of the vanilla Transformer block (left) and the proposed Transformer block (right)

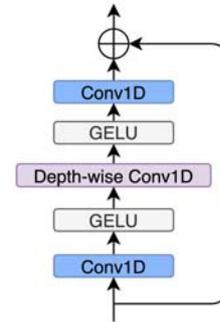


Figure 5 – Architecture of Position-wise FFN

Mathematically, the whole proposed Transformer block is defined as follows (5–6):

$$f^* = CA(BN(f_{in})) + f_{in}, \quad (5)$$

$$f_{out} = DconvFFN(BN(f^*)) + f^*. \quad (6)$$

Inspired by [31], the output of the Encoder is then passed through an average pooling layer and classification head to predict the type of the signal (healthy or faulty).

Table 1 describes the selected hyperparameters of the model that was used in the experiments.

Table 1 – The description of proposed model hyperparameters

Parameter	Value
Input size	1024
N_p	16
Patch embedding channels	64
Number of encoder blocks	6
FFN convolutional channels	64
Number of multi-scale branches	3
Branches' kernel sizes	(301, 401, 501)

4 EXPERIMENTS

To evaluate the performance of the proposed model, the aircraft fault dataset was used. This dataset contains 1158 samples of vibration signal, divided into two classes: 0 (healthy state) and 1 (fault). The amount of healthy and fault states is 865 and 293, respectively. Each sample is a vibration signal that contains 93752 datapoints. For training and testing, each sample was divided into chunks of length 1024 without overlapping, thus each sample provides 91 chunks. Each sample was scaled into the range of $(-1, 1)$ before splitting into chunks.

To analyze the performance of the proposed module under limited data constraints, the number of samples in the training set was chosen to be much smaller than in the test set. More precisely, several training sets were created with 1, 2, 5, and 10 vibration signal samples of each class. The rest of the data was split between validation and testing sets. The information about created datasets is shown in Table 2. For example, the subset “Train 10” contains 20 randomly selected samples (10 samples for the positive class and 10 samples for the negative). Each sample was divided into chunks of 1024 datapoints, thus “Train 10” subset split contains 1820 chunks in total.

Table 2 – The distribution of the created data subsets

Dataset split	Samples	Classes (0/1)	Chunks
Train 1	2	1/1	182
Train 2	4	2/2	364
Train 5	10	5/5	910
Train 10	20	10/10	1820
Validation	113	89/24	10283
Test	1025	766/259	93275

To compare the proposed model with the existing methods, several models were chosen: MCNN-LSTM [35], WDCNN [17], MSCNN [37], MRACNN [38], MA1DCNN [39], ResNet18 (1D) [40].

Each model was trained for 30 epochs in 5 independent runs. After that, the results of each run were averaged. The batch size of the training phase is 128. AdamW was chosen as an optimization algorithm with a learning rate of 0.001. The loss function is binary cross-entropy.

To evaluate the accuracy of the proposed method under a noisy environment, the Gaussian white noise was added to the raw signal to get the signal with different SNR (7) values. In these experiments, noise with SNR from -9dB to 9dB with the step of 3dB was added.

$$SNR = 10 \log_{10} \frac{P_s}{P_n} \quad (7)$$

For comparing results, AP was chosen as a metric. AP is a metric commonly used in binary classification, especially when dataset is highly imbalanced. It is essentially the area under the precision-recall curve, providing a single score that summarize the precision-recall trade-off across various thresholds. The range of AP metric values is (0–1), where 1 means errorless classification.

5 RESULTS

Table 3 shows the results on the test dataset of different methods. Each model is trained on datasets with a different number of vibration samples to evaluate the performance in limited data conditions. Also, the number of parameters of each model is included. The results represent the value of AP metric. The results show that the proposed method (LMCT) trained on a subset of 20 samples (“Train 10”) outperforms the chosen existing methods (AP 0.9941) while having a relatively small number of parameters (0.097M).

Table 3 – The prediction results of selected methods trained on different dataset sizes

Method	N samples per class				N params (M)
	1	2	5	10	
MCNN-LSTM	0.3154	0.4016	0.4342	0.5421	0.093
WDCNN	0.3455	0.6399	0.8322	0.9216	0.041
MSCNN	0.3337	0.7790	0.7998	0.9511	13.78
MRACNN	0.3372	0.5732	0.9182	0.9699	0.599
MA1DCNN	0.3174	0.8043	0.8626	0.9094	0.323
Resnet18 (1D)	0.3346	0.6631	0.8982	0.9556	3.84
LMCT (proposed)	0.3133	0.7314	0.9105	0.9941	0.097

Note that these results represent the performance of the models on the dataset that is collected from samples’ chunks. To analyze the whole vibration sample, the average prediction of each chunk can be used.

Table 4 shows the ablation study results to evaluate the performance and the importance of proposed changes, especially the MSFE module and new Convolutional Transformer Encoder block. The baseline model, the 1D Transformer, follows the standard Transformer architecture. The hyperparameters in the baseline were chosen to make it as close to the proposed method as possible.

Table 4 – The ablation study of the effect of proposed improvements

Method	N samples per class			
	1	2	5	10
1D Transformer	0.3122	0.2976	0.2963	0.3919
1D Transformer + MSFE	0.3122	0.3904	0.4869	0.6063
LMCT w/o MSFE	0.3137	0.6335	0.7523	0.8056
LMCT	0.3133	0.7314	0.9105	0.9941

Figure 6 shows the results (AP) of the selected models under noisy environment conditions (SNR from -9 to 9). For that, models that were trained on 20 samples (subset “Train 10”) were chosen.

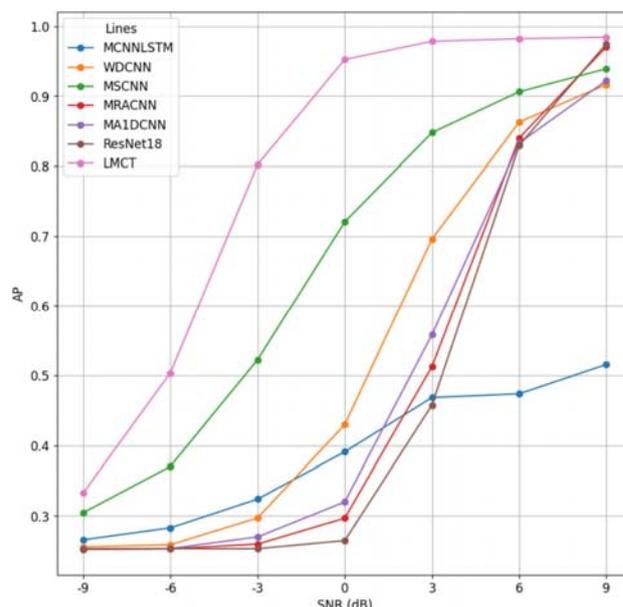


Figure 6 – The results under different SNRs of Gaussian noise

The visualization of the learnt features (features from the last layer of the network) using t-SNE method is shown in the Fig. 7.

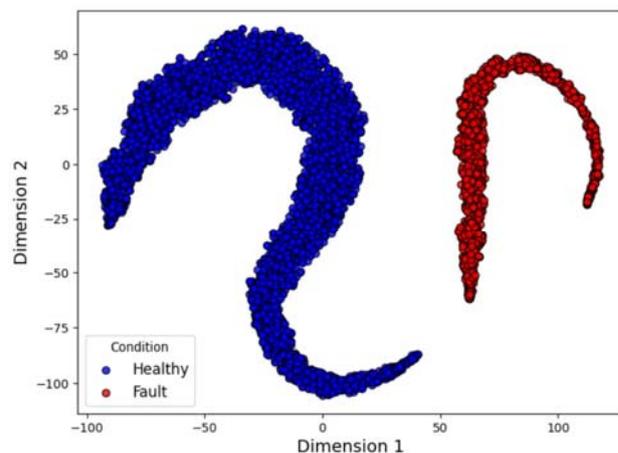


Figure 7 – Distribution of features visualized with t-SNE

6 DISCUSSION

The results from Table 3 show that the proposed model achieves almost ideal accuracy (AP 0.9941). Moreover, the number of parameters of the proposed model is relatively small (0.091M) compared to the existing approaches, which means that a small model is able to achieve good results in vibration-based FD tasks. This can be explained by the fact that the proposed methods employ a combination of convolution and attention mechanisms, which allows the model to examine local features as well as long-term dependencies.

The results of the ablation study in Table 4 suggest the importance of the proposed modules. The MSFE module significantly improves the performance, which means that it is able to extract valuable deep features from the input signal. The Transformer backbone is able to process these features more accurately than the raw input signal. Also, the vanilla Transformer model is not able to accurately classify aircraft sensor vibration signals. It can be explained by the fact that each signal can contain local patterns that can influence the overall prediction results. To extract these patterns, the strength of convolutional inductive locality bias is needed.

Figure 6 shows that the proposed model can be used in noisy environments effectively. The proposed model achieves AP of 0.9783, 0.9821, and 0.9842 at SNR values of 3dB, 6dB, and 9dB, respectively. When SNR is higher than 0, it means that the energy of the original signal is higher than the energy of the added noise and vice versa. In cases when the energy of the noise is higher than the energy of the original signal, classification is difficult to perform. The results, obtained under different SNR values, indicate that wide kernels in MSFE help the model to reduce the impact of the noise in the signal.

Figure 7, which shows the 2D representation of extracted features, suggests that the model is able to learn the feature distribution by extracting patterns and dependencies between similar entries. The fault features and healthy features are clustered, which means that model can separate between both clusters.

CONCLUSIONS

In this paper, the problem of fault diagnosis of aircraft rotating machinery is being solved by applying deep learning method.

The scientific novelty. The conducted experiments show that proposed method can achieve high accuracy having small number of parameters. The proposed changes made to Transformer architecture significantly increase model's performance while reducing it's size. Moreover, the results show that FD methods can be effectively trained with limited size of the data which is important in the field of aircraft FD where fault data samples are extremely rare.

The practical significance. The obtained model allows diagnosing faults in aircraft with high accuracy and can be applied to other rotating machinery FD tasks that use vibration-based analysis.

The prospects for further research lie in developing more efficient model and testing with other limited data training techniques like zero-shot or few-shot learning.

ACKNOWLEDGEMENTS

This work was performed within the scope of the All-Ukrainian competition-hackathon of scientific works of young scientists in the field of intellectual information technologies, 2024. The data was provided by Bechhoefer Eric, PhD, GPMS Inc. (Cornwall, USA).

REFERENCES

1. Baccarini L. M. R., Silva V. V. R., de Menezes B. R. et al. SVM practical industrial application for mechanical faults diagnostic, *Expert Systems with Applications*, 2011, Vol. 38, № 6, pp. 6980–6984. DOI: 10.1016/j.eswa.2010.12.017
2. Lu Q., Shen X., Wang X. et al. Fault diagnosis of rolling bearing based on improved VDM and KNN, *Mathematical Problems in Engineering*, 2021, Vol. 2021, Issue 1, №2530315. DOI: 10.1155/2021/2530315
3. Sun W., Chen J., Li J. Decision tree and PCA-based fault diagnosis of rotating machinery, *Mechanical Systems and Signal Processing*, 2007, Vol. 21, Issue 3, pp. 1300–1317. DOI: 10.1016/j.ymsp.2006.06.010
4. Zhao R., Yan R., Chen Z. et al. Deep learning and its applications to machine health monitoring, *Mechanical Systems and Signal Processing*, 2019, Vol. 115, pp. 213–237. DOI: 10.1016/j.ymsp.2018.05.050
5. Samanta B., Al-Balushi K., and Al-Araimi S. Artificial neural networks and genetic algorithm for bearing fault detection, *Soft Computing*, 2006, Vol. 10, pp. 264–271. DOI: 10.1007/s00500-005-0481-0
6. Wang X., Mao D., Li X. Bearing fault diagnosis based on vibro-acoustic data fusion and 1D-CNN network, *Measurement*, 2021, Vol. 173, №108518. DOI: 10.1016/j.measurement.2020.108518
7. Pinedo-Sanchez L. A., Mercado-Ravell D. A., Carballo-Monsivais C. A. Vibration analysis in bearings for failure prevention using CNN, *Journal of the Brazilian Society of Mechanical Sciences and Engineering*, 2020, Vol. 42, № 12, № 628. DOI: 10.1007/s40430-020-02711-w
8. Yu Z., Zhang C., Liu J. et al. SKND-TSACNN: A novel time-scale adaptive CNN framework for fault diagnosis of rotating machinery, *Knowledge-Based Systems*, 2023, Vol. 275, P. 110682. DOI: 10.1016/j.knosys.2023.110682;
9. Huang Y., Chen C. H., Huang C. J. Motor fault detection and feature extraction using RNN-based variational autoencoder, *IEEE Access*, 2019, Vol. 7, pp. 139086–139096. DOI: 10.1109/ACCESS.2019.2940769
10. Liu H., Zhou J., Zheng Y. et al. Fault diagnosis of rolling bearings with recurrent neural network-based autoencoders, *ISA Transactions*, 2018, Vol. 77, pp. 167–178. DOI: 10.1016/j.isatra.2018.04.005
11. Saghi T., Bustan D., Aphale S. S. Bearing fault diagnosis based on multi-scale CNN and bidirectional GRU, *Vibration*, 2023, Vol. 6, № 1, pp. 11–28. DOI: 10.3390/vibration6010002
12. Qiao M., Yan S., Tang X. et al. Deep Convolutional and LSTM Recurrent Neural Networks for Rolling Bearing Fault Diagnosis Under Strong Noises and Variable Loads, *IEEE Access*, 2020, Vol. 8, pp. 66257–66269. DOI: 10.1109/ACCESS.2020.2985617
13. Vaswani A., Shazeer N., Parmar N. et al. Attention is All you Need [Electronic resource], *Semantic Scholar*, 2017,

- Access mode:
<https://www.semanticscholar.org/paper/Attention-is-All-you-Need-Vaswani-Shazeer/204e3073870fae3d05bcbc2f6a8e263d9b72e776>.
DOI: 10.48550/arXiv.1706.03762
14. Xu J., L. Zhou, W. Zhao et al. Zero-shot learning for compound fault diagnosis of bearings, *Expert Systems with Applications*, 2022, Vol. 190, №116197. DOI: 10.1016/j.eswa.2021.116197
 15. Zhang A., Li S., Cui Y. et al. Limited Data Rolling Bearing Fault Diagnosis With Few-Shot Learning, *IEEE Access*, 2019, Vol. 7, pp. 110895–110904. DOI: 10.1109/ACCESS.2019.2934233
 16. Chen C. C., Liu Z., Yang G. et al. An improved fault diagnosis using 1D-convolutional neural network model, *Electronics*, 2021, Vol. 10, Issue 1, pp. 1–19. DOI: 10.3390/electronics10010059
 17. Zhang W., Peng G., Li C. et al. A New Deep Learning Model for Fault Diagnosis with Good Anti-Noise and Domain Adaptation Ability on Raw Vibration Signals, *Sensors*, 2017, Vol. 17, Issue 2, №425. DOI: 10.3390/s17020425
 18. Zhao M., Zhong S., Fu X. et al. Deep Residual Shrinkage Networks for Fault Diagnosis, *IEEE Transactions on Industrial Informatics*, 2020, Vol. 16, Issue 7, pp. 4681–4690. DOI: 10.1109/TII.2019.2943898
 19. Zhang W., Li X., Ding Q. Deep residual learning-based fault diagnosis method for rotating machinery, *ISA Transactions*, 2019, Vol. 95, pp. 295–305. DOI: 10.1016/j.isatra.2018.12.025
 20. Li X., Wan S., Liu S. et al. Bearing fault diagnosis method based on attention mechanism and multilayer fusion network, *ISA Transactions*, 2021, Vol. 128, Part B, pp. 550–564. DOI: 10.1016/j.isatra.2021.11.020
 21. Verstraete D., Ferrada A., Droguett E. et al. Deep Learning Enabled Fault Diagnosis Using Time-Frequency Image Analysis of Rolling Element Bearings, *Shock and Vibration*, 2017, Vol. 2017, Issue 1, №5067651. DOI: 10.1155/2017/5067651
 22. Pham M. T., Kim J.-M., and Kim C. H. Accurate Bearing Fault Diagnosis under Variable Shaft Speed using Convolutional Neural Networks and Vibration Spectrogram, *Applied Sciences*, 2020, Vol. 10, Issue 18, № 6385. DOI: 10.3390/app10186385;
 23. Dosovitskiy A., Beyer L., Kolesnikov A. et al. An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale [Electronic resource], *ICLR*, 2021. Access mode: <https://arxiv.org/abs/2010.11929>. DOI: 10.48550/arXiv.2010.11929
 24. Zhou H., Zhang S., Peng J. et al. Informer: Beyond Efficient Transformer for Long Sequence Time-Series Forecasting, *Proceedings of the AAAI conference on artificial intelligence*, 2021, Vol. 35, Issue 12, pp. 11106–11115. DOI: 10.48550/arXiv.2012.07436
 25. Weng C., Lu B., Yao J. A One-Dimensional Vision Transformer with Multiscale Convolution Fusion for Bearing Fault Diagnosis, *Global Reliability and Prognostics and Health Management*. N., IEEE, 2021, pp. 1–6. DOI: 10.1109/PHM-Nanjing52125.2021.9612919
 26. Jin Y., Hou L., Chen Y. A Time Series Transformer based method for the rotating machinery fault diagnosis, *Neurocomputing*, 2022, Vol. 494, pp. 379–395. DOI: 10.1016/j.neucom.2022.04.111
 27. Zim A. H., Ashraf A., Iqbal A. et al. A Vision Transformer-Based Approach to Bearing Fault Classification via Vibration Signals, *Asia-Pacific Signal and Information Processing Association Annual Summit and Conference*. C., IEEE, 2022, pp. 1321–1326. DOI: 10.48550/arXiv.2208.07070
 28. Liu W., Zhang Z., Zhang J. et al. A Novel Fault Diagnosis Method of Rolling Bearings Combining Convolutional Neural Network and Transformer, *Electronics*, 2023, Vol. 12, Issue 8, №1838. DOI: 10.3390/electronics12081838
 29. Tang X., Xu Z., Wang Z. A Novel Fault Diagnosis Method of Rolling Bearing Based on Integrated Vision Transformer Model, *Sensors*, 2022, Vol. 22, Issue 10, №3878. DOI: 10.3390/s22103878
 30. Xu Y., Zhang Q., Zhang J. et al. ViTAE: vision transformer advanced by exploring intrinsic inductive bias, *NIPS'21 : 35th International Conference on Neural Information Processing Systems 6–14 December 2021 : proceedings*. Red Hook, NY, Curran Associates Inc., 2021, Vol. 34, pp. 28522–28535. DOI: 10.48550/arXiv.2106.03348
 31. Sun W., Yan R., Jin R. et al. LiteFormer: A Lightweight and Efficient Transformer for Rotating Machine Fault Diagnosis, *IEEE Transactions on Reliability*, 2024, Vol. 73, Issue 2, pp. 1258–1269. DOI: 10.1109/TR.2023.3322860
 32. Fang X., Deng X., Chen J. et al. Convolution Transformer Based Fault Diagnosis Method For Aircraft Engine Bearings, *43rd Chinese Control Conference, Kunming, 28–31 July 2024 : proceedings*. IEEE, 2024, pp. 5038–5042. DOI: 10.23919/CCC63176.2024.10662769
 33. Pei X., Zheng X., Wu J. Rotating Machinery Fault Diagnosis Through a Transformer Convolution Network Subjected to Transfer Learning, *IEEE Transactions on Instrumentation and Measurement*, 2021, Vol. 70, №2515611. DOI: 10.1109/TIM.2021.3119137
 34. Jin Y., Qin C., Zhang Z. et al. A multi-scale convolutional neural network for bearing compound fault diagnosis under various noise conditions, *Science China Technological Sciences*, 2022, Vol. 65, Issue 11, pp. 2551–2563. DOI: 10.1007/s11431-022-2109-4
 35. Chen X., Zhang B., Gao D. Bearing fault diagnosis base on multi-scale CNN and LSTM model, *Journal of Intelligent Manufacturing*, 2021, Vol. 32, Issue 4, pp. 971–987. DOI: 10.1007/s10845-020-01600-2
 36. Hu J., Shen L., Sun G. Squeeze-and-Excitation Networks, *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, 18–23 June 2018 : proceedings*. IEEE, 2018, pp. 7132–7141. DOI: 10.1109/CVPR.2018.00745
 37. Jiang G., He H., Yan J. et al. Multiscale Convolutional Neural Networks for Fault Diagnosis of Wind Turbine Gearbox, *IEEE Transactions on Industrial Electronics*, 2019, Vol. 66, Issue 4, pp. 3196–3207. DOI: 10.1109/TIE.2018.2844805;
 38. Jia L., Chow T. W. S., Wang Y. et al. Multiscale Residual Attention Convolutional Neural Network for Bearing Fault Diagnosis, *IEEE Transactions on Instrumentation and Measurement*, 2022, Vol. 71, №3519413. DOI: 10.1109/TIM.2022.3196742
 39. Wang H., Liu Z., Peng D. et al. Understanding and Learning Discriminant Features based on Multiattention 1DCNN for Wheelset Bearing Fault Diagnosis, *IEEE Transactions on Industrial Informatics*, 2020, Vol. 16, Issue 9, pp. 5735–5745. DOI: 10.1109/TII.2019.2955540
 40. He K., Zhang X., Ren S. et al. Deep Residual Learning for Image Recognition, *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, 27–30 June 2016*. IEEE, 2016, pp. 770–778, DOI: 10.1109/CVPR.2016.90

Received 09.12.2024.
Accepted 03.02.2025.

ЛЕГКОВІСНИЙ БАГАТОМАСШТАБНИЙ ЗГОРТКОВИЙ ТРАНСФОРМЕР ДЛЯ ДІАГНОСТИКИ НЕСПРАВНОСТЕЙ ЛІТАЛЬНИХ АПАРАТІВ ЗА ДОПОМОГОЮ ВІБРАЦІЙНОГО АНАЛІЗУ

Діденко Андрій Є. – аспірант кафедри програмних засобів національного університету «Запорізька Політехніка», Запоріжжя, Україна.

Діденко Артем Є. – аспірант кафедри програмних засобів національного університету «Запорізька Політехніка», Запоріжжя, Україна.

Субботін С. О. – д-р техн. наук, професор, завідувач кафедри програмних засобів національного університету «Запорізька Політехніка», Запоріжжя, Україна.

АНОТАЦІЯ

Актуальність. Діагностика несправностей обертових механізмів, особливо в авіації, відіграє важливу роль в системах моніторингу стану. Своєчасне і точне виявлення несправностей може значно знизити вартість ремонту і збільшити термін служби механізму. Для ефективного виявлення несправностей використовуються інтелектуальні методи, які базуються на традиційних методах машинного та глибинного навчання. Об'єктом дослідження є процес виявлення несправностей в авіаційних апаратах на основі аналізу вібрацій.

Метою роботи є розробка методу глибинного навчання для діагностики несправностей обертових машин з високою точністю.

Метод. Запропонований метод використовує архітектуру трансформера. Першим етапом обробки сигналу вібрації є багатомасштабне вилучення ознак. Цей етап дозволяє моделі розглядати вхідні сигнали в різних масштабах і зменшити вплив шуму. Другий етап – згорткова нейронна мережа з трансформером. Згортка була додана до трансформера, щоб поєднати локальність і вилучення ознак далеких залежностей. Механізм самоуваги трансформера було змінено на механізм каналної уваги, що зменшує кількість параметрів, але зберігає силу уваги. Щоб підсилити цю ідею, аналогічні зміни були зроблені в позиційній мережі прямого поширення.

Результати. Запропонований метод протестовано на наборі даних з вібраціями авіаційного апарату. Для тестування було обрано дві умови: обмеженість обсягу даних та зашумлене середовище. Обмеженість обсягу даних імітується шляхом використання невеликої кількості вибірок до навчального набору даних (максимум 10 на клас). Умова зашумленого середовища імітується шляхом додавання гауссівського шуму до вихідного сигналу. Згідно з отриманими результатами, запропонований метод досягає високої середньої точності при невеликій кількості параметрів. Експерименти також показують важливість запропонованих модулів і змін, підтверджуючи припущення про процес вилучення ознак.

Висновки. Результати проведених експериментів показують, що запропонована модель може виявляти несправності з майже ідеальною точністю, навіть при невеликій кількості параметрів. Запропонована легковісна модель є стійкою в умовах обмеженого обсягу даних та зашумленого середовища. Перспективами подальших досліджень є розробка швидких і точних нейронних мереж для діагностики несправностей та розробка методів навчання на обмежених обсягах даних.

КЛЮЧОВІ СЛОВА: аналіз несправностей, глибинне навчання, обертові механізми, обробка сигналів, трансформер, нейронні мережі.

ЛІТЕРАТУРА

1. SVM practical industrial application for mechanical faults diagnostic / [L. M. R. Baccarini, V. V. R. Silva, B. R. de Menezes et al.] // *Expert Systems with Applications*. – 2011. – Vol. 38, № 6. – P. 6980–6984. DOI: 10.1016/j.eswa.2010.12.017
2. Fault diagnosis of rolling bearing based on improved VDM and KNN / [Q. Lu, X. Shen, X. Wang et al.] // *Mathematical Problems in Engineering*. – 2021. – Vol. 2021, Issue 1. – №2530315. DOI: 10.1155/2021/2530315
3. Sun W. Decision tree and PCA-based fault diagnosis of rotating machinery / W. Sun, J. Chen, J. Li // *Mechanical Systems and Signal Processing*. – 2007. – Vol. 21, Issue 3. – P. 1300–1317. DOI: 10.1016/j.ymsp.2006.06.010
4. Deep learning and its applications to machine health monitoring / [R. Zhao, R. Yan, Z. Chen et al.] // *Mechanical Systems and Signal Processing*. – 2019. – Vol. 115. – P. 213–237. DOI: 10.1016/j.ymsp.2018.05.050
5. Samanta B. Artificial neural networks and genetic algorithm for bearing fault detection / B. Samanta, K. Al-Balushi, and S. Al-Araimi // *Soft Computing*. – 2006. – Vol. 10. – P. 264–271. DOI: 10.1007/s00500-005-0481-0
6. Wang X. Bearing fault diagnosis based on vibro-acoustic data fusion and 1D-CNN network / X. Wang, D. Mao, X. Li // *Measurement*. – 2021. – Vol. 173. – №108518. DOI: 10.1016/j.measurement.2020.108518
7. Pinedo-Sanchez L. A. Vibration analysis in bearings for failure prevention using CNN / L. A. Pinedo-Sanchez, D. A. Mercado-Ravell, C. A. Carballo-Monsivais // *Journal of the Brazilian Society of Mechanical Sciences and Engineering*. – 2020. – Vol. 42, № 12. – №628. DOI: 10.1007/s40430-020-02711-w
8. SKND-TSACNN: A novel time-scale adaptive CNN framework for fault diagnosis of rotating machinery / [Z. Yu, C. Zhang, J. Liu et al.] // *Knowledge-Based Systems*. – 2023. – Vol. 275. – P. 110682. DOI: 10.1016/j.knsys.2023.110682
9. Huang Y. Motor fault detection and feature extraction using RNN-based variational autoencoder / Y. Huang, C. H. Chen, C. J. Huang // *IEEE Access*. – 2019. – Vol. 7. – P. 139086–139096. DOI: 10.1109/ACCESS.2019.2940769
10. Fault diagnosis of rolling bearings with recurrent neural network-based autoencoders / [H. Liu, J. Zhou, Y. Zheng et al.] // *ISA Transactions*. – 2018. – Vol. 77. – P. 167–178. DOI: 10.1016/j.isatra.2018.04.005
11. Saghi T. Bearing fault diagnosis based on multi-scale CNN and bidirectional GRU / T. Saghi, D. Bustan, S. S. Aphale // *Vibration*. – 2023. – Vol. 6, № 1. – P. 11–28. DOI: 10.3390/vibration6010002
12. Deep Convolutional and LSTM Recurrent Neural Networks for Rolling Bearing Fault Diagnosis Under Strong Noises and Variable Loads / [M. Qiao, S. Yan, X. Tang et al.] // *IEEE Access*. – 2020. – Vol. 8. – P. 66257–66269. DOI: 10.1109/ACCESS.2020.2985617
13. Attention is All you Need [Electronic resource] / [A. Vaswani, N. Shazeer, N. Parmar et al.] // *Semantic Scholar*. –

2017. – Access mode: <https://www.semanticscholar.org/paper/Attention-is-Allyou-Need-Vaswani-Shazeer/204e3073870fae3d05bcbc2f6a8e263d9b72e776>. DOI: 10.48550/arXiv.1706.03762
14. Zero-shot learning for compound fault diagnosis of bearings / [J. Xu, L. Zhou, W. Zhao et al.] // *Expert Systems with Applications*. – 2022. – Vol. 190. – №116197. DOI: 10.1016/j.eswa.2021.116197
15. Limited Data Rolling Bearing Fault Diagnosis With Few-Shot Learning / [A. Zhang, S. Li, Y. Cui et al.] // *IEEE Access*. – 2019. – Vol. 7. – P. 110895–110904. DOI: 10.1109/ACCESS.2019.2934233
16. An improved fault diagnosis using 1D-convolutional neural network model / [C. C. Chen, Z. Liu, G. Yang et al.] // *Electronics*. – 2021. – Vol. 10, Issue 1. – P. 1–19. DOI: 10.3390/electronics10010059
17. A New Deep Learning Model for Fault Diagnosis with Good Anti-Noise and Domain Adaptation Ability on Raw Vibration Signals / [W. Zhang, G. Peng, C. Li et al.] // *Sensors*. – 2017. – Vol. 17, Issue 2. – №425. DOI: 10.3390/s17020425
18. Deep Residual Shrinkage Networks for Fault Diagnosis / [M. Zhao, S. Zhong, X. Fu et al.] // *IEEE Transactions on Industrial Informatics*. – 2020. – Vol. 16, Issue 7. – P. 4681–4690. DOI: 10.1109/TII.2019.2943898
19. Zhang W. Deep residual learning-based fault diagnosis method for rotating machinery / W. Zhang, X. Li, Q. Ding // *ISA Transactions*. – 2019. – Vol. 95. – P. 295–305. DOI: 10.1016/j.isatra.2018.12.025
20. Bearing fault diagnosis method based on attention mechanism and multilayer fusion network / [X. Li, S. Wan, S. Liu et al.] // *ISA Transactions*. – 2021. – Vol. 128, Part B. – P. 550–564. DOI: 10.1016/j.isatra.2021.11.020
21. Deep Learning Enabled Fault Diagnosis Using Time-Frequency Image Analysis of Rolling Element Bearings / [D. Verstraete, A. Ferrada, E. Droguett et al.] // *Shock and Vibration*. – 2017. – Vol. 2017, Issue 1. – №5067651. DOI: 10.1155/2017/5067651
22. Pham M. T. Accurate Bearing Fault Diagnosis under Variable Shaft Speed using Convolutional Neural Networks and Vibration Spectrogram / M. T. Pham, J.-M. Kim, and C. H. Kim // *Applied Sciences*. – 2020. – Vol. 10, Issue 18. – №6385. DOI: 10.3390/app10186385
23. An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale [Electronic resource] / [A. Dosovitskiy, L. Beyer, A. Kolesnikov et al.] // *ICLR*. – 2021. – Access mode: <https://arxiv.org/abs/2010.11929>. DOI: 10.48550/arXiv.2010.11929
24. Informer: Beyond Efficient Transformer for Long Sequence Time-Series Forecasting / [H. Zhou, S. Zhang, J. Peng et al.] // *Proceedings of the AAAI conference on artificial intelligence*. – 2021. – Vol. 35, Issue 12. – P. 11106–11115. DOI: 10.48550/arXiv.2012.07436
25. Weng. C. A One-Dimensional Vision Transformer with Multiscale Convolution Fusion for Bearing Fault Diagnosis / C. Weng, B. Lu, J. Yao // *Global Reliability and Prognostics and Health Management*. – N. : IEEE, 2021. – P. 1–6. DOI: 10.1109/PHM-Nanjing52125.2021.9612919
26. Jin Y. A Time Series Transformer based method for the rotating machinery fault diagnosis / Y. Jin, L. Hou, Y. Chen // *Neurocomputing*. – 2022. – Vol. 494. – P. 379–395. DOI: 10.1016/j.neucom.2022.04.111
27. A Vision Transformer-Based Approach to Bearing Fault Classification via Vibration Signals / [A. H. Zim, A. Ashraf, A. Iqbal et al.] // *Asia-Pacific Signal and Information Processing Association Annual Summit and Conference*. – C.:IEEE, 2022. – P. 1321–1326. DOI: 10.48550/arXiv.2208.07070
28. A Novel Fault Diagnosis Method of Rolling Bearings Combining Convolutional Neural Network and Transformer / [W. Liu, Z. Zhang, J. Zhang et al.] // *Electronics*. – 2023. – Vol. 12, Issue 8. – №1838. DOI: 10.3390/electronics12081838
29. Tang X. A Novel Fault Diagnosis Method of Rolling Bearing Based on Integrated Vision Transformer Model / X. Tang, Z. Xu, Z. Wang // *Sensors*. – 2022. – Vol. 22, Issue 10. – №3878. DOI: 10.3390/s22103878
30. ViTAE: vision transformer advanced by exploring intrinsic inductive bias / [Y. Xu, Q. Zhang, J. Zhang et al.] // *NIPS'21 : 35th International Conference on Neural Information Processing Systems 6–14 December 2021 : proceedings*. – Red Hook, NY : Curran Associates Inc., 2021. – Vol. 34. – P. 28522–28535. DOI: 10.48550/arXiv.2106.03348
31. LiteFormer: A Lightweight and Efficient Transformer for Rotating Machine Fault Diagnosis / [W. Sun, R. Yan, R. Jin et al.] // *IEEE Transactions on Reliability*. – 2024. – Vol. 73, Issue 2, P. 1258–1269. DOI: 10.1109/TR.2023.3322860
32. Convolution Transformer Based Fault Diagnosis Method For Aircraft Engine Bearings / [X. Fang, X. Deng, J. Chen et al.] // *43rd Chinese Control Conference, Kunming, 28–31 July 2024 : proceedings*. – IEEE, 2024. – P. 5038–5042. DOI: 10.23919/CCC63176.2024.10662769
33. Pei X. Rotating Machinery Fault Diagnosis Through a Transformer Convolution Network Subjected to Transfer Learning / X. Pei, X. Zheng, J. Wu // *IEEE Transactions on Instrumentation and Measurement*. – 2021. – Vol. 70. – №2515611. DOI: 10.1109/TIM.2021.3119137
34. A multi-scale convolutional neural network for bearing compound fault diagnosis under various noise conditions / [Y. Jin, C. Qin, Z. Zhang et al.] // *Science China Technological Sciences*. – 2022. – Vol. 65, Issue 11. – P. 2551–2563. DOI: 10.1007/s11431-022-2109-4
35. Chen X. Bearing fault diagnosis base on multi-scale CNN and LSTM model / X. Chen, B. Zhang, D. Gao // *Journal of Intelligent Manufacturing*. – 2021. – Vol. 32, Issue 4. – P. 971–987. DOI: 10.1007/s10845-020-01600-2
36. Hu J. Squeeze-and-Excitation Networks / J. Hu, L. Shen, G. Sun // *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, 18–23 June 2018 : proceedings*. – IEEE, 2018. – P. 7132–7141. DOI: 10.1109/CVPR.2018.00745
37. Multiscale Convolutional Neural Networks for Fault Diagnosis of Wind Turbine Gearbox / [G. Jiang, H. He, J. Yan et al.] // *IEEE Transactions on Industrial Electronics*– 2019. – Vol. 66, Issue 4. – P. 3196–3207. DOI: 10.1109/TIE.2018.2844805
38. Multiscale Residual Attention Convolutional Neural Network for Bearing Fault Diagnosis / [L. Jia, T. W. S. Chow, Y. Wang et al.] // *IEEE Transactions on Instrumentation and Measurement*. – 2022. – Vol. 71. – №3519413. DOI: 10.1109/TIM.2022.3196742
39. Understanding and Learning Discriminant Features based on Multiattention 1DCNN for Wheelset Bearing Fault Diagnosis / [H. Wang, Z. Liu, D. Peng et al.] // *IEEE Transactions on Industrial Informatics*. – 2020. – Vol. 16, Issue 9. – P. 5735–5745. DOI: 10.1109/TII.2019.2955540
40. Deep Residual Learning for Image Recognition / [K. He, X. Zhang, S. Ren et al.] // *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, 27–30 June 2016. – IEEE, 2016. – P. 770–778, DOI: 10.1109/CVPR.2016.90*

APPROACH TO DATA DIMENSIONALITY REDUCTION AND DEFECT CLASSIFICATION BASED ON VIBRATION ANALYSIS FOR MAINTENANCE OF ROTATING MACHINERY

Molchanova M. O. – Postgraduate student of the Department of Computer Sciences, Khmelnytskyi National University, Khmelnytskyi, Ukraine.

Didur V. O. – Student of the Department of Computer Sciences, Khmelnytskyi National University, Khmelnytskyi, Ukraine.

Mazurets O. V. – PhD, Associate Professor, Department of Computer Science, Khmelnytskyi National University, Khmelnytskyi, Ukraine.

ABSTRACT

Context. The actual problem of effective intelligent diagnostics of malfunctions of rotating equipment is solved. The object of study is the process of data dimensionality reduction and defect classification based on vibration analysis for maintenance of rotating machines. The subject of study is the methods of dimension reduction and defect classification by vibration analysis.

Objective. Development of an approach to data dimensionality reduction and defect classification based on vibration analysis for maintenance of rotating machines

Method. The comprehensive approach to data dimensionality reduction and defect classification based on vibration analysis is proposed, which solves the problem of data dimensionality reduction for training classifiers and defect classification, and also solves the problem of building a neural network classifier capable of ensuring the speed of fault classification without loss of accuracy on data of reduced dimensionality. The approach differs from the existing ones by the possibility of using optional union and intersection operators when forming a set of significant features, which provides flexibility and allows to adapt to different contexts and data types, ensuring classification efficiency in cases of large-dimensional data.

A denoising method allows to preserve important information, avoiding redundancy and improving the quality of data for further analysis. It involves calculating the signal-to-noise ratio, setting thresholds, and applying a fast Fourier transform that separates relevant features from noise. Applying the LIME method to a set of machine learning models allows to identify significant features with greater accuracy and interpretability. This contributes to more reliable results, as LIME helps to understand the influence of each feature on the final model solution, which is especially important when working with large datasets, where the importance of individual features may not be obvious. The implementation of optional operators of union and intersection of significant features provides additional flexibility in choosing an approach to defining important features. This allows the method to be adapted to different contexts and data types, ensuring efficiency even in cases with a large number of features.

Results. The developed method was implemented in software and examined when solving the problem of defect classification based on vibration analysis for maintenance of rotating machines.

Conclusions. The conducted experimental studies confirmed the high efficiency and workability of the proposed approach for reducing the dimensionality of data and classifying defects based on vibration analysis in the aspect of maintenance of rotating machines. Prospects for further research will be directed to the search for alternative neural network architectures and their training to reduce training time.

KEYWORDS: dimensionality reduction, significant features, defect detection, MobileNetV2.

ABBREVIATIONS

CBM is a Condition-based Maintenance;
PM is a Predictive Maintenance;
ISO is an International Organization for Standardization;
CNC is a computer numerical control;
CNN is a Convolutional Neural Network;
MSVM is a Multiclass Support Vector Machine;
PdM-CNN is a Predictive Maintenance using Convolutional Neural Network;
VMD is a Variational Mode Decomposition;
CWT is a Continuous Wavelet Transform;
SVM is a Support Vector Machine;
CWRU is a Case Western Reserve University;
KNN is a K-Nearest Neighbors;
LIME is a Local Interpretable Model-agnostic Explanations;
FTX is a Feature eXtractor;
NNT is a Neural Network Trainer.

NOMENCLATURE

x is a set of vibration characteristics of machines;
 y is a corresponding classes that include “with fault” and “without fault”;
 F is a structure of the model, which allows to reduce the dimensionality of the vibration data;
 w is a model parameters to be optimized;
 x' is a selection of characteristics from the original set x ;
 x_s is a represents individual vibrational characteristics;
 y' is a corresponding classes for the training data subset;
 y_s is a corresponding class for characteristics x_s ;
 $MLset$ is a set of machine learning models;
 ML_1 is a first machine learning model in $MLset$;
 ML_q is a q -th machine learning model in $MLset$;
 SNR is a signal-to-noise ratio;
 SNR_w is a percentage threshold value for denoising;
 n is a number of models from $MLset$ to generate a list of numbers of significant features;

q is a total number of machine learning models;
 nk is a total number of features in the sample xr ;
 μ is an arithmetic mean value of the features for the sample xr ;
 xr_1 is a first feature in the sample xr ;
 xr_i is a i -th feature in the sample xr ;
 xr_{nk} is a nk -th feature in the sample xr ;
 T is a threshold value;
 FT is a Fourier transform function;
 X is a signal with Fourier transform;
 X' is a signal with Fourier transform filtered according to the T threshold;
 w_1 is a weight of the first feature in the sample xr ;
 w_i is a weight of the i -th feature in the sample xr ;
 w_s is an average value by characteristics;
 Sf is a set of numbers of significant signs;
 Op is a set of optional operations;
 Sf_1 is a first important feature;
 Sf_j is a j -th important feature;
 Sf_n is a last important feature;
 Sf_{res} is a set of important features;
Accuracy is a metric the proportion of correct predictions made by the model across the entire dataset;
 TP is a number of correctly identified positive instances;
 TN is a number of correctly identified negative instances;
 FP is a number of incorrectly identified positive instances;
 FN is a number of incorrectly identified negative instances.
Precision is a metric the proportion of TP predictions among all positive predictions made by the model;
Recall is a metric also known as sensitivity or TP rate, metric the proportion of TP predictions among all actual positive instances;
 $F1$ is a metric that balances is a metric that balances *Precision* and *Recall*.

INTRODUCTION

Complex mechanical systems such as rotating machines require expensive maintenance to prevent accidents that can cost lives or cause serious damage to the system itself [1]. Every year, the industry spends billions of dollars on maintenance, which can account for up to a third of production costs. Improved maintenance can significantly reduce the overall cost of the system over its lifetime. One of the modern approaches to maintenance in complex rotating machinery based on condition-based maintenance techniques is the use of vibration analysis [2]. In this approach, vibration sensors are installed near the rotating components of the system, and signal processing algorithms are used to detect faults and classify their sources.

The object of study is the process of data dimensionality reduction and defect classification based on vibration analysis for maintenance of rotating machines.

The subject of study is methods of dimensionality reduction and defect classification by vibration analysis.

The purpose of the work is to improve the quality of maintenance of rotating machines by reducing the dimensionality of the data, which will allow to reduce computational costs and improve the speed of data processing without deteriorating the accuracy of diagnostics.

1 PROBLEM STATEMENT

In today's rotating machinery maintenance industry, vibration analysis is critical for timely defect detection and accident prevention. However, processing large volumes of vibration data can be challenging due to the high dimensionality, which complicates the defect classification process.

The problem is formulated as follows: given a set of precedents $\langle x, y \rangle$, it is necessary to develop a structure of the model $F()$, which allows to reduce the dimensionality of the vibration data, while preserving the information important for the classification of defects.

Optimize parameters w of the model based on $\langle x', y' \rangle$, where $x' \subset \{xs\}$, $y' = \{ys \mid xs \in x'\}$, which makes it possible to achieve the best results in the classification of defects in two classes: "with a fault" and "without a fault".

Conduct a comparative analysis of model results to confirm the effectiveness of the data dimensionality reduction approach in defect classification tasks.

The task of this study is to develop an effective method of data dimensionality reduction, which will ensure the accuracy and speed of fault classification, thus improving the quality of maintenance of rotating machines.

2 REVIEW OF THE LITERATURE

Vibration measurements are critical for fault diagnosis in industrial equipment because they provide information about the condition of rotating equipment. A review [3] documents state-of-the-art deep learning methods for equipment health monitoring based on vibration signals. Numerous studies from two leading databases, Web of Science and Scopus, were selected for review. After careful analysis, 59 studies were selected and systematically analyzed. The purpose of the review is to provide researchers with an in-depth understanding of fault diagnosis techniques that use deep learning to process vibration signals.

A study [4] discusses the importance of creating an effective maintenance system, in particular CBM and PM that complement each other. The use of CBM and PM in the study focuses on the case study of three pumps in a chili sauce factory in Jakarta, Indonesia, demonstrating that vibration analysis using accelerometers is an effective method for condition monitoring of rotating machinery. Calculation of Root Mean Square and comparison of the result with ISO 10816 allows to determine the current state of the engine, and the fast Fourier transform helps in

grouping vibrations by frequencies for damage analysis. The study shows that technology-enabled service is affordable for small and medium-sized businesses, and the use of machine learning will improve future predictions.

In [5] and [6], the importance of bearings for CNC machines and their role in ensuring the reliable operation of machines is considered. Signals in the conditions of faulty and normal bearings are analyzed, emphasizing the importance of early diagnosis of faults. Despite the importance of the topic, the problem is complicated by the insufficiency of bearing fault databases and the presence of noise in the vibration and acoustic signals, which makes automatic fault identification difficult. A CNN was applied to extract high-level features from vibration and acoustic signals, which enabled the training of MSVM. Experimental analysis showed that the proposed method can diagnose bearing faults of CNC machines with a classification accuracy of 98.9% based on the combined features of vibration and acoustic signals.

The use of the PdM-CNN model is proposed for the automatic classification of rotating equipment malfunctions and recommendations on the need for maintenance [7]. The paper uses data from only one vibration sensor mounted on the bearing on the drive side of the motor, which is the most common location in industry. The study was carried out under controlled conditions with variations in rotational speed, load levels and fault intensities to test whether it is possible to build a model capable of classifying such faults using only one set of vibration sensors. The results showed that the accuracy of the PdM-CNN model was 99.58% and 97.3% when applied to two different public databases.

Research [8] proposed a hybrid bearing fault diagnosis method based on VMD, CWT, CNN, and SVM, which is suitable for processing small samples. First, the resampled data are subjected to VMD processing, after which the reconstructed IMF data are overlaid and sampled to obtain a 2D time-frequency image using CWT. Next, a CNN model is created with the selected hyperparameters, and training samples are fed into the CNN to train the model. A pre-trained CNN model is used to stepwise train test samples to extract fault features, and SVM is used instead of the Softmax function to identify and classify faults.

The effectiveness of the proposed method is confirmed using the vibration data of the CWRU bearings and the test bench for testing the spindle device, where the classification accuracy was on average 99.9% for the former and 90.15% for the latter.

However, the need for a large amount of training data for training neural networks can be noted as a problem for effective intelligent diagnosis of malfunctions of rotating equipment [9, 10]. In addition, this data may not contain samples to identify rare types of damage, so it is advisable to choose pre-trained neural networks that can take advantage of Zero-Shot Learning as fault detection models.

Another problem is the high dimensionality of the data, a large part of which is uninformative. This leads to

an increase in the time of diagnosis and training of models.

3 MATERIALS AND METHODS

To implement the approach to data dimensionality reduction and defect classification based on vibration analysis for the maintenance of rotating machines, the following tasks must be solved:

- 1) data dimensionality reduction for classifier training and defect classification;
- 2) construction of a neural network classifier capable of ensuring the speed of fault classification without loss of accuracy on data of reduced dimensions.

The general scheme of the approach to data dimensionality reduction and defect classification based on vibration analysis is shown in Figure 1.

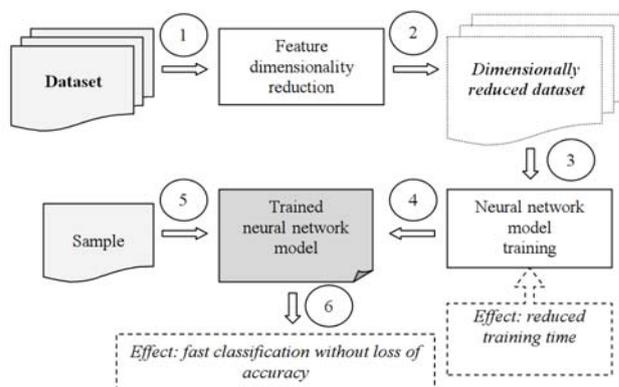


Figure 1 – Scheme of the proposed approach

Figure 1 shows the key points of the proposed approach to data dimensionality reduction and defect classification by vibration analysis. The approach primarily solves the problem of reducing the dimensionality of the data for further training of the neural network model, which allows to achieve the effect of reducing the time spent on training without loss of accuracy. In the future, the trained neural network is able to classify defects based on significant features based on vibration analysis.

To solve the problem of reducing the dimensionality of data for training classifiers and classifying defects, a method of reducing the dimensionality of data is proposed, the scheme of which is shown in Figure 2.

The input data of the dimensionality reduction method is the dataset provided within the framework of the competition “All-Ukrainian competition of young scientists in the field of intellectual IT” (<https://zp.edu.ua/vkiit/>), the set $MLset$ (1):

$$MLset = \{ML_1, \dots, ML_q\}, \quad (1)$$

SNR_w (determined empirically, by default equal to 0.9), an optional operation to create a dataset of significant features based on union or intersection, as well as n models from $MLset$ to create a list of significant feature numbers.

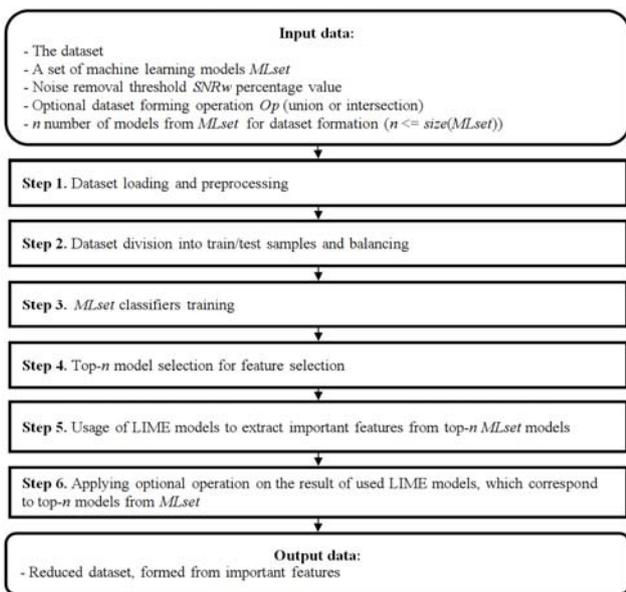


Figure 2 – The method of data dimensionality reduction

The following model architectures were used in the study: SVM, Naive Bayes, RandomForest, Gradient Boosting and KNN. Within these architectures, different sets of models can be created by changing the hyperparameters specific to each model.

In step 1, labels and their corresponding files with vibrations are loaded. Vibration signals are often expressed as noise from various sources, such as electrical interference, environmental factors, and measurement errors. To solve this problem, various methods of signal processing were developed, one of which is the Fourier transform [11]. The loaded dataset is preprocessed by denoising all samples using transformations (2–6). For each sample, a signal-to-noise ratio (2) is calculated:

$$SNR(x) = \frac{\frac{1}{nk} \sum_{i=1}^{nk} xr_i^2}{\frac{1}{nk} \sum_{i=1}^{nk} (xr_i - \mu)^2}, \quad (2)$$

which is then divided by the user's desired ratio in order to set a threshold that would discard noise and not affect potentially important data.

The threshold value T is determined by transformation (3):

$$T = \begin{cases} -\infty, & \text{if } SNRw = 0, \\ \min(|xr_i|) \text{ where } xr_i \neq 0, & \text{if } SNRw = 1, \\ \max(|xr_i|) \times SNR \times (1 - SNRw), & \text{if } 0 < SNRw < 1. \end{cases} \quad (3)$$

Next, the fast Fourier transform FT [12] (4) is calculated for each sample:

$$X(f) = FT(xr_i). \quad (4)$$

After that, the coefficients are filtered according to the T threshold (5):

$$X'(f) = \begin{cases} X(f), & \text{if } |X(f)| \geq T, \\ 0, & \text{if } |X(f)| < T. \end{cases} \quad (5)$$

For the filtered coefficients from (5), the inverse Fourier transform is calculated to obtain the filtered signal (6):

$$x'(i) = F^{-1}(X'(f)). \quad (6)$$

In step 2, the dataset, where the transformation (2–6) was applied to each element, is divided into training and validation samples in the ratio of 80/20. If the dataset is unbalanced, SMOTE balancing is additionally applied to the training sample [13].

In step 3, models are trained from the set $MLset$. Parameters can be modified within each architecture.

In step 4, the top- n models are selected for further obtaining a list of significant features. The selection takes place according to the criterion F1 of the metric [14] in descending order.

In step 5, a set of LIME models, the number of which is equal to n , is used to obtain a list of significant features for the models selected in step 4 from the $MLset$, since LIME can be applied to any model, regardless of its type [15]. The LIME model focuses on explaining individual predictions by generating locally linear models around the prediction point [16]. This allows for detailed analysis of why the model made the prediction it did for a particular case, which is useful for identifying significant features. Significant features are selected according to the condition “if the weight of the feature w_i is greater than or equal to the average value w_s for the entire set of features, it is important”, features with a lower weight are discarded. At this step, the set Sf is obtained from the numbers of significant features (7):

$$\{Sf\} = \{i, | w_i \geq w_s\}. \quad (7)$$

In step 6, provided that $n \geq 2$, an optional operation is performed on the result of the n LIME models used from the set $Op \in \{\cup, \cap\}$ – defined by the user at the input data stage. For all possible pairs of sets of LIME models, sets with numbers of important features are formed by transformation (8):

$$\{Sf_{res}\} = Op_{j-1}^n Sf_j. \quad (8)$$

The output is a reduced dimensionality dataset, where each sample has $size\{Sf_{res}\}$, formed from a list of significant feature numbers. Accordingly, the created dataset will be used as input data to solve the second task of the research – the construction of a neural network

classifier capable of ensuring the speed of fault classification without loss of accuracy.

The method of obtaining a fault classification model is shown in Figure 3.

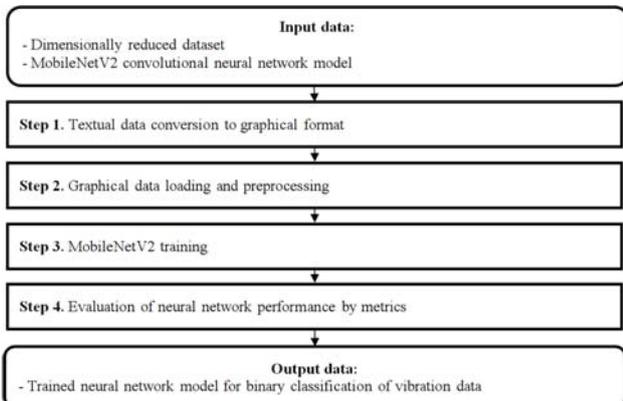


Figure 3 – The method of obtaining a fault classification model

The input data is a reduced dimensionality dataset and a pre-trained MobileNetV2 convolutional neural network model.

MobileNet-v2 is considered to be an efficient CNN model for mobile devices (or embedded systems) with the potential characteristics of small size, low latency, and low power consumption [17], which are important characteristics for vibration analysis defect classification for rotating machinery maintenance. MobileNet-v2 is optimized to achieve high accuracy with significantly reduced computational costs. This is achieved through the use of depth-distributed convolutional layers and inverted residual blocks.

In step 1, the reduced-dimensional text dataset is transformed into a graphical one by plotting the data of each sample on a graph, where the ordinal numbers of the features are on the X axis, and their values are on the Y axis (or in another way: time moments are on X, and powers of vibrations at given moments of time are on Y). An example of the transformation of step 1 is shown in Figure 4.

In step 2, the generated graph dataset is loaded and each sample is processed using the MobileNetV2 convolutional neural network image preprocessor (mobilenet_v2.preprocess_input).

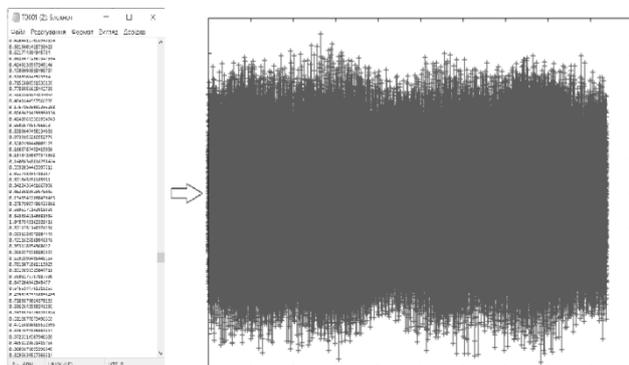


Figure 4 – Converting a text representation into a graphic representation

In step 3, the neural network is trained. MobileNetV2 model, using transfer learning. A neural network is retrained for the task of binary classification of vibration graphs. The pre-trained MobileNetV2 model is used only for feature extraction from images, instead additional GlobalAveragePooling2D and Dense layers are trained (Figure 5).

At step 4, the trained neural network model is evaluated according to the following metrics: Accuracy, Precision, Recall, F1 [18].

The *Accuracy* is calculated by formula (9):

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (9)$$

The *Precision* is calculated by formula (10):

$$Precision = \frac{TP}{TP + FP} \quad (10)$$

The *Recall* is calculated by formula (11):

$$Recall = \frac{TP}{TP + FN} \quad (11)$$

And the *F1* is calculated by formula (12):

$$F1 = 2 \cdot \frac{Precision \cdot Recall}{Precision + Recall} \quad (11)$$

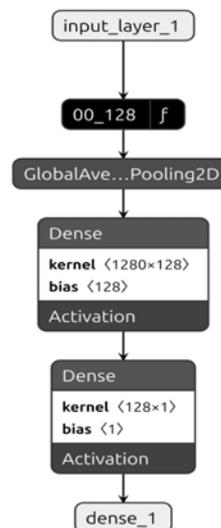


Figure 5 – Architecture of the neural network used

The output of the method is a trained model of the MobileNetV2 convolutional neural network for binary classification of vibration graphs.

So, within the framework of the developed approach to data dimensionality reduction and defect classification based on vibration analysis for the maintenance of rotating machines, the problem of data dimensionality reduction was solved by using agnostic LIME models,

and the problem of building a neural network classifier based on the MobileNetV2 architecture, capable of ensuring the speed of fault classification without loss of accuracy.

4 EXPERIMENT

Based on the proposed approach, a software complex was developed in the form of three applications with a command line interface and a web page. The software implementation and instructions for deployment and use are available on GitHub (https://github.com/Pravetz/mmdv_vkiiit). The interaction between the components of the software complex and data flows is shown in Figure 6. The language for creating the components of the software complex is Python.

“FTX” is an application for the problem of reducing the dimensionality of the data set. It uses libraries sklearn (for training classifiers, dividing the dataset into samples) [19], imblearn (for SMOTE-balancing of the training sample, it happens by default and is not controlled by the user through console parameters) [20], lime (creating explanations, extracting important features) [21], scipy (auxiliary functions for the algorithm for extracting important features) and numpy (working with data arrays) [22].

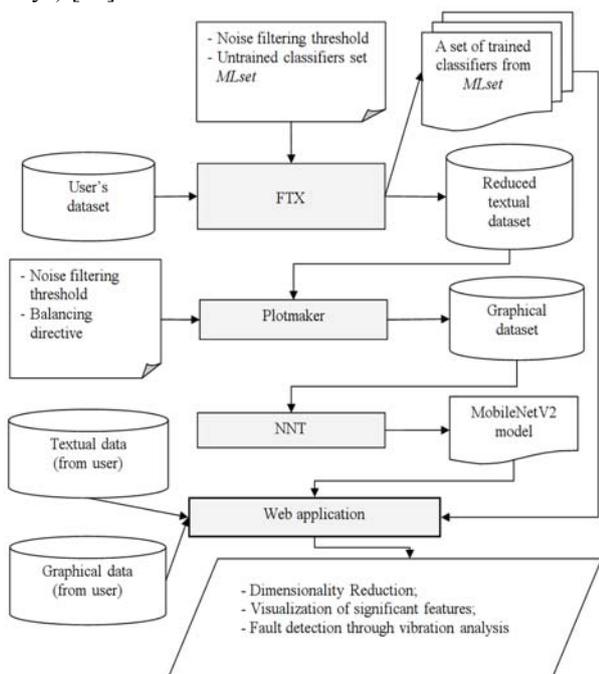


Figure 6 – Interaction between components of the software complex and data flows

The application for performing the task of building a neural network classifier (starting from step 2) “NNT” also does not have a graphical user interface. In the model training mode, the program loads the dataset from the specified path, processes it with the MobileNetV2 neural network image preprocessor, then trains the model for a user-specified number of epochs, evaluates its performance, and saves the trained model according to the specified path.

Libraries such as tensorflow (for training the MobileNetV2 model, loading the dataset) [23], sklearn (dividing the dataset into samples, evaluating the model according to the main metrics), numpy (for working with data arrays) [24] and matplotlib (for saving model training history) [25] are used by it.

The “Plotmaker” application for converting a text dataset into a graphic one performs step 1 of the task of building a neural network classifier.

This application is the smallest of all and performs the preparation of graphical data for MobileNetV2 training. From the libraries, imblearn is used here for SMOTE-balancing of the dataset, graphs are drawn by transferring drawing data using a pipeline to the gnuplot program [26].

The console applications described above were created to be used in the form of a “pipeline”: first, “FTX” creates a dataset of reduced dimensionality, then the created dataset is passed to “Plotmaker”, which creates a graphical data set, which in turn is passed to “NNT” for training a neural network classifier.

The web application (Figure 7) serves to demonstrate the results of the console applications and allows: to use trained versions of the MobileNetV2 model trained on different versions of the dataset (full, reduced by union of significant features, intersection of significant features) to determine the state of the machine component according to the vibration graph, provided by the user; view the important features for each of the classifiers from the *MLset* on an interactive graph, where for each feature its number and weight are written during decision-making.

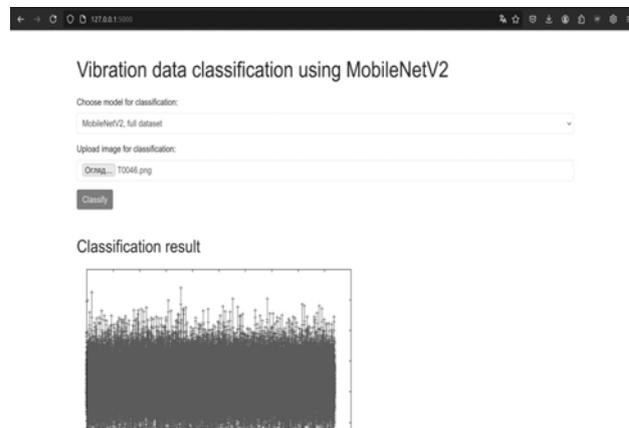


Figure 7 – Web application interface

Based on the developed software complex, a study of the effectiveness of the proposed approach to data dimensionality reduction and defect classification based on vibration analysis for maintenance of rotating machines will be conducted.

The plan for researching the effectiveness of the proposed approach:

1. Study the influence of the application of noise cleaning on the classification of defects of parts of rotating machines.

2. Study significant features for classifiers from the set *MLset*, and find the optimal top-*n* for the selected sample in which the F1 metric will be as close to 1 as possible.

3. Study the ability of MobileNetV2 neural network to learn on data with different numbers of significant features.

During the experiment *MLset* of length 5 was formed, which includes the following models: Gradient Boosting, Random Forest, Naive Bayes, SVM, KNN. These models were trained on the base dataset in 2 variants: without applying transformations (2)–(6) and with their application. The results by metrics are shown in Table 1.

The transformation (7)–(8) is applied to each model from *MLset*, which is included in the top-*n*. The results for values of *n* from 2 to 4 are shown in Table 2.

5 RESULTS

Within the dimensionality reduction task, a set of *MLset* classifiers was trained, the data of the experiment according to the Accuracy, Precision, Recall, F1 metrics, as well as the training time in seconds are shown in Table 1.

Table 1 – Evaluation of classifiers from *MLset* by metrics

Metrics	Gradient Boosting	Random Forest	Naive Bayes	SVM	KNN
<i>Original dataset</i>					
Accuracy	0.784	0.776	0.780	0.797	0.224
Precision	0.761	0.602	0.829	0.776	0.05
Recall	0.784	0.776	0.780	0.797	0.224
F1	0.711	0.678	0.688	0.780	0.082
Training time (seconds)	3670	6	2	109	1
<i>Datasets with transformations (2)–(6) applied</i>					
Accuracy	0.978	0.974	0.961	0.853	0.970
Precision	0.98	0.977	0.967	0.861	0.970
Recall	0.978	0.974	0.961	0.853	0.970
F1	0.979	0.975	0.962	0.829	0.970
Training time	3643	6	2	56	1

Table 2 shows the data of the dimensionality reduction experiment with the original dataset (without applying transformations (2)–(6)). Number of features in the input dataset: 93752.

For *n*=4. Top-4 models by F1 score:

1. SVM.
2. Gradient Boosting.
3. Naive Bayes.
4. Random Forest.

The minimum score of the top-4 model according to F1 metric: 0.678. Table 2 shows the pairwise application of the optional operators “intersection” and “union” between the classifiers included in the top 4 on the dataset with noise.

Table 2 – Application of “intersection” and “union” operators to reduce dimensionality (without (2)–(6))

	SVM	Gradient Boosting	Naive Bayes	Random Forest
<i>For “Intersection” operation</i>				
SVM	39930	16875	16951	16995
Gradient Boosting	16875	39627	16946	16804
Naive Bayes	16951	16946	39838	16962
Random Forest	16995	16804	16962	39758
<i>For “Union” operation</i>				
SVM	39930	62682	62817	62693
Gradient Boosting	62682	39627	62519	62581
Naive Bayes	62817	62519	39838	62634
Random Forest	62693	62581	62634	39758

The intersection of the top 2 models (SVM, Gradient Boosting) reduced the number of significant features to 16875. The intersection of the top 3 models (SVM, Gradient Boosting, Naive Bayes) reduced the number of significant features to 7221. The intersection of the top 4 models (SVM, Gradient Boosting, Naive Bayes, Random Forest) reduced the number of significant features to 3097. At the same time, the application of the “union” operator for the top 2 models (SVM, Gradient Boosting) reduced the number of features to 62682. The union of the top 3 models (SVM, Gradient Boosting Boosting and Naive Bayes) reduced the number of features to 75844. Combining the top 4 models (SVM, Gradient Boosting, Naive Bayes and Random Forest) reduced the number of features to 83390.

Table 3 shows the data of the dimensionality reduction experiment with the denoised dataset (using transformations (2)–(6)). Number of features in the input dataset: 93752.

Table 3 – Application of “intersection” and “union” operators to reduce dimensionality (with (2)–(6))

	Gradient Boosting	Random Forest	KNN	Naive Bayes
<i>For “Intersection” operator</i>				
Gradient Boosting	39676	16835	16728	17655
Random Forest	16835	39730	16844	17643
KNN	16728	16844	39868	17659
Naive Bayes	17655	17643	17659	41704
<i>For “Union” operator</i>				
Gradient Boosting	39676	62571	62816	63725
Random Forest	62571	39730	62754	63791
KNN	62816	62754	39868	63913
Naive Bayes	63725	63791	63913	41704

Top-4 models by F1 score:

1. Gradient Boosting.
2. Random Forest.
3. K-Nearest Neighbors.
4. Naive Bayes.

The minimum rating of the top-4 model according to the F1 metric: 0.962. Table 3 shows the application of the “intersection” and “union” operators among the top 4 classifiers on the noise-free dataset.

The intersection of the top 2 models (Gradient Boosting and Random Forest) reduced the number of

significant features to 16835. The intersection of the top 3 models (Gradient Boosting, Random Forest and Naive Bayes) reduced the number of significant features to 7061. The intersection of the top 4 models (Gradient Boosting, Random Forest, K-Nearest Neighbors and Naive Bayes) reduced the number of significant features to 3102. Combining the top 2 models (Gradient Boosting and Random Forest) reduced the number of features to 62571. Combining the top 3 models (Gradient Boosting, Random Forest and Naive Bayes) reduced the number of features to 75928. Combining the top 4 models (Gradient Boosting, Random Forest, K-Nearest Neighbors and Naive Bayes) reduced the number of features to 83834.

The next experiment was a study of learning the MobileNetV2 neural network on the obtained datasets.

The performance of MobileNetV2 at the intersection of the top-*n* models on the dataset with noise (without transformations (2)–(6)) and without noise, with and without SMOTE balancing is shown in Table 4. These results highlight MobileNetV2’s robustness under varying noise conditions and balancing techniques, offering insights into its reliability across different data scenarios.

The performance of MobileNetV2 on the full dataset with and without noise, with and without SMOTE balancing is shown in Table 5. Number of significant features: 93752.

Table 4 – Evaluation of MobileNetV2 with intersection and union operators of top-*n* models by metrics

	Dataset without (2)–(6)				
	“Intersection” operator, size{Sfres}: 3097		“Union” operator, size{Sfres}: 83390		
Metrics	Balanced data	Unbalanced data	Balanced data	Unbalanced data	
Accuracy	0.8996	0.9071	0.9986	1.0000	
Precision	0.9012	0.8333	0.9985	1.0000	
Recall	0.8818	0.7570	0.9971	1.0000	
F1	0.9215	0.9267	1.0000	1.0000	
Training time (sec)	33.41	23.57	34.24	23.57	
n=4	Dataset with (2)–(6)				
	“Intersection” operator, size{Sfres}: 3102		“Union” operator, size{Sfres}: 83834		
	Metrics	Balanced data	Unbalanced data	Balanced data	Unbalanced data
	Accuracy	0.8996	0.9071	0.9892	0.9946
	Precision	0.9012	0.8333	0.9891	0.9892
	Recall	0.8818	0.7570	0.9884	0.9913
	F1	0.9215	0.9267	0.9898	0.9871
	Training time (sec)	34.60	24.19	34.63	24.57
	n=3	Dataset without (2)–(6)			
		“Intersection” operator, size{Sfres}: 7221		“Union” operator, size{Sfres}: 75844	
Metrics		Balanced data	Unbalanced data	Balanced data	Unbalanced data
Accuracy		0.9487	0.9579	0.9993	1.0000
Precision		0.9493	0.9099	0.9993	1.0000
Recall		0.9327	0.9801	0.9985	1.0000
F1		0.9666	0.8491	1.0000	1.0000
Training time (sec)		31.75	24.20	32.56	24.27
n=3		Dataset with (2)–(6)			
		“Intersection” operator, size{Sfres}: 7061		“Union” operator, size{Sfres}: 75928	
	Metrics	Balanced data	Unbalanced data	Balanced data	Unbalanced data
	Accuracy	0.9487	0.9546	0.9451	0.9968
	Precision	0.9473	0.9005	0.9416	0.9935
	Recall	0.9681	1.0000	0.9984	0.9957
	F1	0.9273	0.8190	0.8910	0.9914
	Training time (sec)	34.43	23.91	33.86	24.30
	n=2	Dataset without (2)–(6)			
		“Intersection” operator, size{Sfres}: 16875		“Union” operator, size{Sfres}: 62682	
Metrics		Balanced data	Unbalanced data	Balanced data	Unbalanced data
Accuracy		0.9393	0.9741	0.9884	1.0000
Precision		0.9395	0.9483	0.9886	1.0000
Recall		0.9588	1.0000	0.9943	1.0000
F1		0.9209	0.9016	0.9831	1.0000
Training time (sec)		34.57	23.96	32.30	23.23
n=2		Dataset with (2)–(6)			
		“Intersection” operator, size{Sfres}: 16835		“Union” operator, size{Sfres}: 62571	
	Metrics	Balanced data	Unbalanced data	Balanced data	Unbalanced data
	Accuracy	0.8844	0.9526	0.9566	0.9569
	Precision	0.8750	0.9076	0.9560	0.9123
	Recall	0.9790	0.9310	0.9939	0.9811
	F1	0.7910	0.8852	0.9209	0.8525
	Training time (sec)	33.10	24.42	34.77	24.06

An experiment was also conducted to study the effectiveness of the dimensionality reduction method with the smallest number of features obtained – 3102 with noise removal ($n=4$, classifiers: Gradient Boosting, Random Forest, K-Nearest Neighbors, Naive Bayes)). The experimental data are shown in Table 6.

All experiments were conducted on a system equipped with an Intel Core i7 processor and 16 GB of RAM, supported by a 32 GB swap file to manage memory-intensive operations. The software environment included Python version 3.10.15, utilizing libraries such as sklearn, imblearn, tensorflow, keras, lime, flask for data processing, modeling, and web functionalities, and Gnuplot 6.0.0, all running on Linux Mint 22 Xfce OS (kernel version 6.8.0–38-generic).

Table 5 – Evaluation of MobileNetV2, noisy and noise-free dataset by metrics

Dataset without transformations (2)–(6)		
Metrics	Balanced data	Unbalanced data
Accuracy	0.9942	1.0000
Precision	0.9944	1.0000
Recall	0.9888	1.0000
F1	1.0000	1.0000
Training time (sec)	35.15	25.70
Dataset with transformations (2)–(6)		
Accuracy	0.9566	0.9612
Precision	0.9575	0.9280
Recall	0.9602	0.9062
F1	0.9548	0.9508
Training time (sec)	35.18	24.73

Table 6 – Evaluation of classifiers from *MLset* by metrics with the number of significant features 3102

Metrics	Gradient Boosting	Random Forest	Naive Bayes	KNN
Original dataset				
Accuracy	0.99	0.98	0.96	0.97
Precision	0.99	0.98	0.96	0.97
Recall	0.99	0.98	0.96	0.97
F1	0.99	0.98	0.96	0.97
Training time (sec)	115	<1	<1	<1

The “Discussion” section details the obtained data from Tables 1–6.

6 DISCUSSION

As can be seen from Table 1, before applying transformations (2)–(6), the metrics show results of less than 83% on all metrics for all models from the *MLset*. The worst results are observed for the KNN model. This is due to the susceptibility of KNN to problems in high-dimensional spaces where the distance between points becomes less informative.

After applying transformations (2)–(6), the metrics significantly improved their performance for all classifiers from the *MLset* set. After denoising for the KNN classifier, the distance between the points became more

informative, which allowed us to obtain metrics at the level of 97%.

However, taking into account the specificity of the problem of defect classification by vibration analysis, the obtained accuracy indicator is not a benchmark and needs improvement.

It is known, that neural networks better track complex dependencies, therefore the MobileNetV2 neural network was chosen as a classifier, which was studied on the obtained data sets by applying the optional operators “intersection” and “union” of significant features selected using LIME models without transformations (2)–(6) (table 2) and with their application (table 3).

The smallest number of features obtained using the optional intersection operator is 3097 without denoising and 3102 with denoising, which is a 30.27-fold reduction in dimensionality.

From the conducted experiment with $n=4$ for classifiers from the *MLset* (table 6), it was observed that reducing the dimensionality by more than 30 times does not worsen the result, but on the contrary, for most classifiers it improves the values of the metrics (compared to the results of table 1). Gradient Boosting managed to improve all metrics to 0.99, while reducing training time from 3643 seconds to 115 (31.67 times).

However, this reduction in dimensionality leads to a loss of accuracy of the MobileNetV2 neural network classifier (table 4).

The corresponding results of the experiment are shown in Table 4. From which it can be seen that, compared to the best F1 indicators of 0.99 for Gradient Boosting, it was possible to raise not only the metrics to 1 on the validation and training data, but also to reduce the training time to 23.23 seconds (compared to 115 seconds).

Such metric results were achieved when applying the optional “union” operator for $n=4$, 3 and 2 in the absence of SMOTE-balancing on the dataset without applying transformations (2)–(6). This is explained by the fact that the MobileNetV2 neural network is capable of reducing the dimensionality and determining meaningful data. However, with $n=2$ for the “union” operator, the number of significant features was reduced to 62,682 (compared to the initial number of 93,752), which in turn contributed to the reduction of the learning time to 23.23 (table 5), compared to 25.70 with the initial dataset in 93752 features.

Compared with analogues, it was possible to improve the accuracy from 0.989 in [6] and from 0.996 in [7] to 1.

CONCLUSIONS

The actual problem of effective intelligent diagnostics of malfunctions of rotating equipment is solved.

The scientific novelty consists in creating a method of reducing the dimensionality of data by applying a complex approach that combines:

- cleaning from noise (transformation (2)–(6)), allows to save important information, avoiding redundancy;
- use of the LIME method on *MLset* to determine significant features, which allows to obtain interpreted re-

sults and ensure greater reliability in determining important features;

– the use of optional operators, which provides flexibility in choosing an approach to determining important features, which can be useful in different contexts and for different types of data, especially in large datasets with a large number of features.

The proposed approach differs from the existing ones by the possibility of using optional operators of union and intersection when forming a set of significant features, which provides flexibility and allows to adapt to different contexts and types of data, ensuring the efficiency of classification in cases of large-dimensional data.

On the basis of the created datasets of reduced dimension, a neural network classifier was built, which reached the values of 1 according to the Accuracy, Precision, Recall and F1 metrics.

The practical significance is that software has been developed that implements the proposed approach to data dimensionality reduction and defect classification based on vibration analysis for maintenance of rotating machines. The results of the experiments show that the developed approach is effective and allows reducing the dimensionality of the data by more than 30 times without losing classification accuracy.

Prospects for further research will be aimed at finding alternative neural network architectures to reduce training time.

ACKNOWLEDGEMENTS

The work was carried out as part of the All-Ukrainian competition-hackathon of scientific works of young scientists in the field of intellectual information technologies in 2024 at Zaporizhzhya National Technical University.

REFERENCES

1. Matania O., Bachar L., Bechhoefer E., Bortman J. Signal Processing for the Condition-Based Maintenance of Rotating Machines via Vibration Analysis: A Tutorial, *Sensors*, 2024, Vol. 24, No. 2, P. 454. DOI: 10.3390/s24020454.
2. Mongia C., Goyal D., Sehgal S. Vibration response-based condition monitoring and fault diagnosis of rotary machinery, *Materials Today: Proceedings*, 2022, Vol. 50, pp. 679–683. DOI: 10.1016/j.matpr.2021.04.395.
3. Tama B. A., Vania M., Lee S., et al. Recent advances in the application of deep learning for fault diagnosis of rotating machinery using vibration signals, *Artificial Intelligence Review*, 2023, Vol. 56, pp. 4667–4709. DOI: 10.1007/s10462-022-10293-3.
4. Nata C., Laurence N., Hartono N., Cahyadi L. Implementation of Condition-based and Predictive-based Maintenance using Vibration Analysis, *2021 4th International Conference of Computer and Informatics Engineering (IC2IE)*, 2021, pp. 90–95. DOI: 10.1109/IC2IE53219.2021.9649400.
5. Iqbal M., Madan A. K., Ahmad N. Vibration and acoustic signal-based bearing fault diagnosis in CNC machine using an improved deep learning, *Iran Journal of Computer Science*, 2024. DOI: 10.1007/s42044-024-00205-9.
6. Althubaiti A., Elasha F., Teixeira J.A. Fault diagnosis and health management of bearings in rotating equipment based on vibration analysis, *Journal of Vibroengineering*, 2022, Vol. 24, No. 1, pp. 46–74. DOI: 10.21595/jve.2021.22100.
7. Souza R.M., Nascimento E.G.S., Miranda U.A., Silva W.J.D., Lepikson H.A. Deep learning for diagnosis and classification of faults in industrial rotating machinery, *Computers & Industrial Engineering*, 2021, Vol. 153, P. 107060. DOI: 10.1016/j.cie.2020.107060.
8. Gu J., Peng Y., Lu H., Chang X., Chen G. A novel fault diagnosis method of rotating machinery via VMD, CWT and improved CNN, *Measurement*, 2022, Vol. 200, P. 111635. DOI: 10.1016/j.measurement.2022.111635.
9. Di Maggio L.G., Brusa E., Delprete C. Zero-Shot Generative AI for Rotating Machinery Fault Diagnosis: Synthesizing Highly Realistic Training Data via Cycle-Consistent Adversarial Networks, *Applied Sciences*, 2023, Vol. 13, No. 22, P. 12458. DOI: 10.3390/app132212458.
10. Ma C., Wang X., Li Y., Cai Z. Broad zero-shot diagnosis for rotating machinery with untrained compound faults, *Reliability Engineering & System Safety*, 2024, Vol. 242, P. 109704. DOI: 10.1016/j.ress.2023.109704.
11. Liang R., Ran W., Chen Y. et al. Fault Diagnosis Method for Rotating Machinery Based on Multi-scale Features, *Chin. J. Mech. Eng.*, 2023, Vol. 36, P. 141. DOI: 10.1186/s10033-023-00966-7.
12. Fast Fourier Transform (FFT) [Electronic resource]. Access mode: <https://pythonnumericalmethods.studentorg.berkeley.edu/notes/ebooks/chapter24.03-Fast-Fourier-Transform.html>.
13. Krak I., Zalutka O., Molchanova M., Mazurets O., Manziuk E., Barmak O. Method for neural network detecting propaganda techniques by markers with visual analytic, *CEUR Workshop Proceedings*, 2024, Vol. 3790, pp. 158–170 [Electronic resource]. Access mode: <https://ceur-ws.org/Vol-3790/paper14.pdf>.
14. Krak I., Zalutka O., Molchanova M., Mazurets O., Bahrii R., Sobko O., Barmak O. Abusive Speech Detection Method for Ukrainian Language Used Recurrent Neural Network, *CEUR Workshop Proceedings*, 2023, Vol. 3387, pp. 16–28. DOI: 10.31110/COLINS/2024-3/002.
15. Hasan M.M. Understanding Model Predictions: A Comparative Analysis of SHAP and LIME on Various ML Algorithms, *Journal of Scientific and Technological Research*, 2024, Vol. 5, No. 1, pp. 17–26. DOI: 10.59738/jstr.v5i1.23(17–26).eaqr5800.
16. Baniya A., Vinod A., Chinta H.S., Vishnu P., T.S. Intrusion Detection based on clustering with ML algorithm and LIME Insights, *2024 International Conference on Integrated Circuits and Communication Systems (ICICACS)*, Raichur, India, 2024, pp. 1–6. DOI: 10.1109/ICICACS60521.2024.10498512.
17. Pham M.T., Kim J.-M., Kim C.H. Deep Learning-Based Bearing Fault Diagnosis Method for Embedded Systems, *Sensors*, 2020, Vol. 20, pp. 6886. DOI: 10.3390/s20236886.
18. Slobodzian V., Molchanova M., Kovalchuk O., Sobko O., Mazurets O., Barmak O., Krak I. An Approach Based on the Visualization Model for the Ukrainian Web Content Classification, *2022 12th International Conference on Advanced Computer Information Technologies (ACIT)*, 2022, pp. 400–405. DOI: 10.1109/ACIT54803.2022.9913162.
19. Scikit-learn: machine learning in Python – Scikit-learn 1.6.0 documentation [Electronic resource], Scikit-learn: machine learning in Python – Scikit-learn 0.16.1

- documentation. Mode of access: <https://scikit-learn.org/stable/>.
20. Imbalanced-learn documentation – Version 0.13.0 [Electronic resource], *Welcome to imbalanced-learn documentation*. Mode of access: <https://imbalanced-learn.org/stable/>.
21. Local Interpretable Model-Agnostic Explanations (lime) – lime 0.1 documentation [Electronic resource], *Local Interpretable Model-Agnostic Explanations (lime) – lime 0.1 documentation*. Mode of access: <https://lime-ml.readthedocs.io/en/latest/>.
22. SciPy [Electronic resource], *SciPy*. Mode of access: <https://scipy.org/>.
23. TensorFlow [Electronic resource], *TensorFlow*. Mode of access: <https://www.tensorflow.org>.
24. NumPy [Electronic resource], *NumPy*. Mode of access: <https://numpy.org/>.
25. Matplotlib – Visualization with Python [Electronic resource], *Matplotlib – Visualization with Python*. Mode of access: <https://matplotlib.org/>.
26. Gnuplot homepage [Electronic resource], *Gnuplot homepage*. Mode of access: <https://gnuplot.sourceforge.net/>.
- Received 02.12.2024.
Accepted 24.01.2025.

УДК 004.8

ПІДХІД ДО ЗМЕНШЕННЯ РОЗМІРНОСТІ ДАНИХ ТА КЛАСИФІКАЦІЇ ДЕФЕКТІВ ЗА АНАЛІЗОМ ВІБРАЦІЙ ДЛЯ ТЕХНІЧНОГО ОБСЛУГОВУВАННЯ ОБЕРТОВИХ МАШИН

Молчанова М. О. – аспірант кафедри комп'ютерних наук Хмельницького національного університету, м. Хмельницький, Україна.

Дідур В. О. – студент кафедри комп'ютерних наук Хмельницького національного університету, м. Хмельницький, Україна.

Мазурець О. В. – канд. техн. наук, доцент кафедри комп'ютерних наук Хмельницького національного університету, м. Хмельницький, Україна.

АНОТАЦІЯ

Актуальність. Вирішується актуальна проблема ефективної інтелектуальної діагностики несправностей обертового обладнання. Об'єктом дослідження є процес зменшення розмірності даних та класифікації дефектів за аналізом вібрацій для технічного обслуговування обертових машин. Предметом дослідження є методи зменшення розмірності та класифікації дефектів за аналізом вібрацій.

Мета роботи – створення підходу до зменшення розмірності даних та класифікації дефектів за аналізом вібрацій для технічного обслуговування обертових машин

Метод. Запропоновано комплексний підхід до зменшення розмірності даних та класифікації дефектів за аналізом вібрацій, що вирішує задачу зменшення розмірності даних для навчання класифікаторів та класифікації дефектів, а також вирішено задачу побудови нейромережевого класифікатора, що спроможний забезпечити швидкість класифікації несправностей без втрати точності на даних зменшеної розмірності. Підхід відрізняється від існуючих можливістю застосування опційних операторів об'єднання та перетину при формуванні множини значущих ознак, що надає гнучкість та дозволяє адаптуватись до різних контекстів та типів даних, забезпечуючи ефективність класифікації у випадках даних великої розмірності.

Метод для очищення від шуму дозволяє зберегти важливу інформацію, уникаючи надлишковості та покращуючи якість даних для подальшого аналізу. Він передбачає обчислення співвідношення сигнал/шум, встановлення порогових значень та застосування швидкого перетворення Фур'є, що забезпечує виокремлення релевантних ознак від шумів. Використання методу LIME до множини моделей машинного навчання дозволяє визначити значущі ознаки з більшою точністю та інтерпретованістю. Це сприяє отриманню надійніших результатів, оскільки LIME допомагає зрозуміти вплив кожної ознаки на кінцеве рішення моделі, що особливо важливо при роботі з великими датасетами, де важливість окремих ознак може бути неочевидною. Впровадження опційних операторів об'єднання та перетину значущих ознак надає додаткову гнучкість у виборі підходу до визначення важливих ознак. Це дозволяє адаптувати метод до різних контекстів та типів даних, забезпечуючи ефективність навіть у випадках з великою кількістю ознак.

Результати. Розроблений метод реалізовано програмно і досліджено при вирішенні задачі класифікації дефектів за аналізом вібрацій для технічного обслуговування обертових машин.

Висновки. Проведені експериментальні дослідження підтвердили високу ефективність та працездатність запропонованого підходу для зменшення розмірності даних та класифікації дефектів за аналізом вібрацій в аспекті технічного обслуговування обертових машин. Перспективи подальших досліджень будуть направлені на пошуки альтернативних нейромережевих архітектур та їх навчання для зниження часу навчання.

КЛЮЧОВІ СЛОВА: зменшення розмірності, значущі ознаки, виявлення дефектів, MobileNetV2.

ЛІТЕРАТУРА

1. Signal Processing for the Condition-Based Maintenance of Rotating Machines via Vibration Analysis: A Tutorial / [O. Matania, L. Bachar, E. Bechhoefer, J. Bortman] // *Sensors*. – 2024. – Vol. 24, No. 2. – P. 454. DOI: 10.3390/s24020454.
2. Mongia C. Vibration response-based condition monitoring and fault diagnosis of rotary machinery / [C. Mongia, D. Goyal, S. Sehgal] // *Materials Today: Proceedings*. – 2022. – Vol. 50. – P. 679–683. DOI: 10.1016/j.matpr.2021.04.395.
3. Recent advances in the application of deep learning for fault diagnosis of rotating machinery using vibration signals / [B. A. Tama, M. Vania, S. Lee et al.] // *Artificial Intelligence Review*. – 2023. – Vol. 56. – P. 4667–4709. DOI: 10.1007/s10462-022-10293-3.
4. Implementation of Condition-based and Predictive-based Maintenance using Vibration Analysis / [C. Nata, N.

- Laurence, N. Hartono, L. Cahyadi] // 2021 4th International Conference of Computer and Informatics Engineering (IC2IE). – 2021. – P. 90–95. DOI: 10.1109/IC2IE53219.2021.9649400.
5. Iqbal M. Vibration and acoustic signal-based bearing fault diagnosis in CNC machine using an improved deep learning / M. Iqbal, A. K. Madan, N. Ahmad // *Iran Journal of Computer Science*. – 2024. DOI: 10.1007/s42044-024-00205-9.
 6. Althubaiti A. Fault diagnosis and health management of bearings in rotating equipment based on vibration analysis – a review / A. Althubaiti, F. Elasha, J. A. Teixeira // *Journal of Vibroengineering*. – 2021. – Vol. 24, No. 1. – P. 46–74 DOI: 10.21595/jve.2021.22100.
 7. Deep learning for diagnosis and classification of faults in industrial rotating machinery / [R. M. Souza, E. G. S. Nascimento, U. A. Miranda et al.] // *Computers & Industrial Engineering*. – 2021. – Vol. 153. – P. 107060. DOI: 10.1016/j.cie.2020.107060.
 8. A novel fault diagnosis method of rotating machinery via VMD, CWT and improved CNN / [J. Gu, Y. Peng, H. Lu et al.] // *Measurement*. – 2022. – Vol. 200. – P. 111635. DOI: 10.1016/j.measurement.2022.111635.
 9. Di Maggio L.G. Zero-Shot Generative AI for Rotating Machinery Fault Diagnosis: Synthesizing Highly Realistic Training Data via Cycle-Consistent Adversarial Networks / L.G. Di Maggio, E. Brusa, C. Delprete // *Applied Sciences*. – 2023. – Vol. 13, № 22. – P. 12458. DOI: 10.3390/app132212458
 10. Ma C. Broad zero-shot diagnosis for rotating machinery with untrained compound faults / C. Ma, X. Wang, Y. Li, Z. Cai // *Reliability Engineering & System Safety*. – 2024. – Vol. 242. – P. 109704. DOI: 10.1016/j.res.2023.109704.
 11. Fault Diagnosis Method for Rotating Machinery Based on Multi-scale Features / [R. Liang, W. Ran, Y. Chen et al.] // *Chinese Journal of Mechanical Engineering*. – 2023. – Vol. 36. – P. 141. DOI: 10.1186/s10033-023-00966-7.
 12. Fast Fourier Transform (FFT) – Python Numerical Methods [Electronic resource]. – Mode of access: <https://pythonnumericalmethods.studentorg.berkeley.edu/notes/ebooks/chapter24.03-Fast-Fourier-Transform.html>.
 13. Method for neural network detecting propaganda techniques by markers with visual analytic / [I. Krak, O. Zalutskya, M. Molchanova et al.] // *CEUR Workshop Proceedings*. – 2024. – Vol. 3790. – P. 158–170 [Electronic resource]. – Access mode: <https://ceur-ws.org/Vol-3790/paper14.pdf>.
 14. Abusive Speech Detection Method for Ukrainian Language Used Recurrent Neural Network / [I. Krak, O. Zalutskya, M. Molchanova et al.] // *CEUR Workshop Proceedings*. – 2023. – Vol. 3387. – P. 16–28. DOI: 10.31110/COLINS/2024-3/002.
 15. Hasan M. M. Understanding Model Predictions: A Comparative Analysis of SHAP and LIME on Various ML Algorithms / M. M. Hasan // *Journal of Scientific and Technological Research*. – 2024. – Vol. 5, No. 1. – P. 17–26. DOI: 10.59738/jstr.v5i1.23(17–26).eaqr5800.
 16. Intrusion Detection based on clustering with ML algorithm and LIME Insights / [A. Baniya et al.] // *2024 International Conference on Integrated Circuits and Communication Systems (ICICACS)*. – Raichur, India, 2024. – P. 1–6. DOI: 10.1109/ICICACS60521.2024.10498512.
 17. Pham M. T. Deep Learning-Based Bearing Fault Diagnosis Method for Embedded Systems / M. T. Pham, J.-M. Kim, C. H. Kim // *Sensors*. – 2020. – Vol. 20, No. 23. – P. 6886. DOI: 10.3390/s20236886.
 18. An Approach Based on the Visualization Model for the Ukrainian Web Content Classification / [V. Slobodzian, M. Molchanova et al.] // *2022 12th International Conference on Advanced Computer Information Technologies (ACIT)*. – 2022. – P. 400–405. DOI: 10.1109/ACIT54803.2022.9913162.
 19. Scikit-learn: machine learning in Python – Scikit-learn 1.6.0 documentation [Electronic resource] // Scikit-learn: machine learning in Python – Scikit-learn 0.16.1 documentation. – Mode of access: <https://scikit-learn.org/stable/>.
 20. Imbalanced-learn documentation – Version 0.13.0 [Electronic resource] // Welcome to imbalanced-learn documentation. – Mode of access: <https://imbalanced-learn.org/stable/>.
 21. Local Interpretable Model-Agnostic Explanations (lime) – lime 0.1 documentation [Electronic resource] // Local Interpretable Model-Agnostic Explanations (lime) – lime 0.1 documentation. – Mode of access: <https://lime-ml.readthedocs.io/en/latest/>.
 22. SciPy [Electronic resource] // SciPy. – Mode of access: <https://scipy.org/>.
 23. TensorFlow [Electronic resource] // TensorFlow. – Mode of access: <https://www.tensorflow.org>.
 24. NumPy [Electronic resource] // NumPy – Mode of access: <https://numpy.org/>.
 25. Matplotlib – Visualization with Python [Electronic resource] // Matplotlib – Visualization with Python. – Mode of access: <https://matplotlib.org/>.
 26. Gnuplot homepage [Electronic resource] // Gnuplot homepage. – Mode of access: <https://gnuplot.sourceforge.net/>.

KEYSTROKE DYNAMICS RECOGNITION USING NINE-VARIATE PREDICTION ELLIPSOID FOR NORMALIZED DATA

Prykhodko S. B. – Dr. Sc., Professor, Head of the Department of Software for Automated Systems, Admiral Makarov National University of Shipbuilding, Mykolaiv, Ukraine.

Trukhov A. S. – Post-graduate student of the Department of Software for Automated Systems, Admiral Makarov National University of Shipbuilding, Mykolaiv, Ukraine.

ABSTRACT

Context. Keystroke dynamics recognition is a crucial element in enhancing security, enabling personalized user authentication, and supporting various identity verification systems. This study investigates the influence of data distribution on the performance of one-class classification models in keystroke dynamics, focusing on the application of a nine-variate prediction ellipsoid. The object of research is the keystroke dynamics recognition process. The subject of the research is a mathematical model for keystroke dynamics recognition. Unlike typical approaches assuming a multivariate normal distribution of data, real-world keystroke datasets often exhibit non-Gaussian distributions, complicating model accuracy and robustness. To address this, the dataset underwent normalization using the multivariate Box-Cox transformation, allowing the construction of a more precise decision boundary based on the prediction ellipsoid for normalized data.

The objective of the work is to increase the probability of keystroke dynamics recognition by constructing a nine-variate prediction ellipsoid for normalized data using the Box-Cox transformation.

Method. This research involves constructing a nine-variate prediction ellipsoid for data normalized using the Box-Cox transformation to improve keystroke dynamics recognition. The squared Mahalanobis distance is applied to identify and remove outliers, while the Mardia test assesses deviations from normality in the multivariate distribution. Estimates for parameters of multivariate Box-Cox transformation are derived using the maximum likelihood method.

Results. The results demonstrate significant performance improvements after normalization, reaching higher accuracy and robustness compared to models built for non-normalized data. The application of the nine-variate Box-Cox transformation successfully accounted for feature correlations, enabling the prediction ellipsoid to better capture underlying data patterns.

Conclusions. For keystroke dynamics recognition, a mathematical model in the form of the nine-variate prediction ellipsoid for data normalized using the multivariate Box-Cox transformation has been developed, which enhances the probability of recognition compared to models constructed for non-normalized data. However, challenges remain in determining the optimal normalization technique and selecting the significance level for constructing the prediction ellipsoid. These findings underscore the importance of careful feature selection and advanced data normalization techniques for further research in keystroke dynamics recognition.

KEYWORDS: keystroke dynamics recognition, multivariate Box-Cox transformation, prediction ellipsoid, normalizing transformation.

ABBREVIATIONS

BCT is the Box-Cox transformation;
SMD is the squared Mahalanobis distance;
TP is true positives;
FP is false positives;
TN is true negatives;
FN is false negatives;
NGD is a prediction ellipsoid for non-Gaussian data;
ND is a prediction ellipsoid for normalized data.

NOMENCLATURE

p is a number of variables;
 m is a number of degrees of freedom;
 N is a number of data points;
 S_X is a sample covariance matrix for initial data;
 S_Z is a sample covariance matrix for normalized data;
 X is a non-Gaussian random vector;
 \bar{X} is a vector of sample means of the X_j variables;
 X_j is a j -th non-Gaussian variable;
 Z is a Gaussian random vector;
 \bar{Z} is a vector of sample means of the Z_j variables;

Z_j is a j -th Gaussian variable that is obtained by transforming the variable;

α is a significance level;

β_1 is a multivariate skewness;

β_2 is a multivariate kurtosis;

$\chi_{m,\alpha}^2$ is the chi-square distribution quantile with m degrees of freedom and significance level α ;

Ψ is a vector of multivariate normalizing transformation.

INTRODUCTION

In recent years, keystroke dynamics, also known as keystroke biometrics or typing biometrics, has emerged as a viable method for biometric authentication. It leverages the unique patterns and rhythms individuals exhibit while typing on a keyboard, capturing characteristics such as keystroke duration and inter-key intervals [1]. These features enable the creation of a distinctive typing profile for each user.

Unlike traditional biometric methods such as fingerprints or facial recognition, keystroke dynamics offers a non-intrusive and continuous form of user authentication

[2]. This makes it particularly attractive for applications in online banking, secure login systems, and access control. A low probability of authenticating individuals can have negative consequences in terms of security and personalization. Therefore, there is a need to develop and improve keystroke dynamics recognition methods.

A significant limitation of existing methods in keystroke dynamics recognition is their sensitivity to non-normal data distribution. Many statistical and machine learning techniques assume multivariate normality in the underlying data, which is often not the case in real-world keystroke dynamics datasets [4]. Non-normal distributions can adversely affect the performance of these techniques, leading to suboptimal recognition results. Therefore, it is necessary to enhance mathematical models to accommodate deviations from the multivariate normal distribution of data.

The object of study is the process of keystroke dynamics recognition.

The keystroke dynamics recognition process includes several important steps to ensure accurate and reliable user authentication. First, a suitable dataset containing keystroke data, such as keypress and release times, is identified. From this raw data, characteristics such as key hold time and key spacing are extracted to represent the unique typing patterns of individuals. Next, outlier detection is performed to identify and remove anomalous data points that may skew the analysis [3]. This step is crucial for refining the dataset and improving the accuracy of the model. Finally, classification is carried out, with the accuracy and efficiency largely depending on the specific model.

The subject of study is a mathematical model for keystroke dynamics recognition. One of the frequently employed methods in pattern recognition involves building decision rules based on prediction ellipsoids.

The purpose of the work is to increase the probability of keystroke dynamics recognition by constructing a nine-variate prediction ellipsoid for normalized data using the multivariate Box-Cox transformation (BCT).

1 PROBLEM STATEMENT

Assume we have an original data sample set consisting of nine keystroke timing features, with a multivariate distribution that is not Gaussian. In this case, there exists a bijective nine-variate normalizing transformation $\Psi = \{\psi_Y, \psi_1, \psi_2, \dots, \psi_9\}^T$ that converts the non-Gaussian random vector $\mathbf{X} = \{X_1, X_2, \dots, X_9\}^T$ into a Gaussian random vector $\mathbf{Z} = \{Z_1, Z_2, \dots, Z_9\}^T$ is given by:

$$\mathbf{Z} = \Psi(\mathbf{X}). \quad (1)$$

It is required to build the prediction ellipsoid for normalized data in the form:

$$(\mathbf{z} - \bar{\mathbf{z}})^T \mathbf{S}_Z^{-1} (\mathbf{z} - \bar{\mathbf{z}}) = \chi_{m,\alpha}^2, \quad (2)$$

where

$$\mathbf{S}_Z = \frac{1}{N} \sum_{i=1}^N (\mathbf{z}_i - \bar{\mathbf{z}})(\mathbf{z}_i - \bar{\mathbf{z}})^T.$$

Also, it is required to construct a prediction ellipsoid for keystroke dynamics recognition based on equation (2) and the transformations (1).

2 REVIEW OF THE LITERATURE

Mathematical modeling methods are utilized in keystroke dynamics recognition, encompassing the process of constructing and refining mathematical models to enhance accuracy and reliability. Recent advancements in this field have employed a range of techniques. Key methods include tree-based methods, such as random forests [5, 6], build hierarchical models that classify data by learning feature splits to effectively separate different classes; support vector-based approaches [7–9], establish optimal decision boundaries by maximizing the margin between classes; neural network-based models [10–12] capture complex patterns in keystroke data through multiple layers of interconnected nodes and others classifiers.

However, in developing user authentication systems, one-class classification is more typical for keystroke dynamics because it focuses on verifying whether a typing pattern belongs to a known user [13]. The primary goal is to create a model based on the unique typing patterns of an individual, and the target class, and then identify whether new typing patterns match this known profile. This is especially useful for authentication, where the system needs to continuously confirm that the current user is indeed the registered individual, rather than distinguishing between multiple users. One-class classification is connected with outlier detection [14]. The model is trained on data from a target class and does not have explicit knowledge of other classes. It identifies whether new data fits within the learned target class pattern, flagging deviations as potential outliers [15].

Popular approaches to one-class classification in keystroke dynamics recognition include one-class support vector machine [16–18], which identifies a boundary that separates target data from outliers; neural network-based classifiers such as autoencoders [19, 20], which learn to reconstruct input data and identify anomalies based on reconstruction errors, and GANs [21], which use generative adversarial networks to model target data and detect deviations. Currently, the mathematical modeling of prediction ellipsoids is widely used for pattern recognition, particularly in one-class classification systems [22, 23]. This technique employs statistical methods to define a multivariate ellipsoid that encompasses the target data points.

The construction of the ellipsoid is based on the assumption that the data follows a multivariate normal distribution. One promising solution to address this limitation is the application of normalizing transformations [24–26]. In the study [27] it was observed that the prediction ellipsoid for normalized data outperformed machine

learning methods such as one-class SVM, isolation forest, and autoencoder in one-class classification for face recognition tasks. We propose applying this mathematical model to keystroke dynamics recognition systems.

Normalization techniques can convert non-normal data into a distribution that more closely approximates multivariate Gaussian distribution, thereby improving the effectiveness of statistical analyses and machine learning models. While univariate normalization techniques transform each feature independently, multivariate normalization techniques consider the relationships between multiple features simultaneously. In this study, we employ the multivariate BCT to normalize keystroke dynamics data, aiming to enhance the overall recognition performance.

3 MATERIALS AND METHODS

In the context of keystroke dynamics-based recognition, the dataset used plays a pivotal role in the effectiveness and accuracy of the algorithms. A keystroke dynamics dataset typically consists of detailed records of an individual's typing behavior, capturing various temporal aspects of keystrokes such as the duration of key presses and the intervals between key presses and releases.

The CMU keystroke dynamics dataset used in this study provides a detailed record of typing from 51 subjects, each of whom typed a static password string: "tie5Roanl". The dataset includes various keystroke timing features, recorded in seconds, such as the duration a key is held down and the time between key presses.

Data collection involved eight separate sessions per subject, with at least one day between each session. During each session, subjects typed the password 50 times, resulting in 400 repetitions per subject. This setup provides a comprehensive dataset with a total of 20,400 samples across all subjects.

The dataset is structured with a subject identifier, session number, repetition number, and a total of 31 timing features. It includes columns with specific naming conventions to represent various keystroke timing metrics. Columns labeled as H.key indicate the hold time for a particular key, capturing the duration from when the key is pressed to when it is released. DD.key1.key2 columns represent the time interval between pressing two consecutive keys, known as keydown-keydown time. Similarly, UD.key1.key2 columns denote the keyup-keydown time, which measures the interval between releasing one key and pressing the next. It is important to note that UD times can sometimes be negative, and the sum of H times and UD times corresponds to the DD time for a given digraph.

To simplify the model, it was decided to focus on 9 key properties, so the feature vector takes the form: $X = \{H.t, H.i, H.e, H.5, H.R, H.o, H.a, H.n, H.l\}$.

After extracting feature vectors, the next crucial step is outlier detection. Identifying and removing anomalies is essential because they can significantly skew the analysis and degrade the accuracy of the recognition model. By detecting outliers, the dataset is refined, ensuring that the

model is trained on data that truly represents typical user behavior.

One popular method for anomaly detection is based on the squared Mahalanobis distance (SMD). A known limitation of the SMD method is its reliance on the assumption that the data follows a multivariate Gaussian distribution. To address this, it is essential to first evaluate the data's normality using statistical tests such as the Mardia test.

The Mardia test is employed to assess whether a dataset adheres to multivariate normality, which is crucial for methods assuming a multivariate Gaussian distribution. This test also applied in the study [27], evaluates two key dimensions of normality: multivariate skewness β_1 and multivariate kurtosis β_2 .

Skewness measures the asymmetry of the data distribution. When skewness is scaled by $N/6$, it follows a chi-square distribution with $p(p+1)(p+2)/6$ degrees of freedom, where p is the number of variables and N is the sample size.

Kurtosis assesses the tailedness of the distribution or the extent to which the tails differ from those of a normal distribution. The Mardia test calculates kurtosis coefficients and compares them to a normal distribution with a mean of $p(p+2)$ and a variance of $8p(p+2)/N$.

If the data deviates significantly from multivariate normality, it is imperative to apply a normalizing transformation (1) on a non-Gaussian random vector $\mathbf{X} = \{X_1, X_2, \dots, X_9\}^T$ to convert it into a Gaussian random vector $\mathbf{Z} = \{Z_1, Z_2, \dots, Z_9\}^T$.

Normalizing transformations are essential tools in statistical analysis and machine learning, primarily because they help stabilize variance, reduce skewness, and align data more closely with a multivariate Gaussian distribution. This process is particularly critical for methods that rely on the assumption of multivariate normality, as such transformations can greatly enhance the validity and interpretability of the results. There are two main types of normalizing transformations: univariate and multivariate.

Univariate transformations include methods like the logarithm and the univariate BCT. The logarithm is commonly employed to stabilize variance, especially in data with positive skewness. On the other hand, the univariate BCT offers greater flexibility by addressing both positive and negative skewness through the selection of an optimal parameter λ . However, this method can be more complex to implement due to the need for parameter estimation, which may also be sensitive to outliers.

In contrast, multivariate transformations, such as the multivariate BCT, overcome the limitations of univariate methods by considering the interrelationships between multiple variables. While univariate transformations are more straightforward, they can fall short in situations where the interdependencies among variables are significant. Multivariate transformations provide a more

comprehensive approach but come with increased computational complexity.

The multivariate BCT builds on the principles of the univariate BCT approach but extends its application to multiple variables simultaneously.

$$Z_j = x(\lambda_j) = \begin{cases} \left(\frac{X_j^{\lambda_j} - 1}{\ln(X_j)} \right) / \lambda_j, & \lambda_j \neq 0; \\ \ln(X_j), & \lambda_j = 0. \end{cases} \quad (3)$$

This transformation maintains the correlations between variables while normalizing their distributions, making it especially effective for multivariate data. However, applying this transformation demands considerable effort due to the complexity of the required parameter estimation. A widely used approach for estimating parameters for each feature in a set involves maximizing the log-likelihood of the transformed data, as described in the study [27].

Applying the multivariate BCT can greatly improve the approximation of the data's distribution to normality. After normalization, it's crucial to run the Mardia test again to assess the transformation's success. If the test confirms that the data now follows a multivariate normal distribution, the dataset is ready for further analysis. If not, further adjustments or alternative methods may be necessary to meet the assumptions required for subsequent statistical procedures.

The next step is the construction of the prediction ellipsoid. A prediction ellipsoid is a tool in multivariate analysis used to determine whether a data point belongs to a specific target class. This method involves calculating the squared Mahalanobis distance for each point, which corresponds to the left side of the comparison. This distance is then compared against a critical value from the chi-square distribution, representing the right side of the equation:

$$(\mathbf{x} - \bar{\mathbf{x}})^T \mathbf{S}_{\mathbf{x}}^{-1} (\mathbf{x} - \bar{\mathbf{x}}) = \chi_{9, 0.005}^2, \quad (4)$$

where

$$\mathbf{S}_{\mathbf{x}} = \frac{1}{N} \sum_{i=1}^N (\mathbf{x}_i - \bar{\mathbf{x}})(\mathbf{x}_i - \bar{\mathbf{x}})^T.$$

The SMD follows a chi-square distribution with degrees of freedom equal to the number of features in the data, in our case is 9. This connection allows the determination of a critical value based on the desired significance level, for one-class classification tasks, a common choice is 0.005. Specifically, if a data point's SMD exceeds the critical value from the chi-square distribution, it is classified as an anomaly (an instance of another class). Conversely, if the distance is below the critical value, the data point is classified as an instance of the target class.

In cases where the data is non-normal, following the normalization process, a nine-variate prediction ellipsoid is constructed based on (4):

$$(\mathbf{z} - \bar{\mathbf{z}})^T \mathbf{S}_{\mathbf{z}}^{-1} (\mathbf{z} - \bar{\mathbf{z}}) = \chi_{9, 0.005}^2. \quad (5)$$

The chi-square distribution quantile value is 23.59 for 9 degrees of freedom at a significance level of 0.005. If the SMD falls below the chi-square critical value, the point is considered to lie within the ellipsoid, indicating its membership in the target class.

4 EVALUATION METRICS

In one-class classification, where the goal is to distinguish between target and anomaly instances, evaluation metrics such as accuracy, specificity, precision, recall, and the F1 score play an important role in evaluating model performance [28]. These metrics are derived from classification outcomes categorized as true positives (TP), false positives (FP), true negatives (TN), and false negatives (FN).

In this context, true positives (TP) indicate correctly identified anomalies, while false positives (FP) represent cases mistakenly classified as anomalies or false alarms. True negatives (TN) denote correctly identified target instances, and false negatives (FN) reflect cases where actual anomalies were incorrectly classified as targets.

Accuracy measures the overall correctness of the classification, considering both target instances and anomalies: $\text{Accuracy} = (\text{TP} + \text{TN}) / (\text{TP} + \text{TN} + \text{FP} + \text{FN})$.

Specificity evaluates the model's capacity to correctly classify target instances. It measures the proportion of identified target instances relative to all target instances: $\text{Specificity} = \text{TN} / (\text{TN} + \text{FP})$.

Precision focuses on the reliability of the model when identifying anomalies, showing the proportion of true anomalies among all instances that the model classified as anomalies: $\text{Precision} = \text{TP} / (\text{TP} + \text{FP})$.

Recall (sensitivity) assesses the model's ability to detect all actual anomalies, it measures the proportion of true anomalies that were correctly identified out of all existing anomalies: $\text{Recall} = \text{TP} / (\text{TP} + \text{FN})$.

The F1 score offers a balanced evaluation by calculating the harmonic mean between precision and recall: $\text{F1 score} = 2 * (\text{Precision} * \text{Recall}) / (\text{Precision} + \text{Recall})$.

5 EXPERIMENTS

Data with the identifier s015 was randomly selected to construct the prediction ellipsoid. Meanwhile, s004 was employed in testing to verify the recognition of keystroke dynamics from a different individual.

It is essential to identify and remove outliers, which requires verifying whether the data follows a multivariate normal distribution. The Mardia test indicated that the set with the identifier s015 shows deviations from multivariate normality, as the test statistic for multivariate skewness $N\beta_1/6$ measuring 391.54, exceeds the chi-square distribution threshold of 215.53 for 165 degrees of freedom

at a significance level of 0.005. Similarly, the test statistic for multivariate kurtosis β_2 with a value of 113.32, surpasses the normal distribution quantile of 102.62, given a mean of 99, variance of 1.98, and 0.005 significance level. Suggesting that further normalization is needed before proceeding with analysis.

The initial step in normalization involves determining the optimal parameters. Through the application of the maximum likelihood method, the following parameter estimates were obtained: $\hat{\lambda}_1 = 0.9939$, $\hat{\lambda}_2 = 1.3605$, $\hat{\lambda}_3 = 1.2202$, $\hat{\lambda}_4 = 1.7521$, $\hat{\lambda}_5 = 2.2965$, $\hat{\lambda}_6 = 1.0447$, $\hat{\lambda}_7 = 1.6466$, $\hat{\lambda}_8 = 1.3512$, $\hat{\lambda}_9 = 2.0599$.

After applying the nine-variate BCT, the resulting set with components (3) was analyzed using the Mardia test. The test statistic for multivariate skewness $N\beta_1/6$, which is 212.07, is below the chi-square distribution's critical value of 215.53 for 165 degrees of freedom at a 0.005 significance level. However, the test statistic for multivariate kurtosis β_2 , with a value of 109.01, remains above the normal distribution quantile of 102.62, given a mean of 99, variance of 1.98, and 0.005 significance level. Despite the transformation, the set still exhibits non-normality, primarily due to the influence of outliers, that distort the distribution and prevent normalization. Nevertheless, using the normalized set is preferable, as it brings the distribution closer to multivariate normality, improving the performance of the Mahalanobis distance method.

Next, the SMD is computed for each feature vector to identify potential outliers. The calculated distances are

compared against the chi-square distribution critical value of 23.59 for 9 degrees of freedom at a significance level of 0.005. Any vectors exceeding this threshold are considered outliers. In this iteration, the vector with number 295 with the maximum SMD of 37.44 is removed from the set.

This iterative process continues until all outliers are removed from the set. After removing 6 outliers, the test statistic for multivariate kurtosis of the normalized set, according to the Mardia test, falls below the critical value. This highlights the significant impact of outliers on the set's distribution.

Table 1 presents the SMDs and corresponding indices for each outlier that was removed. This iterative process continued until no further significant outliers remained, resulting in a dataset that was more refined and less influenced by extreme values.

Table 1 – Removed outliers

№	SMD	Index	№	SMD	Index
1	37.44	295	6	26.963	323
2	36.962	160	7	26.868	45
3	30.742	306	8	25.776	263
4	28.833	388	9	24.515	294
5	28.662	214	10	23.972	204

As a result of the iterative outlier removal, the set with the vector of means $\bar{\mathbf{X}} = \{0.07525; 0.07022; 0.07823; 0.063; 0.06911; 0.08829; 0.08605; 0.07505; 0.0751\}$ was retrieved. The covariance matrix of the final set is provided in Table 2.

Table 2 – The covariance matrix of the set after removal of outliers

0.0 ³ 19	0.0 ⁴ 26	0.0 ⁴ 54	0.0 ⁴ 13	0.0 ⁵ 49	0.0 ⁴ 12	0.0 ⁴ 26	-0.0 ⁵ 6	-0.0 ⁴ 17
0.0 ⁴ 26	0.0 ³ 21	-0.0 ⁴ 19	0.0 ⁵ 51	0.0 ⁵ 82	0.0 ⁴ 14	0.0 ⁴ 4	-0.0 ⁴ 1	0.0 ⁵ 14
0.0 ⁴ 50	-0.0 ⁴ 19	0.0 ³ 25	0.0 ⁶ 1	0.0 ⁵ 35	0.0 ⁵ 11	-0.0 ⁴ 26	0.0 ⁴ 27	0.0 ⁶ 1
0.0 ⁴ 13	0.0 ⁵ 51	0.0 ⁶ 1	0.0 ³ 19	0.0 ⁴ 16	-0.0 ⁵ 52	0.0 ⁴ 31	-0.0 ⁴ 18	-0.0 ⁵ 57
0.0 ⁵ 49	0.0 ⁵ 82	0.0 ⁵ 35	0.0 ⁴ 16	0.0 ³ 13	0.0 ⁴ 14	-0.0 ⁵ 62	0.0 ⁵ 56	0.0 ⁵ 48
0.0 ⁴ 12	0.0 ⁴ 14	0.0 ⁵ 11	-0.0 ⁵ 52	0.0 ⁴ 14	0.0 ³ 12	-0.0 ⁵ 36	0.0 ⁵ 36	0.0 ⁵ 86
0.0 ⁴ 26	0.0 ⁴ 4	-0.0 ⁴ 26	0.0 ⁴ 31	-0.0 ⁵ 62	-0.0 ⁵ 36	0.0 ³ 35	-0.0 ⁴ 38	0.0 ⁵ 19
-0.0 ⁵ 6	-0.0 ⁴ 1	0.0 ⁴ 27	-0.0 ⁴ 18	0.0 ⁵ 56	0.0 ³ 36	-0.0 ⁴ 38	0.0 ⁵ 27	0.0 ⁵ 7
-0.0 ⁴ 17	0.0 ⁵ 14	0.0 ⁶ 1	-0.0 ⁵ 57	0.0 ⁵ 48	0.0 ⁵ 86	0.0 ⁵ 19	0.0 ⁵ 7	0.0 ⁵ 2

The resulting sample was randomly shuffled to ensure that the data points were distributed evenly across the training and test sets, reducing any potential bias due to the order of the data. After shuffling, the dataset was split into training and test sets, each containing 50% of the data, equating to 195 vectors per set. The training set is used to construct the prediction ellipsoid, allowing the model to learn patterns and relationships from the data. The test set, on the other hand, is used to evaluate the model's performance on a distinct subset of the data that was not seen during training. This approach helps in achieving a representative evaluation of the model's performance by providing an unbiased assessment of how well the model generalizes to new, unseen data.

Table 3 displays the covariance matrix of the training set, which has a vector of means $\bar{\mathbf{X}} = \{0.07635; 0.07052; 0.07875; 0.06254; 0.06955; 0.08806; 0.08752; 0.07447; 0.07495\}$.

Based on the Mardia test results, the multivariate distribution of the training sample shows deviations from multivariate normality. The test statistic for multivariate skewness $N\beta_1/6$ is 286.99, which exceeds the critical value of 215.53 from the chi-square distribution for 165 degrees of freedom at a 0.005 significance level. The test statistic for multivariate kurtosis β_2 is 105.43, which exceeds the critical value of 104.19 for a normal distribution with a mean of 99, a variance of 4.062, and a significance level of 0.005.

Table 3 – The covariance matrix of the training set

0.0 ³ 2	0.0 ⁴ 16	0.0 ⁴ 62	0.0 ⁴ 42	0.0 ⁵ 53	0.0 ⁵ 71	0.0 ⁴ 38	0.0 ⁵ 14	-0.0 ⁵ 75
0.0 ⁴ 16	0.0 ³ 21	-0.0 ⁴ 19	-0.0 ⁵ 26	-0.0 ⁶ 35	0.0 ⁴ 14	0.0 ⁴ 62	-0.0 ⁴ 1	0.0 ⁵ 15
0.0 ⁴ 62	-0.0 ⁴ 19	0.0 ³ 26	-0.0 ⁶ 84	0.0 ⁴ 11	0.0 ⁵ 71	-0.0 ⁴ 51	0.0 ⁵ 99	0.0 ⁴ 17
0.0 ⁴ 42	-0.0 ⁵ 26	-0.0 ⁶ 84	0.0 ³ 21	0.0 ⁴ 24	-0.0 ⁶ 75	0.0 ⁴ 37	-0.0 ⁴ 28	-0.0 ⁵ 87
0.0 ⁵ 53	-0.0 ⁶ 35	0.0 ⁴ 11	0.0 ⁴ 24	0.0 ³ 13	0.0 ⁴ 12	-0.0 ⁵ 36	-0.0 ⁵ 53	0.0 ⁵ 72
0.0 ⁵ 71	0.0 ⁴ 14	0.0 ⁵ 71	-0.0 ⁶ 75	0.0 ⁴ 12	0.0 ³ 12	0.0 ⁴ 1	0.0 ⁵ 16	0.0 ⁴ 18
0.0 ⁴ 38	0.0 ⁴ 62	-0.0 ⁴ 51	0.0 ⁴ 37	-0.0 ⁵ 36	0.0 ⁴ 1	0.0 ³ 36	-0.0 ⁴ 61	0.0 ⁶ 69
0.0 ⁵ 14	-0.0 ⁴ 1	0.0 ⁵ 99	-0.0 ⁴ 28	-0.0 ⁵ 53	0.0 ⁵ 16	-0.0 ⁴ 61	0.0 ³ 27	0.0 ⁴ 23
-0.0 ⁵ 74	0.0 ⁵ 15	0.0 ⁴ 17	-0.0 ⁵ 87	0.0 ⁵ 72	0.0 ⁴ 18	0.0 ⁶ 69	0.0 ⁴ 23	0.0 ³ 22

The training set is normalized using a nine-variate BCT. Optimal parameters for this transformation are determined through the maximum likelihood method: $\hat{\lambda}_1 = 1.3676$, $\hat{\lambda}_2 = 1.4807$, $\hat{\lambda}_3 = 1.078$, $\hat{\lambda}_4 = 1.7393$, $\hat{\lambda}_5 = 2.1004$, $\hat{\lambda}_6 = 1.1498$, $\hat{\lambda}_7 = 1.566$, $\hat{\lambda}_8 = 1.1685$, $\hat{\lambda}_9 = 2.1146$.

As a result of applying the BCT to the training set with components (3), the resulting sample has a vector of means: $\bar{Z} = \{0.70932; 0.66184; 0.86764; -0.57016; -0.47427; 0.81642; 0.62417; -0.81443; -0.47084\}$. The covariance matrix of the normalized sample is presented in Table 4.

Table 4 – The covariance matrix of the normalized set

0.0 ⁴ 29	0.0 ⁵ 14	0.0 ⁴ 2	0.0 ⁵ 2	0.0 ⁶ 13	0.0 ⁵ 21	0.0 ⁵ 39	0.0 ⁷ 73	-0.0 ⁷ 68
0.0 ⁵ 14	0.0 ⁴ 16	-0.0 ⁵ 48	-0.0 ⁷ 21	0.0 ⁷ 71	0.0 ⁵ 27	0.0 ⁴ 43	-0.0 ⁵ 16	0.0 ⁷ 31
0.0 ⁴ 2	-0.0 ⁵ 48	0.0 ³ 18	-0.0 ⁶ 42	0.0 ⁶ 43	0.0 ⁴ 42	-0.0 ⁴ 11	0.0 ⁵ 51	0.0 ⁶ 65
0.0 ⁵ 2	-0.0 ⁷ 21	-0.0 ⁶ 42	0.0 ⁵ 31	0.0 ⁶ 15	-0.0 ⁶ 15	0.0 ⁷ 14	-0.0 ⁵ 2	-0.0 ⁷ 32
0.0 ⁶ 13	0.0 ⁷ 71	0.0 ⁶ 43	0.0 ⁶ 15	0.0 ⁶ 33	0.0 ⁶ 58	0.0 ⁸ 8	-0.0 ⁶ 17	0.0 ⁶ 13
0.0 ⁵ 21	0.0 ⁵ 27	0.0 ⁴ 42	-0.0 ⁶ 15	0.0 ⁶ 58	0.0 ⁵ 55	0.0 ⁶ 15	0.0 ⁶ 92	0.0 ⁶ 72
0.0 ⁵ 39	0.0 ⁴ 43	-0.0 ⁴ 11	0.0 ⁵ 14	0.0 ⁸ 8	0.0 ⁵ 15	0.0 ⁴ 22	-0.0 ⁴ 1	0.0 ⁷ 12
0.0 ⁷ 73	-0.0 ⁵ 16	0.0 ⁵ 51	-0.0 ⁵ 2	-0.0 ⁶ 17	0.0 ⁶ 92	-0.0 ⁴ 1	0.0 ³ 11	0.0 ⁶ 89
-0.0 ⁷ 68	0.0 ⁷ 31	0.0 ⁶ 66	-0.0 ⁷ 32	0.0 ⁷ 13	0.0 ⁶ 72	0.0 ⁷ 12	0.0 ⁶ 89	0.0 ⁶ 57

According to the Mardia test, the normalized training set conforms to the multivariate normal distribution. The test statistic for multivariate skewness $N\beta_1/6$ is 175.47, which is below the critical value of 215.53 from the chi-square distribution with 165 degrees of freedom at a 0.005 significance level. Additionally, the test statistic for multivariate kurtosis β_2 is 99.76, which does not exceed the critical value of 104.19 for a normal distribution with a mean of 99, a variance of 4.062, and a significance level of 0.005.

6 RESULTS

After applying the normalization, nine-variate prediction ellipsoids were constructed for both the non-Gaussian data (NGD) (4) and the normalized data (ND) (5). The computer program implementing the constructed models was developed to conduct experiments. The program was written in the Python language.

Table 5 presents a comparative analysis of the performance metrics for prediction ellipsoid models in one-class classification.

Table 5 – The covariance matrix of the training set

Model	Accuracy	Specificity	Precision	Recall	F1 score
NGD	0.9412	0.9795	0.9893	0.9225	0.9547
ND	0.9782	0.9949	0.9974	0.9700	0.9835

The metrics for the nine-variate prediction ellipsoids for non-Gaussian data reflect good performance, which is likely attributed to the effective keystroke dynamics recognition. However, the nine-variate prediction ellipsoid for normalized data significantly outperforms it. Normalization has led to better detection of anomalies, reduced false positives, and improved overall classification performance. The results highlight the importance of applying normalization techniques to achieve a more reliable and accurate model for keystroke dynamics recognition.

7 DISCUSSION

The results demonstrate that applying the nine-variate BCT significantly improved the model's performance, underscoring the importance of normalization in addressing non-Gaussian data distributions. Choosing the appropriate normalization technique is vital, as it significantly affects the model's effectiveness and reliability. Multivariate normalization methods capture complex variable relationships, enhancing prediction ellipsoid accuracy and providing a more precise representation of data patterns.

The choice of significance level is an important factor in constructing the prediction ellipsoid. In this study, a

significance level of 0.005 was selected, aligning with common practices in one-class classification and outlier detection tasks [29].

Despite these advantages, using the prediction ellipsoid for normalized data comes with some disadvantages. It is generally considered essential to have at least 100 instances for building a high-quality model. Additionally, selecting the appropriate normalization transformation can be challenging, particularly for sets that have complex distributions or many outliers. The last is the need to choose a significance level that affects the efficiency and reliability of the prediction ellipsoid.

Since 10 data points were removed as outliers, the model might miss some underlying patterns. This limitation might be addressed with a more complex normalization, such as the Johnson transformation, which could better handle the nuances of the data distribution and improve the model's ability to represent all relevant data points.

CONCLUSIONS

The study examined the influence of data distribution on keystroke dynamics recognition, with a particular focus on the application of a nine-variate prediction ellipsoid for normalized data using the multivariate BCT.

The research evaluated the transformation's impact on performance by comparing it with a prediction ellipsoid model developed for non-Gaussian data. The findings demonstrated that applying the BCT significantly improved model performance. Normalizing the data led to a more accurate and robust prediction ellipsoid, outperforming the non-normalized model across various evaluation metrics. The nine-variate BCT not only enhanced overall model performance but also offered deeper insights into feature relationships by accounting for correlations. This underscores the critical importance of selecting an appropriate normalization technique when dealing with non-Gaussian distributions.

Despite the benefits of normalization, the study also identified disadvantages, particularly in determining the optimal normalization transformation. Additionally, selecting an appropriate significance level is important, as it directly impacts the reliability and effectiveness of the prediction ellipsoid.

The scientific novelty of the obtained results is that the nine-variate prediction ellipsoid for normalized data for solving keystroke dynamics recognition tasks is first built using the nine-variate BCT. The application of a constructed prediction ellipsoid for normalized data allowed to increase accuracy.

The practical significance of the obtained results is that the software implementing the constructed ellipsoid is developed in the Python language. The experimental results allow us to recommend the constructed model for use in practice.

Prospects for further research could explore alternative normalization techniques, such as the Johnson transformation, to further refine model performance. Additionally, investigating the effects of model complexity and

feature selection could provide valuable insights into improving methods for keystroke dynamics recognition.

ACKNOWLEDGEMENTS

The work is of an initiative nature and was carried out within the scope of the authors' scientific activities during their regular working hours in their primary positions.

REFERENCES

1. Dias T., Vitorino J., Maia E., Sousa O., Praça I. KeyRecs: A keystroke dynamics and typing pattern recognition dataset, *Data in Brief*, 2023, Vol. 50, pp. 1–8. DOI: 10.1016/j.dib.2023.109509
2. Alshehri A., Coenen F., Bollegala D. Accurate continuous and non-intrusive user authentication with multivariate keystroke streaming, *9th International Conference on Knowledge Discovery and Information Retrieval, Funchal, 1–3 November, 2017 : proceedings*. Funchal, SciTePress, 2017, pp. 61–70. DOI: 10.5220/0006497200610070
3. Ismail M., Salem M., Abd El Ghany M. et al. Outlier detection for keystroke biometric user authentication, *PeerJ Computer Science*, 2024, Vol. 50. DOI: 10.7717/peerjcs.2086
4. Oyebola O. Examining the distribution of keystroke dynamics features on computer, tablet and mobile phone platforms, *Mobile Computing and Sustainable Informatics: Proceedings of ICMCSI 2023, Nepal, 11–12 January, 2023 : proceedings*. Singapore, Springer Nature Singapore 2023, pp. 613–620 DOI: 10.1007/978-981-99-0835-6_43
5. Saini B., Singh P., Nayyar A., et al. A three-step authentication model for mobile phone user using keystroke dynamics, *IEEE Access*, 2020, Vol. 10, pp. 125909–125922. DOI: 10.1109/ACCESS.2020.3008019
6. Anusas-Amornkul N. Strengthening password authentication using keystroke dynamics and smartphone sensors, *ICICM 2019: Proceedings of the 9th International Conference on Information Communication and Management, Prague, 23–26 August, 2019 : proceedings*. New York, Association for Computing Machinery, 2019, pp. 70–74. DOI: 10.1145/3357419.3357425
7. Fierrez J., Pozo A., Martinez-Diaz M. et al. Benchmarking touchscreen biometrics for mobile authentication, *IEEE Transactions on Information Forensics and Security*, 2018. Vol. 13, № 11, pp. 2720–2733. DOI: 10.1109/TIFS.2018.2833042
8. Li Q. Chen H. CDAS: A continuous dynamic authentication system, *Proceedings of the 2019 8th International Conference on Software and Computer Applications, Penang, 19–21 February, 2019, proceedings*. New York, Association for Computing Machinery, 2019, pp. 447–452. DOI: 10.1145/3316615.3316691
9. Chen L., Zhong Y., Ai W. et al. Continuous authentication based on user interaction behavior / [L. Chen.] // *2019 7th International Symposium on Digital Forensics and Security (ISDFS), Barcelos, 10–12 June, 2019 : proceedings*. Barcelos, IEEE, 2019, pp. 1–6. DOI: 10.1109/ISDFS.2019.8757539
10. Baynath P., Soyjaudah K., Khan M. Keystroke recognition using neural network, *2017 5th International Sympo-*

- sium on Computational and Business Intelligence (ISCBI), Dubai, 11–14 August, 2017 : proceedings.* Dubai, IEEE, 2017, pp. 86–90. DOI: 10.1109/ISCBI.2017.8053550
11. Andrian A., Jayabalan M., Thiruchelvam V. Keystroke dynamics based user authentication using deep multilayer perceptron, *International Journal of Machine Learning and Computing*, 2020, Vol. 10, № 1, pp. 134–139. DOI: 10.18178/ijmlc.2020.10.1.910
 12. Sharma A., Jureček M., Stamp M. Keystroke dynamics for user identification, *arXiv preprint arXiv:2307.05529*, 2023, pp. 1–22. DOI: 10.48550/arXiv.2307.05529
 13. Raul N., Shankarmani R., Joshi P. A comprehensive review of keystroke dynamics-based authentication mechanism, *International Conference on Innovative Computing and Communications. Advances in Intelligent Systems and Computing, Delhi, 20–21 February, 2021 : proceedings.* Singapore, Springer, 2021, Vol. 1059, pp. 149–162 DOI: 10.1007/978-981-15-03245_13
 14. Antal M., Szabó L. An evaluation of one-class and two-class classification algorithms for keystroke dynamics authentication on mobile devices, *2015 20th International Conference on Control Systems and Computer Science, Bucharest, 27–29 May, 2015 : proceedings.* Bucharest, IEEE, 2015, pp. 343–350. DOI: 10.1109/CSCS.2015.16
 15. Marques H., Swersky L., Sander J. et al. On the evaluation of outlier detection and one-class classification: a comparative study of algorithms, model selection, and ensembles, *Data Mining and Knowledge Discovery*, 2023, Vol. 37, pp. 1–45. DOI: 10.1007/s10618-023-00931-x
 16. Toosi R., Akhaee M. Time-frequency analysis of keystroke dynamics for user authentication, *Future Generation Computer Systems*, 2021, Vol. 115, pp. 438–447. DOI: 10.1016/j.future.2020.09.027
 17. Jawed H., Ziad Z., Khan M. et al. Anomaly detection through keystroke and tap dynamics implemented via machine learning algorithms, *Turkish Journal of Electrical Engineering and Computer Sciences*, 2018, Vol. 26, № 4, pp. 1698–1709. DOI: 10.3906/elk-1711-410
 18. Ceker H., Upadhyaya S. User authentication with keystroke dynamics in long-text data, *2016 IEEE 8th International Conference on Biometrics Theory, Applications and Systems (BTAS), Niagara Falls, 6–9 September, 2016.* Niagara Falls, IEEE, 2016, pp. 1–6. DOI: 10.1109/BTAS.2016.7791182
 19. Patel Y., Ouazzane K., Vassilev V. et al. Keystroke dynamics using auto encoders, *2019 International Conference on Cyber Security and Protection of Digital Services, Oxford, 3–4 June, 2019 : proceedings.* Oxford, IEEE, 2019, pp. 1–8. DOI: 10.1109/CyberSecPODS.2019.8885203
 20. Centeno M., Moorsel A., Castruccio S. Smartphone continuous authentication using deep learning autoencoders, *15th Annual Conference on Privacy, Security and Trust (PST), Calgary, 28–30 August, 2017 : proceedings.* Calgary, IEEE, 2017, pp. 1470–1478. DOI: 10.1109/PST.2017.00026
 21. Eizaguirre-Peral I., Seguro-Gil L., Zola F. Conditional generative adversarial network for keystroke presentation attack, *arXiv preprint arXiv:2212.08445*, 2022, pp. 1–4. DOI: 10.48550/arXiv.2212.08445
 22. Kim S., Park D., Jung J. Evaluation of one-class classifiers for fault detection: Mahalanobis classifiers and the Mahalanobis-Taguchi system, *Processes*, 2021, Vol. 9, № 8, P. 1450. DOI: 10.3390/pr9081450
 23. Bezerra V., Da Costa V., Barbon Junior S. et al. IoTDS: A one-class classification approach to detect botnets in Internet of Things devices, *Sensors*, 2019, Vol. 19, № 14, P. 3188. DOI: 10.3390/s19143188
 24. Prykhodko S. B., Shutko I. S., Prykhodko A. S. A nonlinear regression model to estimate the size of web apps created using the CakePHP framework, *Radio Electronics, Computer Science, Control*, 2021, Vol. 59, № 4, pp. 129–139. DOI: 10.15588/1607-3274-2021-4-12
 25. Lam K., Meijer K., Loonstra F. et al. Real-world keystroke dynamics are a potentially valid biomarker for clinical disability in multiple sclerosis, *Multiple Sclerosis Journal*, 2021, Vol. 27, № 4, pp. 1421–1431. DOI: 10.1177/1352458520968797
 26. Prykhodko S., Trukhov A. Face recognition using the ten-variate prediction ellipsoid for normalized data based on the Box-Cox transformation, *Radio Electronics, Computer Science, Control*, 2024, Vol. 2, pp. 82–89. DOI: 10.15588/1607-3274-2024-2-9
 27. Prykhodko S., Trukhov A. Application of a ten-variate prediction ellipsoid for normalized data and machine learning algorithms for face recognition, *CMIS-2024: Seventh International Workshop on Computer Modeling and Intelligent Systems, Zaporizhzhia, 3 May, 2024 : proceedings.* Aachen, CEUR, 2024, pp. 362–375.
 28. Nurmaini S., Darmawahyuni A., Sakti Mukti A. et al. Deep learning-based stacked denoising and autoencoder for ECG heartbeat classification, *Electronics*, 2020, Vol. 9, № 1, P. 135. DOI: 10.3390/electronics9010135
 29. Prykhodko S., Prykhodko N., Makarova L. et al. Detecting outliers in multivariate non-Gaussian data on the basis of normalizing transformations, *2017 IEEE First Ukraine Conference on Electrical and Computer Engineering, Kyiv, 29 May – 2 June, 2017 : proceedings.* New York, IEEE, 2017, pp. 846–849. DOI: 10.1109/UKRCON.2017.8100366

Received 11.11.2024.
Accepted 10.01.2025.

РОЗПІЗНАВАННЯ КЛАВІАТУРНОГО ПОЧЕРКУ ЗА ДОПОМОГОЮ ДЕВ'ЯТИВИМІРНОГО ЕЛІПСОЇДА ПРОГНОЗУВАННЯ ДЛЯ НОРМАЛІЗОВАНИХ ДАНИХ

Приходько С. Б. – д-р техн. наук, професор, завідувач кафедри програмного забезпечення автоматизованих систем Національного університету кораблебудування ім. адмірала Макарова, Миколаїв, Україна.

Трухов А. С. – аспірант кафедри програмного забезпечення автоматизованих систем Національного університету кораблебудування ім. адмірала Макарова, Миколаїв, Україна.

АНОТАЦІЯ

Актуальність. Розпізнавання клавіатурного почерку є важливим елементом у підвищенні безпеки, що дозволяє реалізувати персоналізовану автентифікацію користувачів та підтримує різні системи перевірки особистості. Це дослідження вивчає вплив розподілу даних на ефективність моделей однокласової класифікації в задачах розпізнавання клавіатурного почерку, зосереджуючи увагу на застосуванні дев'ятивимірного еліпсоїда прогнозування. Об'єктом дослідження є процес розпізнавання клавіатурного почерку. Предметом дослідження є математичні моделі для розпізнавання клавіатурного почерку. На відміну від типових підходів, що передбачають багатовимірний нормальний розподіл даних, реальні набори даних часто відхиляються від нього, що ускладнює побудову точних і надійних моделей. Для вирішення цієї проблеми дані були нормалізовані за допомогою багатовимірного перетворення Бокса-Кокса, що дозволило покращити вірогідність розпізнавання клавіатурного почерку за допомогою застосування еліпсоїда прогнозування для нормалізованих даних.

Метою роботи є підвищення ймовірності розпізнавання клавіатурного почерку шляхом побудови дев'ятивимірного еліпсоїда прогнозування для нормалізованих даних із використанням багатовимірного перетворення Бокса-Кокса.

Метод. Дослідження включає побудову дев'ятивимірного еліпсоїда прогнозування для даних, нормалізованих за допомогою перетворення Бокса-Кокса. Квадрат відстані Махаланобіса застосовується для виявлення та видалення викидів, а тест Мардіа оцінює відхилення багатовимірного розподілу від нормального. Оцінки параметрів багатовимірного перетворення Бокса-Кокса отримані методом максимальної правдоподібності.

Результати. Результати показують значне підвищення вірогідності розпізнавання після нормалізації, що полягає у збільшенні точності та надійності порівняно з моделями, побудованими для ненормалізованих даних. Застосування дев'ятивимірного перетворення Бокса-Кокса дозволило краще врахувати кореляції між ознаками, що дозволило еліпсоїду прогнозування краще захоплювати складні закономірності даних.

Висновки. Для розпізнавання клавіатурного почерку була розроблена математична модель у формі дев'ятивимірного еліпсоїда прогнозування для даних, нормалізованих із використанням багатовимірного перетворення Бокса-Кокса, що підвищує ймовірність розпізнавання в порівнянні з моделями, побудованими для ненормалізованих даних. Однак залишаються труднощі у визначенні оптимального методу нормалізації та виборі рівня значущості для побудови еліпсоїда прогнозування. Ці висновки підкреслюють важливість ретельного вибору ознак та застосування вдосконалених методів нормалізації даних для подальших досліджень у сфері розпізнавання клавіатурного почерку.

КЛЮЧОВІ СЛОВА: розпізнавання клавіатурного почерку, багатовимірне перетворення Бокса-Кокса, еліпсоїд прогнозування, нормалізуюче перетворення.

ЛІТЕРАТУРА

1. KeyRecs: A keystroke dynamics and typing pattern recognition dataset / [T. Dias, J. Vitorino, E. Maia et al.] // Data in Brief. – 2023. – Vol. 50. – P. 1–8. DOI: 10.1016/j.dib.2023.109509
2. Alshehri A. Accurate continuous and non-intrusive user authentication with multivariate keystroke streaming / A. Alshehri, F. Coenen, D. Bollegala // 9th International Conference on Knowledge Discovery and Information Retrieval, Funchal, 1–3 November, 2017 : proceedings. – Funchal : SciTePress, 2017. – P. 61–70. DOI: 10.5220/0006497200610070
3. Ismail M. Outlier detection for keystroke biometric user authentication / [M. Ismail, M. Salem, M. Abd El Ghany et al.] // PeerJ Computer Science. – 2024. – Vol. 50. DOI: 10.7717/peerj-cs.2086
4. Oyebola O. Examining the distribution of keystroke dynamics features on computer, tablet and mobile phone platforms / O. Oyebola // Mobile Computing and Sustainable Informatics: Proceedings of ICMCSI 2023, Nepal, 11–12 January, 2023 : proceedings. – Singapore: Springer Nature Singapore 2023. – P. 613–620 DOI: 10.1007/978-981-99-0835-6_43
5. A three-step authentication model for mobile phone user using keystroke dynamics / [B. Saini, P. Singh, A. Nayyar, et al.] // IEEE Access. – 2020. – Vol. 10. – P. 125909–125922. DOI: 10.1109/ACCESS.2020.3008019
6. Anusas-Amornkul N. Strengthening password authentication using keystroke dynamics and smartphone sensors / N. Anusas-Amornkul // ICICM 2019: Proceedings of the 9th International Conference on Information Communication and Management, Prague, 23 – 26 August, 2019 : proceedings. – New York : Association for Computing Machinery, 2019. – P. 70–74. DOI: 10.1145/3357419.3357425
7. Benchmarking touchscreen biometrics for mobile authentication / [J. Fierrez, A. Pozo, M. Martinez-Diaz et al.] // IEEE Transactions on Information Forensics and Security. – 2018. – Vol. 13, № 11. – P. 2720–2733. DOI: 10.1109/TIFS.2018.2833042
8. Li Q. CDAS: A continuous dynamic authentication system / Q. Li, H. Chen // Proceedings of the 2019 8th International Conference on Software and Computer Applications, Penang, 19–21 February, 2019 : proceedings. – New York : Association for Computing Machinery, 2019. – P. 447–452. DOI: 10.1145/3316615.3316691
9. Continuous authentication based on user interaction behavior / [L. Chen, Y. Zhong, W. Ai et al.] // 2019 7th International Symposium on Digital Forensics and Security (ISDFS), Barcelos, 10–12 June, 2019 : proceedings. – Barcelos : IEEE, 2019. – P. 1–6. DOI: 10.1109/ISDFS.2019.8757539

10. Baynath P. Keystroke recognition using neural network / P. Baynath, K. Soyjaudah, M. Khan // 2017 5th International Symposium on Computational and Business Intelligence (ISCBI), Dubai, 11–14 August, 2017 : proceedings. – Dubai : IEEE, 2017. – P. 86–90. DOI: 10.1109/ISCBI.2017.8053550
11. Andrian A. Keystroke dynamics based user authentication using deep multilayer perceptron / A. Andrian, M. Jayabalan, V. Thiruchelvam // International Journal of Machine Learning and Computing. – 2020. – Vol. 10, № 1. – P. 134–139. DOI: 10.18178/ijmlc.2020.10.1.910
12. Sharma A. Keystroke dynamics for user identification / A. Sharma, M. Jureček, M. Stamp // arXiv preprint arXiv:2307.05529, 2023. – P. 1–22. DOI: 10.48550/arXiv.2307.05529
13. Raul N. A comprehensive review of keystroke dynamics-based authentication mechanism / N. Raul, R. Shankarmani, P. Joshi // International Conference on Innovative Computing and Communications. Advances in Intelligent Systems and Computing, Delhi, 20–21 February, 2021 : proceedings. – Singapore : Springer, 2021. – Vol. 1059. – P. 149–162. DOI: 10.1007/978-981-15-03245_13
14. Antal M. An evaluation of one-class and two-class classification algorithms for keystroke dynamics authentication on mobile devices / M. Antal, L. Szabó // 2015 20th International Conference on Control Systems and Computer Science, Bucharest, 27–29 May, 2015 : proceedings. Bucharest : IEEE, 2015. – P. 343–350. DOI: 10.1109/CSCS.2015.16
15. On the evaluation of outlier detection and one-class classification: a comparative study of algorithms, model selection, and ensembles / [H. Marques, L. Swersky, J. Sander et al.] // Data Mining and Knowledge Discovery. – 2023. – Vol. 37. – P. 1–45. DOI: 10.1007/s10618-023-00931-x
16. Toosi R. Time-frequency analysis of keystroke dynamics for user authentication / R. Toosi, M. Akhaee // Future Generation Computer Systems. – 2021. – Vol. 115. – P. 438–447. DOI: 10.1016/j.future.2020.09.027
17. Jawed H. Anomaly detection through keystroke and tap dynamics implemented via machine learning algorithms / [H. Jawed, Z. Ziad, M. Khan et al.] // Turkish Journal of Electrical Engineering and Computer Sciences. – 2018. – Vol. 26, № 4. – P. 1698–1709. DOI: 10.3906/elk-1711-410
18. Ceker H. User authentication with keystroke dynamics in long-text data / H. Ceker, S. Upadhyaya // 2016 IEEE 8th International Conference on Biometrics Theory, Applications and Systems (BTAS), Niagara Falls, 6–9 September, 2016. – Niagara Falls : IEEE, 2016. – P. 1–6. DOI: 10.1109/BTAS.2016.7791182
19. Keystroke dynamics using auto encoders / [Y. Patel, K. Ouazzane, V. Vassilev et al.] // 2019 International Conference on Cyber Security and Protection of Digital Services, Oxford, 3–4 June, 2019 : proceedings. – Oxford : IEEE, 2019. – P. 1–8. DOI: 10.1109/CyberSecPODS.2019.8885203
20. Centeno M. Smartphone continuous authentication using deep learning autoencoders / M. Centeno, A. Moorsel, S. Castruccio // 15th Annual Conference on Privacy, Security and Trust (PST), Calgary, 28–30 August, 2017 : proceedings. – Calgary : IEEE, 2017. – P. 1470–1478. DOI: 10.1109/PST.2017.00026
21. Eizaguirre-Peral I. Conditional generative adversarial network for keystroke presentation attack / I. Eizaguirre-Peral, L. Seguro-Gil, F. Zola // arXiv preprint arXiv:2212.08445 2022. – P. 1–4. DOI: 10.48550/arXiv.2212.08445
22. Kim S. Evaluation of one-class classifiers for fault detection: Mahalanobis classifiers and the Mahalanobis–Taguchi system / S. Kim, D. Park, J. Jung // Processes. – 2021. – Vol. 9, № 8. – P. 1450. DOI: 10.3390/pr9081450
23. IoTDS: A one-class classification approach to detect botnets in Internet of Things devices / [V. Bezerra, V. da Costa, S. Barbon Junior et al.] // Sensors. – 2019. – Vol. 19, № 14. – P. 3188. DOI: 10.3390/s19143188
24. Prykhodko S. B. A nonlinear regression model to estimate the size of web apps created using the CakePHP framework / S. B. Prykhodko, I. S. Shutko, A. S. Prykhodko // Radio Electronics, Computer Science, Control. – 2021. – Vol. 59, № 4. – P. 129–139. DOI: 10.15588/1607-3274-2021-4-12
25. Real-world keystroke dynamics are a potentially valid biomarker for clinical disability in multiple sclerosis / [K. Lam, K. Meijer, F. Loonstra et al.] // Multiple Sclerosis Journal. – 2021. – Vol. 27, № 4. – P. 1421–1431. DOI: 10.1177/1352458520968797
26. Prykhodko S. Face recognition using the ten-variate prediction ellipsoid for normalized data based on the Box-Cox transformation / S. Prykhodko, A. Trukhov // Radio Electronics, Computer Science, Control. – 2024. – Vol. 2. – P. 82–89. DOI: 10.15588/1607-3274-2024-2-9
27. Prykhodko S. Application of a ten-variate prediction ellipsoid for normalized data and machine learning algorithms for face recognition / S. Prykhodko, A. Trukhov // CMIS-2024: Seventh International Workshop on Computer Modeling and Intelligent Systems, Zaporizhzhia, 3 May, 2024 : proceedings. – Aachen : CEUR, 2024. – P. 362–375.
28. Deep learning-based stacked denoising and autoencoder for ECG heartbeat classification / [S. Nurmaini, A. Darmawatyuni, A. Sakti Mukti et al.] // Electronics. – 2020. – Vol. 9, № 1. – P. 135. DOI: 10.3390/electronics9010135
29. Detecting outliers in multivariate non-Gaussian data on the basis of normalizing transformations / [S. Prykhodko, N. Prykhodko, L. Makarova et al.] // 2017 IEEE First Ukraine Conference on Electrical and Computer Engineering, Kyiv, 29 May – 2 June, 2017 : proceedings. – New York : IEEE, 2017. – P. 846–849. DOI: 10.1109/UKRCON.2017.8100366

METHOD OF NEURAL NETWORK DETECTION OF DEFECTS BASED ON THE ANALYSIS OF ROTATING MACHINES VIBRATIONS

Sobko O. V. – Postgraduate student of the Department of Computer Science, Khmelnytskyi National University Khmelnytskyi, Ukraine.

Dydo R. A. – Student of the Department of Computer Science, Khmelnytskyi National University Khmelnytskyi, Ukraine.

Mazurets O. V. – PhD, Associate Professor, Department of Computer Science, Khmelnytskyi National University, Khmelnytskyi, Ukraine.

ABSTRACT

Context. The paper proposes a solution to the urgent problem of detecting equipment defects by analyzing the vibrations of rotating machines. The object of study is the process of detecting defects by analyzing the vibrations of rotating machines. The subject of study is artificial intelligence methods for detecting defects by analyzing the vibrations of rotating machines.

Objective. Improving the accuracy of detecting defects in the analysis of rotating machine vibrations by creating a method for neural network detection of defects in the analysis of rotating machine vibrations and a corresponding neural network model that can detect defects in the analysis of rotating machine vibrations without removal preliminary noise in order to preserve important features for more accurate classification.

Method. A method of neural network defect detection based on the analysis of vibrations of rotating machines is proposed, which is capable of predicting the presence or absence of a defect based on the input data of vibrations with the implementation of preliminary processing, namely the creation of a two-dimensional time-frequency image. The method differs from the existing ones in that the defect analysis is performed without removing noise by fine-tuning the model parameters.

Results. The proposed method of neural network detection of defects based on the analysis of rotating machines vibrations is implemented in the form of a web application and the effectiveness of the neural network model obtained by performing the steps of the method is studied.

Conclusions. The study results show that the model has achieved high accuracy and consistency between training and validation data, which is confirmed by high values of such indicators as Accuracy, Precision, Recall i F1-Score on the validation dataset, as well as minimal losses. The cross-validation confirmed the stable efficiency of the model, demonstrating high averaged metrics and insignificant deviations from the obtained metrics. Thus, the neural network model detects defects in rotating machines with high efficiency even without cleaning vibration signals from noise. Prospects for further research are to test the described method and the resulting neural network model on larger data sets.

KEYWORDS: defects, analysis, vibrations, rotating machines, neural network, ResNet50.

ABBREVIATIONS

CNN is a Convolutional Neural Network;
RNN is a Recurrent Neural Network;
SVM is a Support Vector Machine;
k-NN is a k-Nearest Neighbors;
AIoT is an Artificial Intelligence of Things;
1D-CNN is a One-Dimensional Convolutional Neural Network;
LMD is a Local Mean Decomposition;
EMD is an Empirical Mode Decomposition;
DT is a Decision Tree;
PNG is a Portable Network Graphics;
RNT is a ResNet Training.

NOMENCLATURE

X is a set of vibration signals;
 x_1 is a first vibration signal;
 x_2 is a second vibration signal;
 x_n is a n -th vibration signal;
 n is a number of vibration signals in the set;
 x_i is a vibration signal;
 R^T is a time series;
 T is a time series length;
 m_i is a class label;

$Model_{\theta}()$ is a neural network model;
 θ is a set of parameters $Model_{\theta}()$;
 \hat{m}_i is a predicted class probability for the sample under study;
 \bar{x} is a vector of vibration signals;
 x_{1t1} is a first vibration signal at a given time in vector of vibration signals;
 x_{2t2} is a second vibration signal at a given time in vector of vibration signals;
 x_{nT} is a n -th vibration signal at a given time in vector of vibration signals;
 $f(t_i)$ is a function that projects indices onto the signal amplitude;
 t_i is a moment in time;
 I is a pixel matrix;
 H is an image height;
 W is an image width;
 Γ is an image of vibration signals;
 k is a number of folds for cross-validation.

INTRODUCTION

Rotating machines, including electric motors, pumps, turbines, generators, and components of airplanes, heli-

copters, and other mechanisms, are central to modern industry and transportation. They perform important functions such as energy conversion and transmission, as well as provide essential mobility and safety in aviation and automotive systems. The reliability of these machines is an important factor, as malfunctions can lead not only to financial losses but also to serious accidents. Regular monitoring of the condition of rotating machines through vibration analysis helps to detect early signs of defects, which in turn allows for timely maintenance and prevents disasters [1].

The vibration signals generated during the operation of these machines contain information about the condition of their components – bearings, shafts, gears, etc. Changes in the frequency, amplitude, and pattern of vibrations can signal defects such as cracks, wear, or imbalance.

Thanks to the introduction of new technologies, there is a continuous evolution in the areas of machine condition monitoring and performance evaluation. One of the main reasons for this progress is the significant reduction in the cost of sensors and data storage systems. Today, high-precision vibration, temperature, and other sensors have become available for a wide range of industries, allowing for continuous collection of important information about the condition of equipment [2].

On the other hand, the ever-increasing computing power of modern computers makes it possible to process and analyze huge amounts of data in real time, allowing for the application of sophisticated machine learning algorithms and neural networks to detect anomalies that may indicate potential malfunctions. Early detection of problems in rotating machines or other mechanisms allows timely elimination of them, which significantly reduces the risk of accidents and increases the overall efficiency of production [3].

The development of data-driven methods, as well as big data and computational intelligence technologies, is opening up new approaches to diagnosing rotating machinery faults. Modern approaches to data analysis use machine learning algorithms and artificial neural networks. These algorithms allow not only to detect faults but also to predict their occurrence, as well as to classify different types of defects based on the analysis of vibration data and other parameters. Thanks to these intelligent solutions, it is possible to proactively monitor the condition of rotating machines, which significantly increases their reliability and efficiency.

The object of study is the process of detecting defects by analyzing the vibrations of rotating machines.

The subject of study is artificial intelligence methods for detecting defects by analyzing the vibrations of rotating machines.

The purpose of the work is to improve the accuracy of defect detection by analyzing the vibrations of rotating machines.

1 PROBLEM STATEMENT

Rotating machines are key components in various industrial sectors. They are prone to wear and tear and defects that can cause unwanted vibrations, which in turn can lead to accidents and production stoppages. Therefore, timely detection of defects in them is of great importance to ensure equipment reliability, reduce maintenance costs, and increase overall efficiency. The development of a method for neural network detection of defects based on the analysis of vibration signals of rotating machines will allow detecting defects to avoid the emergency state of rotating machines.

Let $X = \{x_1, x_2, \dots, x_n\}$ is a set of vibration signals, where each $x_i \in R^T$ is a time series of length T , representing data from the vibration sensor. Everyone x_i label corresponds to $m_i \in \{0, 1\}$, where 0 indicates a good condition of the machine, and 1 indicates a faulty condition. It is necessary to determine the presence or absence of a defect for the input data from the vibration sensor. Thus, there is a need to develop a neural network method for detecting defects based on the analysis of vibrations of rotating machines and a neural network model is required $Model_{\theta}()$ with parameters θ at which each input value of the vibration signal x_i is processed and fed to the input of the neural network model to obtain the predicted probability of the class (“without defect” or “with defect”) $\hat{m}_i = Model_{\theta}(x_i)$. To study the effectiveness of the obtained neural network model, calculate the following metrics Accuracy, Precision, Recall, F1-score.

2 REVIEW OF THE LITERATURE

The urgency of the task of detecting defects in rotating machinery is due to the need to reduce the risk of failure, minimize downtime, and reduce maintenance costs. By detecting defects such as bearing wear, overheating, or mismatches in the lubrication systems, more serious failures can be prevented. The use of modern methods, including vibration sensors and infrared thermography, allows for effective diagnostics, and the integration of artificial intelligence and machine learning also increases the accuracy and speed of defect analysis [4].

Next, consider scientific research devoted specifically to the detection of defects in rotating machine parts using vibration methods.

In the article [5] The article considers the application of deep learning to diagnose malfunctions of rotating machines based on the analysis of vibration signals. In particular, considerable attention is paid to the use of CNN for feature extraction and classification of vibration data. CNN can detect defects, such as damage to bearings and gearboxes, using time-frequency signal analysis. In addition to CNN, the article discusses the use of auto-codes that are capable of self-learning and analyzing incoming vibration data. This approach makes it possible to detect defects even in difficult conditions, including in the presence of significant noise. In addition, the analysis of se-

quences of signals that change over time can be performed using RNN, which makes it possible to track the development of defects in real time.

In the article [6] the authors solved the problem of detecting faults in rotating machines by analyzing vibration signals. They compared different classifiers, such as support vector machine SVM, naive Bayesian classifier, and k-NN nearest neighbors, to find the most effective approach for accurate diagnosis. According to the study, SVM proved to be the most effective classifier due to its high accuracy in detecting faults.

Authors [7] applied an AIoT approach to detect anomalies in rotating machines by monitoring vibrations. In their approach, vibration signals are collected using an MPU6050 accelerometer connected to a Raspberry Pi 4B device that acts as a peripheral computing device. For efficient implementation, this device uses SVM, which provides the ability to analyze data directly on the periphery.

Paper [8] describes the use of 1D-CNN to diagnose the condition of rotating machines. The authors propose to use 1D-CNN to automatically detect defects in gears based on vibration signals, which allows identifying various types of damage, such as wear or chipped teeth. The authors claim that this approach reduces the need for manual feature extraction, which is usually required for traditional machine learning algorithms. During the experiments, the authors compared 1D-CNN with other machine learning methods, including decision trees, random forests, and SVM, and demonstrated that 1D-CNN outperformed these algorithms, reaching 97.11% accuracy.

Article [9] discusses the diagnostics of rotating machines using adaptive processing of vibration signals. Experimental results show that the use of stochastic resonant wavelet decomposition and LMD morphological filtering reduces noise and increases fault identification accuracy by up to 95%. This approach demonstrates the effectiveness of adaptive processing of vibration signals to improve the accuracy of diagnostics of rotating machines.

In [10], the authors also investigate methods for detecting malfunctions of rotating machines. First, the vibration signals were pre-processed using EMD to extract the most important components while eliminating redundant information. After that, significant characteristics in the time and frequency domains were extracted from each channel of the processed signal, which were combined to create a clear representation of the features of each class. The characteristics were selected from a set including time, frequency, and statistical data based on long-term experiments. The combined feature vector was used to train and test the SVM classifier using 10-fold cross-validation. In addition, the basic SVM classifier was compared with the k-NN and DT methods. The proposed approach demonstrated the best results, reaching an accuracy of 99%.

Thus, the task of neural network defect detection based on the analysis of rotating machine vibrations is relevant, as timely diagnosis of equipment malfunctions is

important for preventing accidents and optimizing maintenance costs. In modern research, neural networks are being actively studied, which have demonstrated high efficiency in classifying the state of machines based on vibration signals. Existing approaches for detecting defects in rotating machines are usually based on traditional signal processing methods with the removal of noise present in vibration signals. However, it is difficult to find the threshold for noise removal, which leads to the removal of important features that would indicate the presence of defects along with the noise. Thus, there is a need to develop a new method based on neural networks that can learn complex features in vibration data, even in the presence of noise, which allows for more accurate and reliable diagnostics.

3 MATERIALS AND METHODS

To solve this problem, propose a method of neural network detection of defects based on the analysis of rotating machines vibrations. The method scheme is shown in Figure 1.

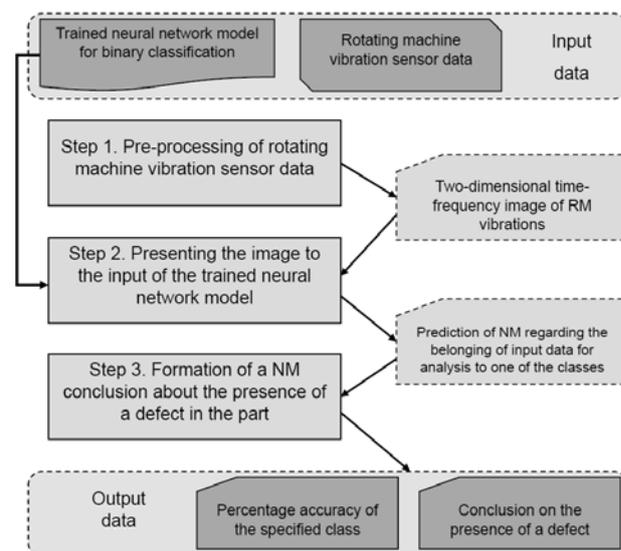


Figure 1 – Scheme of the method of neural network detection of defects based on the analysis of rotating machines vibrations

The input data of the method are $Model_{\theta}$ with θ for binary classification, as well as the set X and has the form shown in Figure 2.

The first step involves pre-processing the data from the vibration sensor, namely converting the vibration signal into a two-dimensional image. Data from the sensor is converted into an image by creating a graph of signal amplitudes over time. The set of vibration signals X is represented as a vector (1) with numerical values that will be used to plot the signal amplitudes over time, x_i represents the amplitude of the signal at a given time t_i .

$$\bar{x} = [x_{1t1}, x_{2t2}, \dots, x_{nt}] \quad (1)$$

The x -signal is represented as a line graph, where the amplitude values are displayed along the Y and the time indices are displayed along the X . Thus, on the graph, each point (t_i, x_i) corresponds to the value of the signal at a given time t_i . This can be expressed as a function that projects the indices onto the signal amplitude: (2)

$$f(t_i) = x_i, i = 1, 2, \dots, T . \quad (2)$$

```
T0002.txt: Блокнот
Файл Редагування Формат Вигляд Довідка
-1.9190626999999999e-001
6.1933791999999999e-001
9.0571493000000003e-001
-2.3381837999999999e-001
8.3622079999999999e-001
6.1709183000000001e-001
-7.1722269000000005e-001
1.6831778300000000e+000
7.9228717000000004e-001
-1.1242148900000000e+000
1.5641797799999999e+000
4.4104302000000001e-001
1.2651348099999999e+000
3.2017171400000000e+000
-1.0754298000000000e-001
4.3879690999999998e-001
2.1243556000000000e-001
-1.1615450400000000e+000
2.4128887699999999e+000
-2.8583792000000002e-001
```

Figure 2 – Example of sample vibration sensor input data

The graph is built by connecting the points $(t_i, f(t_i))$ lines, which visually represents the signal as a continuous curve. When a graph is created, it is saved as a PNG image. The image is saved as a raster representation, where each pixel contains information about the signal amplitude at the corresponding time. Let's denote the created image as a pixel matrix I , size $H \times W$. The image is then converted to an array of pixels and scaled to fit 224×224 pixels. When plotting, the dimensionality of the data is reduced, which is due to the peculiarities of working with images in most neural network architectures. The Matplotlib library automatically scales and arranges the data so that the plot fits into the dimensions of the generated image. At this point, the data is compressed to 50176 points. The image is formed as a two-dimensional tensor $I \in R^{224 \times 224}$. The following image is then transmitted to the input $Model_0()$.

The second step is to feed the resulting image I to enter the previously trained $Model_0()$ for classification. $Model_0()$ with parameters θ , receives at the entrance I and predicts the class \hat{m}_i , where $\hat{m}_i \in [0, 1]$ is the probability of a signal belonging to one of the classes.

At step 3, based on the forecast, a conclusion is formed on the condition of rotating machine parts. If $\hat{m}_i \geq 0.4$, the conclusion indicates that the part has a defect, otherwise – without a defect. The value of 0.4 was chosen as a threshold that balances the sensitivity and specificity of the model, avoiding false negatives (when

parts with defects are classified as defect-free) while minimizing false positives. This value was chosen empirically, taking into account the analysis of the model's performance on training and validation datasets, to ensure reliable diagnostics and timely detection of defects. This step also calculates the percentage accuracy of belonging to a particular class (3), which is the output of the described method of neural network defect detection based on the analysis of rotating machine vibrations:

$$\hat{m}_i \times 100\% . \quad (3)$$

To receive $Model_0()$, it is necessary to determine the type and architecture of the neural network and train it on the data set. The scheme for obtaining a typical neural network model used for the method of neural network defect detection, based on the analysis of rotating machine vibrations, is shown in Figure 3. This scheme illustrates the overall process of developing a neural network model that can detect defects in rotating machinery by analyzing the vibration data collected from the machines. Figure 3 provides a clear visual representation of the typical workflow involved in creating such a model.

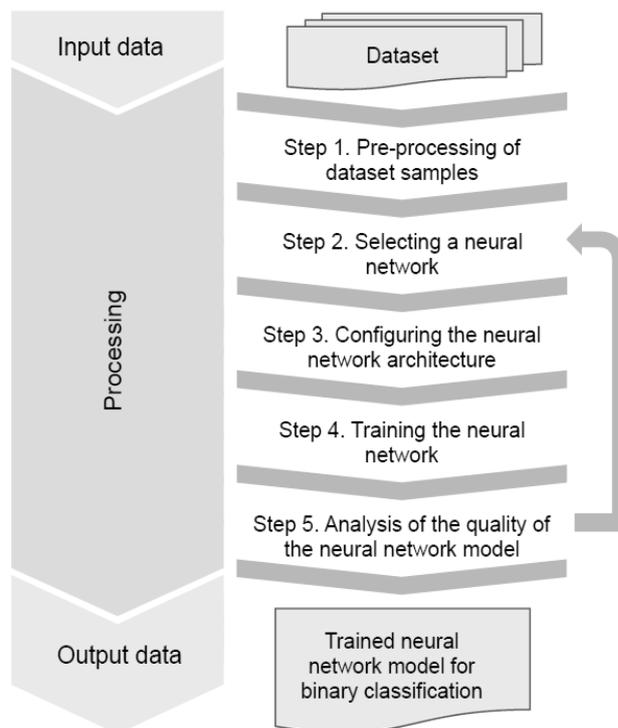


Figure 3 – Scheme of obtaining a typical neural network model for the method of neural network defect detection based on the analysis of vibrations of rotating machines

Input data to receive $Model_0()$ there is a dataset received for participation in the competition “All-Ukrainian competition of young scientists in the field of intelligent IT” (<https://zp.edu.ua/vkiit/>).

The first step is to preprocess the dataset. Since the dataset contains text files with numerical values of vibration that were taken from the sensors (Figure 1) and the

key.txt file, which is the key to which class each file with signals from the vibration sensor belongs to, it is first necessary to represent the values of the vibration signals in the form I . This is done in the same way as the transformation of the input data in step 1 of the neural network defect detection method described above for analyzing the vibrations of rotating machines. Figure 4 shows the image before and after preprocessing.

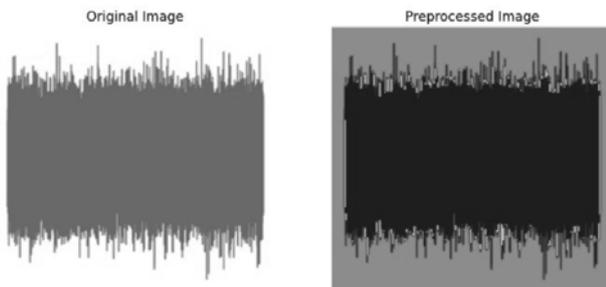


Figure 4 – Images before and after preprocessing

After converting data into images according to the key.txt file, the images are labeled with their “no defect” and “defect” classes. To train the neural network model, 80% of the data is selected from the dataset, and the remaining 20% is used to validate the model.

The second step is to select the architecture of the neural network model. In the task of classifying defects from vibration data converted to images, the choice of neural network architecture is important to achieve high accuracy. Among the various options, such as ResNet50, VGG, Inception, MobileNet, and DenseNet [11], the most reasonable solution is to use ResNet50. This architecture is characterized by its ability to maintain performance even with significant network depth, which is achieved by using residual blocks. Residual blocks avoid the problem of gradient fading, which ensures stable training and efficient feature extraction, even when the data is complex or contains implicit patterns [12], as in the case of time signals of vibration data.

In addition, ResNet50 has a relatively moderate computational complexity compared to heavier models such as DenseNet, making it more practical for applications with limited computing resources [12]. Compared to models designed for lower computational cost, such as MobileNet, ResNet50 provides significantly higher accuracy, which is important for flaw detection where minimizing false positives is important, such as in rotating machinery. In general, thanks to the balance between accuracy and efficiency, as well as the ability to highlight complex features of the input data, ResNet50 is the optimal choice for the task of classifying defects from vibration signal analysis.

The second step is to configure the parameters of the chosen neural network architecture. First of all, the model architecture includes the number of layers and their configuration, which determines the depth and complexity of the network. Each layer contains a certain number of neurons involved in extracting features from the input data. It

is also important to use convolutional layers, where the size of filters is customized. For example, the classic ResNet50 model also includes certain hyperparameters, such as the learning rate and the number of epochs. The learning rate determines how quickly the model adapts to new data during training, and the number of epochs indicates how many times the model will go through the entire training set [13]. It is also possible to define a loss function, in this case the loss function is a binary cross-entropy, which allows you to measure the quality of model predictions and adjust its weights during training.

The third step involves training the neural network. For training, 926 samples from the dataset are used. As a result of training, a neural network model is obtained, and then the model's effectiveness is evaluated in step 4. To evaluate the model's performance, various metrics are used, such as the confusion matrix, Accuracy, Precision, Recall, F1-score, and additionally, Accuracy and Loss graphs are built [14]. If the results are unsatisfactory, you should return to step 2. If the results of the metrics are satisfactory, a trained model is obtained for neural network defect detection based on the analysis of rotating machine vibrations.

Thus, these steps allow you to get $Model_{\theta}()$, for the above method of neural network defect detection based on the analysis of vibrations of rotating machines.

4 EXPERIMENTS

To test the described method of neural network detection of defects based on the analysis of rotating machines vibrations, a software application has been developed that is available for download and deployment via the link on GitHub (https://github.com/OlenaSobko/Sobko_Dydo/). To develop the software application and conduct experimental studies, we used hardware with the following parameters:

- Intel Core i5 processor;
- the amount of RAM is 8 GB;
- Software:
 - Python 3.11.6 interpreter;
 - Library scikit-learn 1.3.2, imbalanced-learn 0.12.4, TensorFlow 2.14.0, Pillow (PIL) 10.1.0, NumPy 1.26.2, Flask 3.0.3;
 - Windows 10 Home version 22H2 (kernel version 10.0.19045).

The software application was developed as a web application (Figure 5), and the interaction of its components is shown in Figure 6.

The diagram illustrates the layout of the components of a software application for detecting defects in rotating machines using vibration analysis. One of the input data is a data dataset containing vibration signals characterizing the state of the rotating machine. This data is processed and visualized using the Matplotlib library to create a graphical dataset that presents visual information about the vibrations. Next, a graphical dataset is passed to the component and the neural network model is trained. After that, a trained ResNet50 model is obtained, to which the

user can provide either a graphical representation of vibrations from sensors or a numerical one in the form of a text file to detect defects in rotating machines. The web application is an “intermediary” between the user and the neural network model, which allows to upload input data from the user and provide a conclusion on the presence of defects in rotating machines.

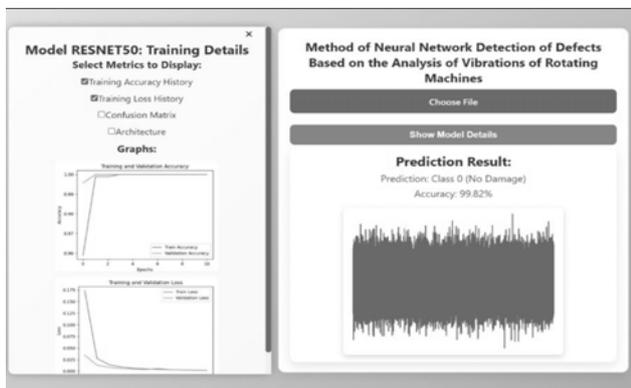


Figure 5 – A web application for testing the method of neural network defect detection based on the analysis of vibrations of rotating machines

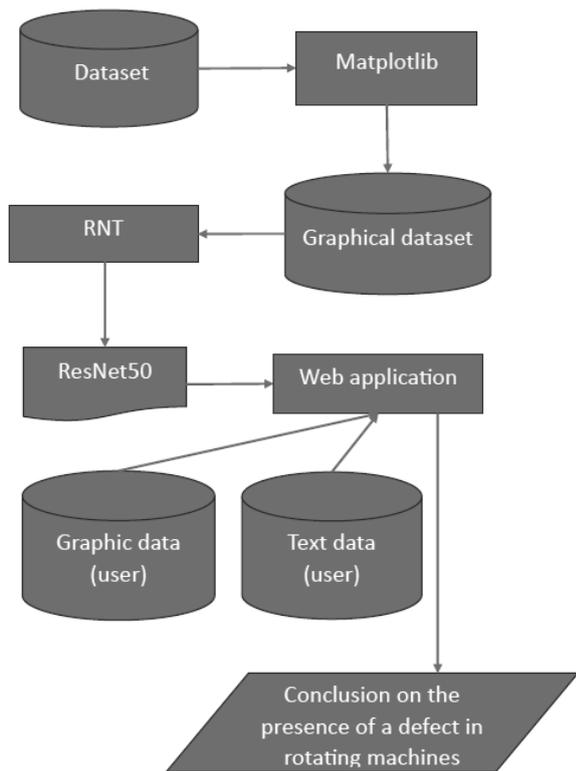


Figure 6 – Diagram of the components of the software application

To test the method of neural network detection of defects based on the analysis of rotating machines vibrations, a neural network model based on ResNet50 was obtained according to the steps of Figure 3, the architecture of which is shown in Figure 7.

The input is a 224×224 image with three channels corresponding to RGB color components. The first layer is the input layer, which accepts data in the format $(None, 224, 224, 3)$, where None indicates the size of the packet.

The next component is the ResNet50 kernel, which consists of pre-trained deep neural network layers. This part of the network converts input images into high-level features, and its output tensor has the dimension $(None, 7, 7, 2048)$, where 2048 is the number of extracted features. After that, a global feature averaging layer is applied (GlobalAveragePooling2D), which aggregates features from the entire field of view, reducing the dimensionality to $(None, 2048)$, which reduces the amount of data and avoids overtraining by keeping only the key generalized characteristics.

Next, two Dense layers are added to perform classification functions. The first dense layer with 256 neurons takes an input of 2048 features and performs a linear transformation with nonlinear activation, which helps to extract and classify relevant defect features. The last dense layer contains only one neuron with an activation function, which allows for the final binary classification that determines the presence or absence of a defect in the system.

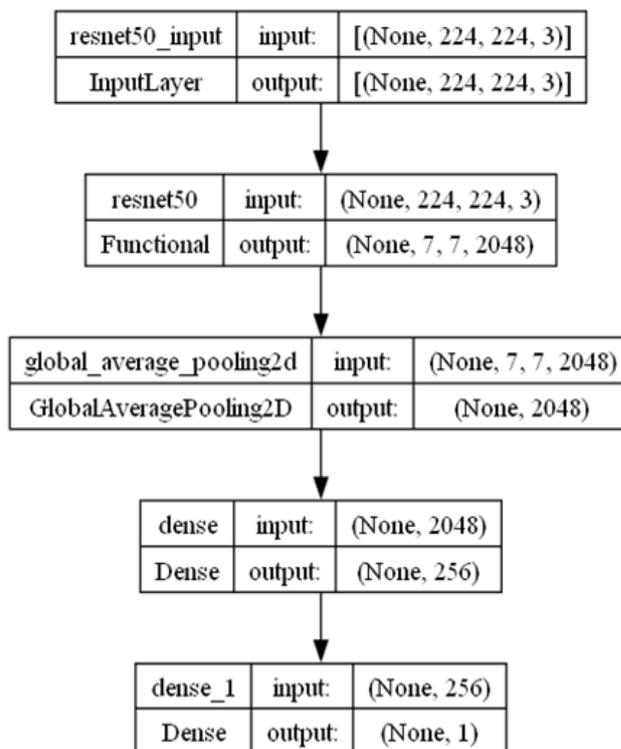


Figure 7 – Architecture of the resulting neural network model based on ResNet50

As for the dataset used for testing, in this case, randomly by means of scikit-learn (<https://scikit-learn.org/stable/>) 926 samples were selected for training, of which 689 were “without defect”, and 237 were “with

defect” and 232 for validation, of which 176 were “without defect” and 56 were “with defect”.

To study the effectiveness of the proposed method, the following steps should be taken:

1. Build Accuracy and Loss graphs.
2. Build the entanglement matrix.
3. Calculate metrics Accuracy, Precision, Recall, F1-score.
4. Conduct a cross-validation test to assess the accuracy of the obtained neural network model.

The results of the study are presented below in Section 5.

5 RESULTS

The Accuracy and Loss graphs, presented in Figure 8 below, serve as fundamental tools for assessing the performance of a machine learning model throughout the training process. These graphs offer a detailed and dynamic representation of the model’s progression as it learns from the training data over successive epochs. By tracking the changes in accuracy and loss values during training, they provide insights into the effectiveness of the model’s learning process, the behavior of the optimization algorithm, and the ability of the model to generalize its learned patterns to unseen data.

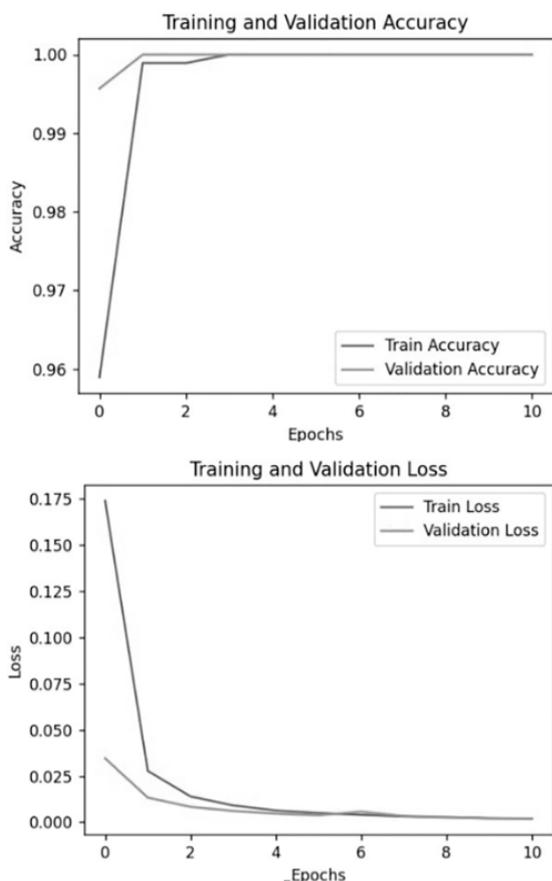


Figure 8 – Graphs of accuracy and losses of the neural network model based on ResNet50

Next, we constructed a confusion matrix for the validation data, which is shown on Figure 9.

As a result, the model received the following metric values: Accuracy 1.0, Precision 1.0, Recall 1.0, F1-Score 1.0.

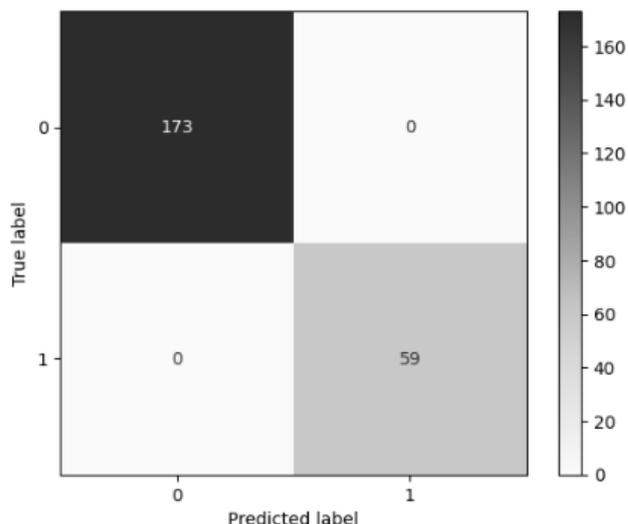


Figure 9 – Confusion matrix for validation data

The next step is cross-validation. Cross-validation is a method of assessing the accuracy of machine learning models by dividing data into several parts, or folds. In the classical case, for example, in k-fold cross-validation, the dataset is divided into k equal parts. The model is trained on k-1 parts of the data and then tested on the remaining part. This process is repeated k times, each time using a different fold for validation and the rest for training. After that, the results of each cycle are averaged to get an overall assessment of the model’s performance [15]. The results of such cross-validation for the ResNet50 neural network model are shown in Table 1.

Table 1 – Metrics results for different folds

Fold	Accuracy	Precision	Recall	F1 Score
1	0.9855	0.9773	0.9806	0.9855
2	0.9982	0.9974	0.9913	0.9943
3	0.9986	0.9999	0.9913	0.9955
4	0.9838	0.9828	0.9896	0.9791
5	0.9887	0.9870	0.9851	0.9861
6	0.9904	0.9971	0.9988	0.9980
7	0.9841	0.9918	0.9894	0.9840
8	0.9986	0.9960	0.9985	0.9972
9	0.9886	0.9876	0.9844	0.9823
10	0.9945	0.9943	0.9981	0.9962

The average values of five folds are Accuracy 0.9911, Precision 0.9911, Recall 0.9907, F1-Score 0.9898, and the deviations from the initially calculated metric values are obtained: Accuracy 0.0089, Precision 0.0089, Recall 0.0093, F1-Score 0.0102.

Compared to the works considered [8, 9, 10] in Section 2, the trained neural network model for the method of neural network detection of defects based on the analysis

of rotating machines vibrations received a higher accuracy rate of 99.11% than in the cited works (Accuracy rates 97.11%, 95%, 99%).

Further, in Section 6, we discuss the results of evaluating the effectiveness of the obtained neural network model for testing the method of neural network defect detection based on the analysis of rotating machines' vibrations.

6 DISCUSSION

The accuracy and loss plots in Figure 8 show that the model's accuracy increases rapidly and reaches 1.0 already at the initial epochs. The accuracy on the training and validation sets is almost identical, which indicates that they are consistent. The losses decrease quite quickly, reaching very low values after a few epochs. The difference between the losses on the training and validation sets is insignificant, indicating that the model does not overlearn.

The resulting confusion matrix demonstrates that the model achieved perfect results on the validation data, as all model predictions are correct. The model classified 173 samples as class "0", i.e. without defects and 59 samples as a class "1", i.e. with defects, without making a single error. All model quality indicators, including Accuracy, Precision, Recall, and F1-Score, have a value of 1.0, which indicates no false positives and full coverage of both classes. This indicates the perfect ability of the model to distinguish between classes on the validation data, i.e., it not only correctly predicts class membership, but also does not allow false classifications, fully covering both classes.

To confirm the model's effectiveness, a cross-validation was conducted to ensure that the high performance was the result of a good generalization and not a coincidence. The results of the cross-validation demonstrate the model's consistently high performance across all key metrics in each of the five folds. Accuracy in all folds ranges from 0.9838 to 0.9887, indicating a high generalization ability of the model. Similarly, Precision, Recall, and F1 Score in each fold are close to 1.0, which confirms the absence of a significant number of false positives and false negatives.

The average metric values for the five folds are: Accuracy – 0.98614, Precision – 0.9853, Recall – 0.98582, and F1-Score – 0.9834, which once again confirms the overall reliability and efficiency of the model. A small deviation of the metrics from the initially selected values indicates the stability of the model, since these deviations are minimal and do not significantly affect its overall performance.

Thus, the neural network model obtained as a result of the experimental study is focused on detecting defects by analyzing the vibrations of rotating machines without preliminary noise removal. The ResNet50 base model is responsible for extracting high-level features, which allows the neural network to learn with a high level of detail. The use of a global feature averaging layer reduces the number of parameters, which reduces the risk of

overfitting, allowing for better generalization of information. Due to the sequence of dense layers at the output, the network is able to convert the selected features into a binary decision (for example, the presence or absence of a defect), which is the result of the task. The above is confirmed by an experiment to study the effectiveness of the resulting model.

The study demonstrates that the proposed method of neural network defect detection based on the analysis of vibrations of rotating machines allowed us to create a neural network model as a result of the experiment, which achieved high accuracy and consistency between training and validation data, as evidenced by higher Accuracy, Precision, Recall, F1-Score indicators compared to [8, 9] and [10].

CONCLUSIONS

The paper solves an urgent problem of neural network defect detection based on the analysis of rotating machine vibrations.

The scientific novelty is to develop a method of neural network detection of defects by analyzing the vibrations of rotating machines based on the ResNet50 neural network model, which achieves high results of accuracy of defect detection by analyzing the vibrations of rotating machines. A distinctive feature of the method is that it does not require preliminary cleaning of vibration signals from noise, which is achieved by adjusting the parameters of the neural network architecture.

The practical significance the results obtained indicate that the method proposed in this work and the developed software can be used to detect defects by analyzing the vibrations of rotating machines, since it showed high performance on the proposed dataset.

Prospects for further research is to test the resulting neural network model on larger data sets.

ACKNOWLEDGEMENTS

The paper was prepared within the framework of the All-Ukrainian hackathon competition for young scientists in the field of intellectual information technologies in 2024 at Zaporizhzhia National Technical University. We are grateful to the organizers of the hackathon for providing the data set for testing the proposed method.

REFERENCES

1. Maurya M., Panigrahi I., Dash D., Malla C. Intelligent fault diagnostic system for rotating machinery based on IoT with cloud computing and artificial intelligence techniques: a review, *Soft Computing*, 2023, Vol. 28, No. 1, pp. 477–494. DOI: 10.1007/s00500-023-08255-0.
2. Shaalan A. A., Mefteh W., Frihida A. M. Review on deep learning classifiers for faults diagnosis of rotating industrial machinery, *Service Oriented Computing and Applications*, 2024, pp. 1–19. DOI: 10.1007/s11761-024-00418-7.
3. Souza R. M., Nascimento E. G., Miranda U. A., Silva W. J., Lepikson H. A. Deep learning for diagnosis and classification of faults in industrial rotating machinery, *Computers & Industrial Engineering*, 2021, Vol. 153, P. 107060. DOI: 10.1016/j.cie.2020.107060.

4. Yadav E., Chawla V. K. Role and Significance of Defect Detection Methods for Rotating Machines: An Explicit Literature Review, *Journal of The Institution of Engineers (India): Series C*, 2024, pp. 1–18. DOI: 10.1007/s40032-024-01109-0.
5. Tama B. A., Vania M., Lee S., Lim S. Recent advances in the application of deep learning for fault diagnosis of rotating machinery using vibration signals, *Artificial Intelligence Review*, 2023, Vol. 56, No. 5, pp. 4667–4709. DOI: 10.1007/s10462-022-10293-3.
6. Kafeel A., Aziz S., Awais M., Khan M. A., Afaq K., Idris S. A., Mostafa S. M. An expert system for rotating machine fault detection using vibration signal analysis, *Sensors*, 2021, Vol. 21, No. 22, p. 7587. DOI: 10.3390/s21227587.
7. Mian T., Choudhary A., Fatima S., Panigrahi B. K. Artificial intelligence of things based approach for anomaly detection in rotating machines, *Computers and Electrical Engineering*, 2023, Vol. 109, P. 108760. DOI: 10.1016/j.compeleceng.2023.108760.
8. Ong P., Tan Y. K., Lai K. H., Sia C. K. A deep convolutional neural network for vibration-based health-monitoring of rotating machinery, *Decision Analytics Journal*, 2023, Vol. 7, p. 100219. DOI: 10.1016/j.dajour.2023.100219.
9. Zhen J. Rotating Machinery Fault Diagnosis Based on Adaptive Vibration Signal Processing under Safety Environment Conditions, *Mathematical Problems in Engineering*, 2022, Vol. 2022, No. 1, P. 1543625. DOI: 10.1155/2022/1543625.
10. Aziz S., Ahmed M., Abbas I., Naqvi S. Z. H., Khan M. U. Vibration Signal Analysis Towards Early Detection of Machine Faults, *2020 First International Conference of Smart Systems and Emerging Technologies (SMARTTECH)*, Riyadh, Saudi Arabia, 2020, pp. 89–93. DOI: 10.1109/SMART-TECH49988.2020.00034.
11. MobileNet vs ResNet50 – Two CNN Transfer Learning Light Frameworks, Analytics India Magazine. [Electronic resource]. – Access mode: <https://analyticsindiamag.com/ai-mysteries/mobilenet-vs-resnet50-two-cnn-transfer-learning-light-frameworks>.
12. Tchatchoua P., Graton G., Ouladsine M., Christaud J. F. Application of 1D ResNet for Multivariate Fault Detection on Semiconductor Manufacturing Equipment, *Sensors*, 2023, Vol. 23, No. 22, P. 9099. DOI: 10.3390/s23229099.
13. Improving the performance of ResNet50 Graffiti Image Classifier with Hyperparameter Tuning in Keras, Towards Data Science. [Electronic resource]. – Access mode: <https://towardsdatascience.com/improving-the-performance-of-resnet50-graffiti-image-classifier-with-hyperparameter-tuning-in-keras-dbb59f43c6f7>.
14. Slobodzian V., Molchanova M., Kovalchuk O., Sobko O., Mazurets O., Barmak O., Krak I. An Approach Based on the Visualization Model for the Ukrainian Web Content Classification, *2022 12th International Conference on Advanced Computer Information Technologies (ACIT)*, 2022, pp. 400–405. DOI: 10.1109/ACIT54803.2022.9913162.
15. Bates S., Hastie T., Tibshirani R. Cross-validation: what does it estimate and how well does it do it?, *Journal of the American Statistical Association*, 2024, Vol. 119, No. 546, pp. 1434–1445. DOI: 10.1080/01621459.2023.2197686.

Received 16.12.2024.
Accepted 04.02.2025.

УДК 004.8

МЕТОД НЕЙРОМЕРЕЖЕВОЇ ВИЯВЛЕННЯ ДЕФЕКТІВ НА ОСНОВІ АНАЛІЗУ ВІБРАЦІЙ ОБЕРТОВИХ МАШИН

Собко О. В. – аспірант кафедри комп'ютерних наук Хмельницького національного університету, м. Хмельницький, Україна.

Дидо Р. А. – студент кафедри комп'ютерних наук Хмельницького національного університету, м. Хмельницький, Україна.

Мазурець О. В. – канд. техн. наук, доцент кафедри комп'ютерних наук Хмельницького національного університету, м. Хмельницький, Україна.

АНОТАЦІЯ

Актуальність. У роботі пропонується вирішення актуальної проблеми виявлення дефектів обладнання за аналізом вібрацій обертових машин. Об'єктом дослідження є процес виявлення дефектів за аналізом вібрацій обертових машин. Предметом дослідження є методи штучного інтелекту для виявлення дефектів за аналізом вібрацій обертових машин.

Мета роботи. Підвищення точності виявлення дефектів за аналізом вібрацій обертових машин шляхом створення методу нейромережевого виявлення дефектів за аналізом вібрацій обертових машин та відповідної нейромережевої моделі, яка здатна виявляти дефекти за аналізом вібрацій обертових машин без попереднього видалення шумів з метою збереження важливих ознак для точнішої класифікації.

Метод. Запропоновано метод нейромережевої виявлення дефектів на основі аналізу вібрацій обертових машин, який здатен за вхідними даними вібрацій з виконанням попередньої обробки, а саме створення двовимірного часово-частотного зображення, зробити прогноз щодо наявності чи відсутності дефекту. Метод відрізняється від існуючих тим, що аналіз на дефекти проводиться без видалення шумів за рахунок тонкого налаштування параметрів моделі.

Результати. Запропонований у роботі метод нейромережевої виявлення дефектів на основі аналізу вібрацій обертових машин реалізовано у вигляді беззастосунку та проведено дослідження ефективності нейромережевої моделі, отриманої шляхом виконання кроків методу.

Висновки. Результати дослідження показують, що модель досягла високої точності та узгодженості між тренувальними та валідаційними даними, що підтверджується високими значеннями таких показників, як Accuracy = 1.0, Precision = 1.0, Recall = 1.0 і F1-Score = 1.0 на валідаційному наборі даних, а також мінімальними втратами. Проведена крос-валідація підтвердила стабільну ефективність моделі, продемонструвавши високі усереднені метрики та незначні відхилення від отриманих метрик. Таким чином, нейромережева модель виявляє дефекти обертових машин з високою ефективністю навіть без очищення вібраційних сигналів від шумів. Перспективи подальших досліджень полягають в апробації описаного метода та отриманої нейромережевої моделі на більших наборах даних.

КЛЮЧОВІ СЛОВА: дефекти, аналіз, вібрації, обертові машини, нейромережа, ResNet50.

ЛІТЕРАТУРА

1. Intelligent fault diagnostic system for rotating machinery based on IoT with cloud computing and artificial intelligence techniques: a review / [M. Maurya, I. Panigrahi, D. Dash, C. Malla] // *Soft Computing*. – 2023. – Vol. 28, № 1. – P. 477–494. DOI: 10.1007/s00500-023-08255-0
2. Shaalan A. A. Review on deep learning classifiers for faults diagnosis of rotating industrial machinery / A. A. Shaalan, W. Meftteh, A. M. Frihida // *Service Oriented Computing and Applications*. – 2024. – P. 1–19. DOI: 10.1007/s11761-024-00418-7
3. Deep learning for diagnosis and classification of faults in industrial rotating machinery / [R. M. Souza et al.] // *Computers & Industrial Engineering*. – 2021. – Vol. 153. – P. 107060. DOI: 10.1016/j.cie.2020.107060
4. Yadav E. Role and Significance of Defect Detection Methods for Rotating Machines: An Explicit Literature Review / E. Yadav, V. K. Chawla // *Journal of The Institution of Engineers (India): Series C*. – 2024. – P. 1–18. DOI: 10.1007/s40032-024-01109-0
5. Recent advances in the application of deep learning for fault diagnosis of rotating machinery using vibration signals / [B. A. Tama, M. Vania, S. Lee, S. Lim] // *Artificial Intelligence Review*. – 2022. – Vol. 56, № 5. – P. 4667–4709. DOI: 10.1007/s10462-022-10293-3
6. An Expert System for Rotating Machine Fault Detection Using Vibration Signal Analysis / [A. Kafeel, S. Aziz, M. Awais et al.] // *Sensors*. – 2021. – Vol. 21, No. 22. – P. 587. DOI: 10.3390/s21227587
7. Artificial intelligence of things based approach for anomaly detection in rotating machines / [T. Mian, A. Choudhary, S. Fatima, B. K. Panigrahi] // *Computers and Electrical Engineering*. – 2023. – Vol. 109. – P. 108760. DOI: 10.1016/j.compeleceng.2023.108760
8. A deep convolutional neural network for vibration-based health-monitoring of rotating machinery [Electronic resource] / [P. Ong, Y. K. Tan, K. H. Lai, C. K. Sia] // *Decision Analytics Journal*. – 2023. – P. 100219. DOI: 10.1016/j.dajour.2023.100219
9. Zhen J. Rotating Machinery Fault Diagnosis Based on Adaptive Vibration Signal Processing under Safety Environment Conditions / J. Zhen // *Mathematical Problems in Engineering*. – 2022. – Vol. 2022, № 1. – P. 1543625. DOI: 10.1155/2022/1543625
10. Vibration Signal Analysis Towards Early Detection of Machine Faults / [S. Aziz, M. Ahmed, I. Abbas et al.] // *2020 First International Conference of Smart Systems and Emerging Technologies (SMARTTECH)*, Riyadh, 3–5 November 2020. – [S. I.], 2020 – P. 89–93. DOI: 10.1109/SMARTTECH49988.2020.00034
11. MobileNet vs ResNet50 – Two CNN Transfer Learning Light Frameworks [Electronic resource] // *Analytics India Magazine*. – Mode of access: <https://analyticsindiamag.com/ai-mysteries/mobilenet-vs-resnet50-two-cnn-transfer-learning-light-frameworks>
12. Application of 1D ResNet for Multivariate Fault Detection on Semiconductor Manufacturing Equipment [Electronic resource] / [P. Tchatchoua, G. Graton, M. Ouladsine, J. F. Christaud] // *Sensors*. – 2023. – Vol. 23, No. 22. – P. 9099. DOI: 10.3390/s23229099
13. Konda S. Improving the performance of ResNet50 Graffiti Image Classifier with Hyperparameter Tuning in Keras [Electronic resource] / S. Konda // *Medium*. – Mode of access: <https://towardsdatascience.com/improving-the-performance-of-resnet50-graffiti-image-classifier-with-hyperparameter-tuning-in-keras-dbb59f43c6f7>
14. An Approach Based on the Visualization Model for the Ukrainian Web Content Classification / [Slobodzian V., Molchanova M., Kovalchuk O. et al.] // *2022 12th International Conference on Advanced Computer Information Technologies (ACIT)*, Ruzomberok, Slovakia, 26–28 September 2022. – [S. I.], 2022. – P. 400–405. DOI: 10.1109/ACIT54803.2022.9913162
15. Bates S. Cross-validation: what does it estimate and how well does it do it? / S. Bates, T. Hastie, R. Tibshirani // *Journal of the American Statistical Association*. – 2023. – P. 1–22. DOI: 10.1080/01621459.2023.2197686

DATA-DRIVEN DIAGNOSTIC MODEL BUILDING FOR HELICOPTER GEAR HEALTH AND USAGE MONITORING

Subbotin S. A. – Dr. Sc., Professor, Head of the Department of Software Tools, National University “Zaporizhzhia Polytechnic”, Zaporizhzhia, Ukraine.

Bechhoefer E. – PhD, CEO and Founder, GPMS Inc., USA.

ABSTRACT

Context. Modern technical objects (in particular vehicles) are extremely complex and place high demands on reliability. This requires automation of condition monitoring and fault diagnosis of objects and their components. The predictive maintenance improves operational readiness of technical objects. The object of study is a technical object health and usage monitoring process. The subject of study is a methods of computational intelligence for data-driven model building and related data processing tasks for health and usage monitoring system.

Objective. The purpose of the work is to formulate data processing problems, to form a data set for data-driven model building and construct simple method for automatic diagnostic model building on example of helicopter health and usage monitoring system.

Method. The method is proposed for the mapping of multidimensional data into a two-dimensional space preserving local properties of class separation, allowing for the visualization of multidimensional data and the production of simple diagnostic models for the automatic classification of diagnostic objects. The proposed method allows obtaining highly accurate diagnostic model with small training samples, provided that the frequency of classes in the samples is preserved. A method for synthesizing diagnostic models based on a two-layer feed-forward neural network is also proposed, which allows obtaining models in a non-iterative mode.

Results. A sample of observations of the state of helicopter gears was obtained, which can be used to compare data-driven diagnostic methods and data processing methods that solve the problems of data dimensionality reduction. The Software has been developed that allows displaying a sample from a multidimensional to a two-dimensional space, which makes it possible to visualize data and reduces the dimensionality of the data. Diagnostic models have been obtained that allow automating the decision-making process on whether the diagnosed object (helicopter gear) belongs to one of two classes of states.

Conclusions. The results of conducted experiments allow to conclude that the proposed method provides a significant reduction in the data dimensionality (in particular, for the considered problem of constructing a model for helicopter gear diagnosis, it reduces the data dimensionality due to the compression of features by 46876 times). As the results of the conducted experiments for randomly selected instances in a two-dimensional system of artificial features obtained on the basis of the proposed method showed a significant reduction of the sample for individual tasks may allow to provide acceptable accuracy. And taking into account individual estimates of the instance significance will allow, even for small samples, to ensure the topological representativeness of the formed sample in relation to the original sample.

The prospects for further research are to compare methods for constructing data-driven models, as well as methods for reducing the dimensionality of data based on the proposed sample. Additionally, it may be of interest to study a possible combination of the proposed method with methods for sample forming using metrics of the value of instances.

KEYWORDS: data-driven diagnosis, health and usage monitoring system, data dimensionality reduction, classification.

ABBREVIATIONS

DM is a diagnostic model;
HUMS is a Health and Usage Monitoring System;
MAE is a mean absolute error;
MSE is a mean squared error;
SSE is a sum squared error.

NOMENCLATURE

η is a neural network layer number;
 $\psi^{(\eta,i)}$ is an activation function of i -th neuron of η -th layer;
 $\varphi^{(\eta,i)}$ is a weight (postsynaptic) function of i -th neuron of η -th layer;
[] is a designation of a non-obligatory parameter;
< > is a designation of a tuple;
 C is a center relative to which the distance is determined from the instances;
 C^+ is a sample center;
 C^{avg} is an average between class centers;
 C^q is a q -th class center;
 C_j^q is a value of j -th coordinate (feature) of the q -th class center;

$d(a, b)$ is a distance between points a and b ;
 E is a model error function;
 f is a criterion (goal function);
 $F()$ is a diagnostic (recognition) model structure;
 F_i is an i -th model structure;
 $f()$ is a user criterion characterizing the argument quality relatively to the problem being solved;
 j is a feature number;
 K is a number of classes;
 M is a number of layers in neural network;
 N is a number of input features;
 N' is a number of features in a reduced set (subsample);
 N_η is a number of neurons in the η -th layer of neural network;
opt is an optimal (desired or acceptable) value of the functional $f()$ for the problem being solved;
 $P_{correct}$ is a correct decision making probability;
 $P_{incorrect}$ is an incorrect decision making probability;
 $P_{incorrect\ test}$ is a incorrect decision making probability for a test set;

$P_{\text{incorrect train}}$ is a incorrect decision making probability for a training set;
 q is a class number;
 S is a number of instances in the original sample;
 s is an instance number;
 S' is a number of precedents (examples) in a reduced sample (subsample);
 S^q is a number of exemplars of a sample belonging to the q -th class;
 S_{test} is a volume of a test set;
 S_{train} is a volume of a training set;
 w is a set of controlled (adjusted) parameters of a model;
 w' is a set of controlled (adjusted) parameters of a reduced model;
 $w_j^{(\eta,i)}$ is a value of j -th adjustable parameter (or weight) of j -th input of i -th node of η -th layer of a neural network;
 $w^{(\eta,i)}$ is a set adjustable parameters (or weights) of i -th node of η -th layer of a neural network;
 x' is a input features of reduced set of instances or features;
 x is a set of instances (examples, cases);
 x_j is a j -th input feature;
 x'_j is a j -th input feature is a reduced set;
 x^s is an s -th instance of a sample;
 x_j^s is a value of j -th input (descriptive) feature of s -th instance;
 $x_j^{(\eta,i)}$ is a j -th input value of i -th neuron of η -th layer of a neural network;
 y is a set of values of output feature;
 y' is an output feature for reduced set of instances;
 y^s is a value of an output feature for s -th instance x^s ;
 y^{s*} is a calculated output feature value for the s -th instance on the model output;
 y_i^s is a value of i -th output feature of s -th instance;
 $y^{(\eta,i)}$ is an output value of i -th neuron of η -th layer of a neural network.

INTRODUCTION

Modern technical objects (in particular vehicles) are extremely complex and place high demands on reliability. This requires automation of condition monitoring and fault diagnosis of objects and their components. The predictive maintenance improves operational readiness of technical objects [1, 2].

The **object of study** is a technical object health and usage monitoring process.

The process of technical object health and usage monitoring involves measuring equipment characteristics, processing measurement data and determining the equipment state [3]. This may be provided by the HUMS, which is a sensor-based system that measure the health and performance of mission-critical components in diagnosed objects. It provides actionable information so that maintainers can make data-informed decisions [1].

The HUMS are widely used for condition monitoring of rotating equipment such as helicopters trucks or power plants. Typically, for vehicles the sensors placed on the

equipment such as the transmission or the engine and then vibration measured and used to determine the state of the equipment or it's details [2].

The HUMS for helicopters may provide such benefits as: reduction in inspection times, unneeded maintenance, unscheduled and scheduled maintenance, unscheduled flight cancellation, test flights, vibration related wear damage and avionics removals, and increasing in safety, sense of safety, performance, mission, confidence, ease of troubleshooting, and morale [4].

The basis for automation of decision making in HUMS is the use of DM. The DM is a formalized description of diagnosing object required to solve the problems of diagnosis [3, 5].

To build diagnostic systems (particularly HUMS), it is possible to use the paradigms of model-based diagnosis (requires the presence of an analytical or physical model of the diagnosed object) [5] or data-driven diagnosis (based on the AI-model created on the set of observations) [6, 7].

Since vehicles are extremely complex systems that operate in changing conditions of an aggressive environment and are characterized by variability, the creating accurate analytical (physical) models for them is an extremely complex task. Therefore, in practice, the use of data-driven diagnosis methods is more convenient.

The **subject of study** is a methods of computational intelligence for data-driven model building and related data processing tasks for HUMS.

Model building based on computational intelligence [7] requires the availability of a sample of observations containing instances (cases, exemplars) of good and defective items or details.

The equipment faults are relatively rare in a practice. So the class-imbalanced training dataset should be used to train the model for fault detection.

The **purpose of the work** is to formulate data processing problems, to form a data set for data-driven model building and construct simple method for automatic DM building on example of helicopter HUMS.

1 PROBLEM STATEMENT

The instance (case, exemplar) is a set of data that describes the observation of an object (process) at a particular point in time. Instances are characterized by the values of features (attributes, properties).

The training sample is a sample on the basis of which the DM is built. The test sample is a sample used to test the performance of the DM.

The input data for DM building is a training sample $\langle x, y \rangle$ characterized by N features set $\{x_j\}, j = 1, 2, \dots, N$, consisting of S instances (examples, cases): $x = \{x^s\}, s = 1, 2, \dots, S$, each of which described by the feature values $x^s = \{x_j^s\}$, where x_j^s is a value of j -th input (descriptive) feature of s -th instance, $y = \{y^s\}$ is a set of values of output feature, where y^s is a value of an output feature for s -th instance [7, 8].

The problem of dependency approximation: given: recognized instance x^s , model of the dependence $y=f(x)$,

find: y^s the estimated value of output feature for x^s . If the output variable is continuous then the problem of dependency approximation is a output real value estimation (regression). If the output variable is discrete then the problem of dependency approximation is a classification (pattern recognition).

During the process of DM building we typically need to address some of the following problems.

The problem of feature selection [9]: given: $\langle x, y \rangle$, find: $\langle x', y \rangle, x' \in \{x_j\}, N' < N, S' = S, f(\langle x', y \rangle, \langle x, y \rangle) \rightarrow \text{opt}$.

The feature extraction problem (construction of artificial features): given: $\langle x, y \rangle$, find: $\langle x', y \rangle, x' = \{x'_i\}, x'_i = F_i(\{x_j\}), S' = S, f(\langle x', y \rangle, \langle x, y \rangle) \rightarrow \text{opt}$.

The instance selection problem [9]: given: $\langle x, y \rangle$, find: $\langle x', y' \rangle, x' \in \{x^s\}, y' = \{y^s | x^s \in x'\}, S' < S, N' = N, f(\langle x', y' \rangle, \langle x, y \rangle) \rightarrow \text{opt}$.

The problem of DM synthesis (model building) [8]: given: $\langle x, y \rangle$, find: $\langle F(), w \rangle: y^s = F(w, x^s), f(F(), w, \langle x, y \rangle) \rightarrow \text{opt}$.

Particularly, for the case of feed-forward neural or integrated feed-forward neuro-fuzzy network basis the $\langle F(), w \rangle$ defined as $\langle M, \{N_\eta\}, \{y^{(\eta,i)}(x^{(\eta,i)})\} \rangle$ and described by the $y_i^s = y^{(M,i)}(y^{(M-1,i)}(\dots y^{(1,i)}(x^s)))$, $i=1,2,\dots,N_{\eta-1}$, $w^{(\eta,i)} = \{w_j^{(\eta,i)}\}$, $y^{(0,j)} = \psi^{(0,j)} = x_j^s$, $N_0 = N$, $x_j^{(1,i)} = x_j^s$, $i=1,2,\dots,N_{\eta-1}$, $\eta=1,2,\dots,M$, $j=1,2,\dots,N$.

Here $\{y^{(\eta,i)} = \psi^{(\eta,i)}(w^{(\eta,i)})\}$ and $w = \{w_j^{(\eta,i)}\} = \{w_j^{(\eta,i)}\}$ are structural blocks and parameters of DM based on neural network, respectively.

The problem of DM simplification: given: $[\langle x, y \rangle, \langle F(), w \rangle]$, find: $\langle F'(), w' \rangle: f(\langle F'(), w' \rangle, \langle F(), w \rangle, \langle x, y \rangle) \rightarrow \text{opt}$.

The problem of DM additional training: given: $\langle x', y' \rangle, \langle x, y \rangle, \langle F(), w \rangle$, find: $\langle F'(), w' \rangle: f(\langle F'(), w' \rangle, \langle F(), w \rangle, \langle x', y' \rangle, \langle x, y \rangle) \rightarrow \text{opt}$.

For each problem considered above, it is possible to use a wide range of criteria for DM quality (performance) estimation [10, 11]. However, in practice, criterion f usually defined on the base of model error function E .

In general case (for discrete and continuous outputs) we can use:

– MSE:

$$E = \frac{1}{S} \sum_{s=1}^S (y^s - F(x^s))^2;$$

– SSE:

$$E = \sum_{s=1}^S (y^s - F(x^s))^2;$$

– MAE:

$$E = \frac{1}{S} \sum_{s=1}^S |y^s - F(x^s)|.$$

In case of discrete output feature we can use an error formula:

$$E = \sum_{s=1}^S \{1 | y^s \neq F(x^s)\} \rightarrow \min.$$

For binary and discrete output features we can compute the probability estimates of decisions:

- of incorrect decision making: $P_{\text{incorrect}} = E/S$;
- of correct decision making: $P_{\text{correct}} = 1 - P_{\text{incorrect}}$.

2 REVIEW OF THE LITERATURE

The generalized system of technical diagnostics (in particular, HUMS) [1, 5, 6] is shown on the Fig. 1.

The signals from the diagnosed object are measured by the measurement devices and transformed into data in the computer memory through the input devices.

For example, in the tasks of diagnosing helicopter gears, the data set of observations is a large vibration data set.

The signal processing is performed and object's state is determined.

There are two main paradigms for the object's state determination. They are model-based diagnosis [5] and data-driven diagnosis [6].

Since the analytical (physical) model of the diagnosed object state is absent in many practical cases or is not enough accurate due to the lack of expert knowledge, the data-driven diagnostics [6] has become widely used in practice. It is used a sample of observations and based on constructing a model of diagnosed object state using machine learning methods and models (statistical [12], metric [13], soft computing [14, 15], logical [16], rule-based [17], and others). A comparative characteristics description of machine learning methods for constructing data-driven diagnostic models is given in Table 1. Here, for each comparison criterion, cells with the best content are highlighted with a green background, cells with the worst content are highlighted with a red background, and cells with average content are highlighted with a yellow background.

As can be seen from Table 1, there is no group of data-driven methods that would be the best by all criteria. Therefore, for each specific task, based on its characteristics and available resources, it is necessary to specifically solve the problem of selecting applicable methods, as well as the problem of choosing the best method among them.

Additionally, the optimization methods (hybrid, genetic [18], swarm [19], classic direct search and gradient methods [20]) have particular interest for different machine learning problems. These methods have a strong impact on the speed and accuracy of problem solving in the process of constructing a diagnostic model.

A significant impact on simplifying and accelerating the process of diagnostic model constructing, as well as reducing the complexity of the model and increasing its interpretability can be exerted by methods of data dimensionality reduction [21] applied to the sample before the model constructing.

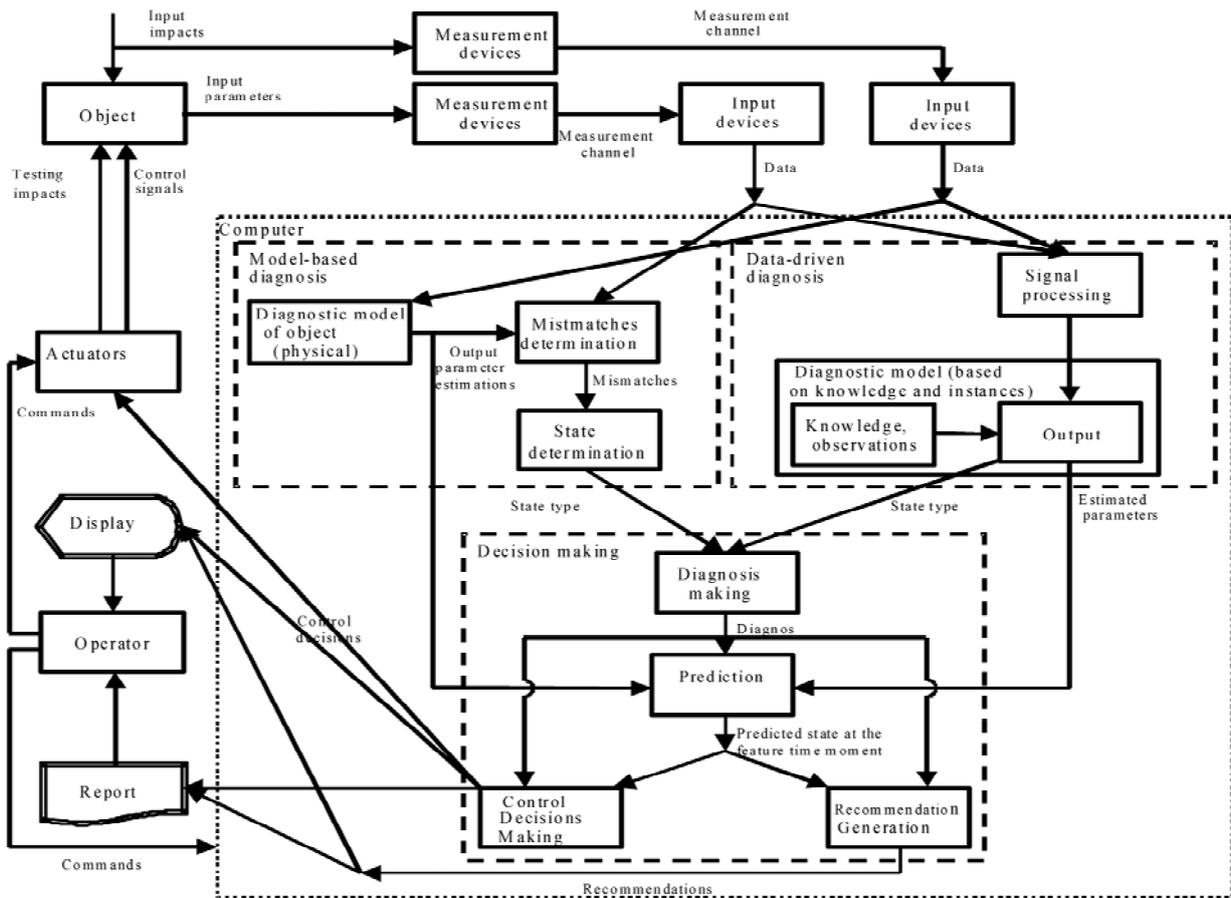


Figure 1 – Schema of generalized system of technical diagnostics

Methods of data dimensionality reduction include methods of the informativeness estimation and selecting informative features (feature selection) [22, 23], methods of artificial features forming (feature extraction) [24, 25] including locality-sensitive hashing [26, 27], significance estimation and selection of instances (sample forming) [28, 29]. In addition, methods of digital processing and signal analysis (Fourier and wavelet transformation, statistical, spectral, cepstral analysis, etc.) [30–33] can be used to reduce the dimensionality of data.

Information about the structure of samples, useful for constructing models and solving related problems, can be provided by cluster analysis methods [34]. Cluster analysis or clustering is the task of grouping a set of objects (instances) in such a way that objects in the same group (called a cluster) are more similar (in some sense) to each other than to those in other groups (clusters).

Additional information can also be provided by data mapping and visualization methods. Known data visualization methods [35] seem to be extremely computationally complex and their application to large volumes of data can be difficult.

This paper considers the problem of creating a simple heuristic method for data mapping and visualization, data dimensionality reduction and DM building,

that does not require solving optimization problems and complex calculations.

3 MATERIALS AND METHODS

Since generally the data sample is multidimensional and the number of features is much greater than three, its visualization is an extremely difficult task. The use of various data transformation methods will also require extremely large computational resources. And the process of building a model will also be extremely resource-intensive.

Therefore, it is proposed to use a combined approach to solving the problem, combining a simple transformation of data from a multidimensional space to a two-dimensional one, which will allow, on the one hand, to visualize a data sample, and, on the other hand, will significantly reduce the dimensionality of the data and will significantly simplify and speed up the process of building diagnostic models, as well as reduce their complexity and the number of adjustable parameters. The model simplifying will also improve their interpretability and generalizing properties.

Let consider a heuristic approach that involves mapping the sample from the original space of N features into a two-dimensional space of artificial features. These features will essentially be similar to locality-sensitive hashes defined without solving the optimization problem, based on the hypothesis of compactness of the arrangement of classes.

Table 1 – Comparative characteristics of the groups of data-driven methods of diagnosis

Group of methods	Statistical	Separation in the feature space	Metrical	Based on soft computing			Based on assumption of class of decision functions	Logical	Rule-based expert systems
				Neural networks	Neuro-fuzzy networks	Fuzzy models			
Input data required for DM synthesis:									
– training set	+	+	+	+	+	–	+	+	–
– human expert knowledge	–	–	–	–	–	+	–	–	+
– model architecture	–	–	–	+	+	+	–	–	–
– other	Density distribution of feature values	–	Metric type	–	–	–	Type and quality functional of the decision function	–	–
Specific requirements to the task	Knowledge of the distribution of feature values and their densities	The need to transform the feature space	High informative features	–	Formation of fuzzy terms and their membership functions	Formation of fuzzy terms and their membership functions	Orthonormality of feature system	Discrete and categorical data	Consistency and completeness of knowledge
The requirement for compactness of classes	–	+	+	–	–	–	+	–	–
Processing of complex symbolic data	–	–	–	–	–	+	–	+	+
Applicability to large-scale feature space problems	+	+	+	+	+	–	–	–	–
Applicability poorly studied subject areas	–	+	+	+	+	–	–	+	–
Interoperability of DM	Low	Low	Low	Low	High	High	Middle	High	High
Automatic generalization of data	Low	High	High	High	Middle	Low	Low	Low	Low
The level of automation of DM building	High	High	High	Middle	Middle	Low	High	High	Low
Adaptiveness	Middle	High	High	High	Middle	Low	Middle	Low	Low
The need for a large amount of observations	+	–	–	–	–	–	–	–	–
The computational complexity of DM building	High	Middle	Middle	High	Middle	Low	High	High	Low
The use of local search in the DM construction	–	–	–	+	+	–	+	–	–

For the given training sample $\langle x, y \rangle$, where $y \in \{0, 1\}$, find coordinates of class centers (etalons as averaged instances representing corresponding classes):

$$C_j^q = \frac{1}{Sq} \sum_{s=1}^S \{x_j^s | y^s = q\}, j = 1, 2, \dots, N, q=0, 1,$$

as well as the sum by the coordinates:

$$C_j^+ = \frac{1}{S} \sum_{s=1}^S \{x_j^s\}, j = 1, 2, \dots, N,$$

and the average of the class centers:

$$C_j^{avg} = \frac{C_j^0 + C_j^1}{2}, j = 1, 2, \dots, N.$$

For each s -th instance of a sample, $s = 1, 2, \dots, S$, determine the distance from it to the sample center C^+ or to the center of each class C^q , $q=0, 1$, or to the average of class centers:

– as an Euclidean distance:

$$d^s(x^s, C) = \sqrt{\sum_{j=1}^N (x_j^s - C_j)^2},$$

– as a cosine of angle between vectors:

$$d^s(x^s, C) = \frac{\sum_{i=1}^N \sum_{j=1}^N x_i^s C_j}{\sqrt{\sum_{i=1}^N \sum_{j=1}^N x_i^s x_j^s} \sqrt{\sum_{i=1}^N \sum_{j=1}^N C_i C_j}}.$$

Here as C we conventionally note the center relative to which the distance is determined.

There are many other methods of determining distances that can be used.

Map the sample $\{\langle x_j^s, y^s \rangle\}$ to the two-dimensional coordinate system: $\{\langle d^s(C^0), d^s(C^1), y^s \rangle\}$, where y^s used as class label to marking instances.

Such a two-dimensional coordinate system will allow to visualize data sample and to estimate complexity of a problem.

If the separation of classes in a two-dimensional coordinate system be close to linear, then it seems possible to

use a modification of the metric classification method based on class centers.

The basic metric classification method assumes that at the training stage, class centers are defined in the system of original features as average coordinates of instances of the corresponding classes, and at the recognition stage, the recognized instance is assigned to the class whose distance to the corresponding class center from the given instance is the smallest.

Unlike the basic method, the initial set of features (signal acquisitions) is proposed to be replaced by distances in a two-dimensional coordinate system. This will significantly reduce the dimensionality of the data set, as well as make the model extremely simple.

The modified metric classification method can be formulated as such set of stages.

Phase of model training.

Stage 1. Initialization. Give a training data set $\langle x, y \rangle$ in original N -dimensional set of features. Specify the method of C and d computing.

Step 2. Model Building. Using given data set $\langle x, y \rangle$ compute class centers $\{C^q\}$ as a DM parameters.

Phase of recognition.

Stage 1. Initialization. Give data set $\langle x, y \rangle$ in original N -dimensional set of features and. Provide DM parameters (C coordinates) and d computing method.

Stage 2. Data Mapping. Using specified method of C computing map the sample $\{\langle \{x^s\}, y^s \rangle\}$ to the two-dimensional coordinate system: $\{\langle d^s(C^p), d^s(C^q), y^s \rangle\}$.

Step 3. Decision Making. For each recognized instance x^s the decision on it's classifying can be computed using the formula:

$$y^s = \arg \min_{p,q} (d^s(x^s, C^p), d^s(x^s, C^q))$$

or the same in other form:

$$y^s = \{p \mid d^s(x^s, C^p) \leq d^s(x^s, C^q)\}.$$

Such idea may be used also for a feed-forward neural network building. Simple shallow two-layer feed-forward neural network (Fig. 2) [36] may be constructed and trained in a non-iterative mode using following method.

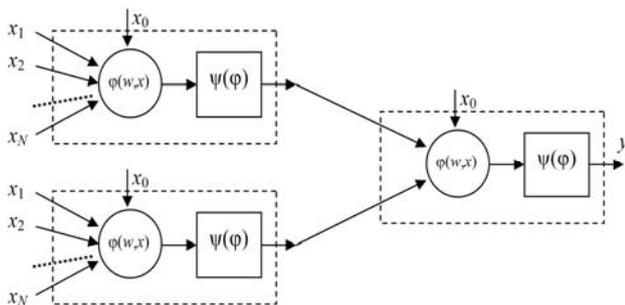


Figure 2 – Schema of the two-layer feed-forward neural network

The network input is a set of original N features $\{x_j\}$. Each i -th neuron of a η -th layer of a network perform computation: $y^{(\eta,i)} = \psi^{(\eta,i)}(\varphi^{(\eta,i)})$.

The first hidden layer of a network is represented by two neurons mapping instances from the original feature space to the two dimensional space. Each neuron has discriminative (postsynaptic) function Euclidean distance:

$$\varphi^{(1,i)} = \sum_{j=1}^N (x_j - w_j^{(1,i)})^2, \quad i = 1, 2,$$

and transfer (activation) function Gaussian:

$$\psi^{(1,i)}(a^2) = \exp(-a^2),$$

or sigmoid:

$$\psi^{(1,i)}(a) = \frac{1}{1 + \exp(-a)},$$

or pure linear:

$$\psi^{(1,i)}(a) = a.$$

First neuron compute distance from the recognized object to the C^0 and the second neuron computes distance to the C^1 . So weights of first neuron will be equal to the values of C^0 coordinates, and the second neuron weights will be equal to the values of C^1 coordinates:

$$w_j^{(1,i)} = C_j^{i-1}, \quad j = 1, 2, \dots, N; i = 1, 2.$$

The second layer of a network contains one neuron, which separates classes in a two-dimensional space of distance coordinates. This neuron use weighted sum as discriminative (postsynaptic) function:

$$\varphi^{(2,1)} = w_0^{(2,1)} + \sum_{j=1}^2 w_j^{(2,1)} \psi^{(1,j)},$$

and hard-lim (Heaviside step) function as a transfer function:

$$\psi^{(2,1)}(a) = \begin{cases} 0, & a < 0; \\ 1, & a \geq 0, \end{cases}$$

or as a sigmoid function:

$$\psi^{(2,1)}(a) = \frac{1}{1 + \exp(-a)}.$$

The weights of this neuron may be evaluated using the formula:

– for the case, when first layer neurons use Gaussians:

$$w_j^{(2,1)} = \begin{cases} 0, & j = 0; \\ 1, & j = 1; \\ -1, & j = 2; \end{cases}$$

– for the case, when first layer neurons use sigmoid or pure linear:

$$w_j^{(2,1)} = \begin{cases} 0, & j = 0; \\ -1, & j = 1; \\ 1, & j = 2. \end{cases}$$

After network creation it is possible to adjust its parameters to improve its accuracy, if needed, or for additional training, when additional set of training instances should be used. For this purpose, the steepest descent method [20] with a backpropagation technique [37] or evolutionary search methods may be used.

4 EXPERIMENTS

Let's consider a real problem of constructing a diagnostic model for diagnosing helicopter gears.

Detection of gear faults is challenging. Undetected gear faults can result in catastrophic gearbox failures, which depending on the criticality of the application, can be life threatening. For vibration monitoring, it is identifying a feature within the background noise of a gearbox which is indicative of fault. For critical system, the risk is a missed detection that allows a mission to proceed when it should be aborted.

The data set provided by GPMS Inc., USA contains 1158 raw vibration signals measured for helicopter gears.

The roughly two-thirds of these instances are nominal data while the other third have a gear fault. Each sample instance provided as an input signal represented by 93752 acquisitions (discrete counts) associated with a value of the output binary variable representing class of the instance: good (nominal) or bad (having a gear fault).

The examples of graphs for good and bad gear signals and their parameters are shown at Fig. 3. As can be seen from Fig. 3, the signal parameters [2] can be used to construct a decision-making model for the signal belonging to one of two classes. However, determining these parameters requires significant computational resources and requires solving the problem of selecting informative features among the signal parameters.

Therefore, we will use the method proposed in this paper, which maps instances into a two-dimensional space with subsequent construction of a DM for a two-dimensional data sample.

As a main indicator for method or model performance evaluation we should use a model error probability estimation (percent of incorrect decisions) or model correct decision probability estimation (percent of correct decisions).

5 RESULTS

The fragment of the results of conducted experiments is presented at Fig. 4–Fig. 9. Here the C^+ designated as C , and C^{avg} designated as C^m .

As it can be seen from Fig. 4–Fig. 9, the proposed method allows map multidimensional data into two-dimensional space, ensuring linear separability of classes in a two-dimensional space. On the one hand, it allows visualizing data and getting view on the number and shape of clusters, as well as the significance of instances, and, on the other hand, clearly confirms the possibility of using the proposed method for constructing a DM in the problem of diagnosing helicopter gears.

For the experimental study of the applicability of the proposed method for DM building the training and test samples were formed from the initial sample by randomly selecting instances of each class, maintaining the ratio of the shares of classes in the initial sample. Then for the obtained training samples, DM were constructed based on the proposed method, which were tested on the corresponding test samples. The results of experiments are given in Table 2.

Table 2 – Results of experiments on constructing DM based on the proposed method

S_{train}	S_{test}	$P_{incorrect\ train}$	$P_{incorrect\ test}$
10	1148	0	0/0043
50	1108	0	0.0009
100	1058	0	0
250	908	0	0
500	658	0	0

As can be seen from Table 2, the proposed method allows obtaining highly accurate DM even with small training samples, provided that the frequency of classes in the samples is preserved.

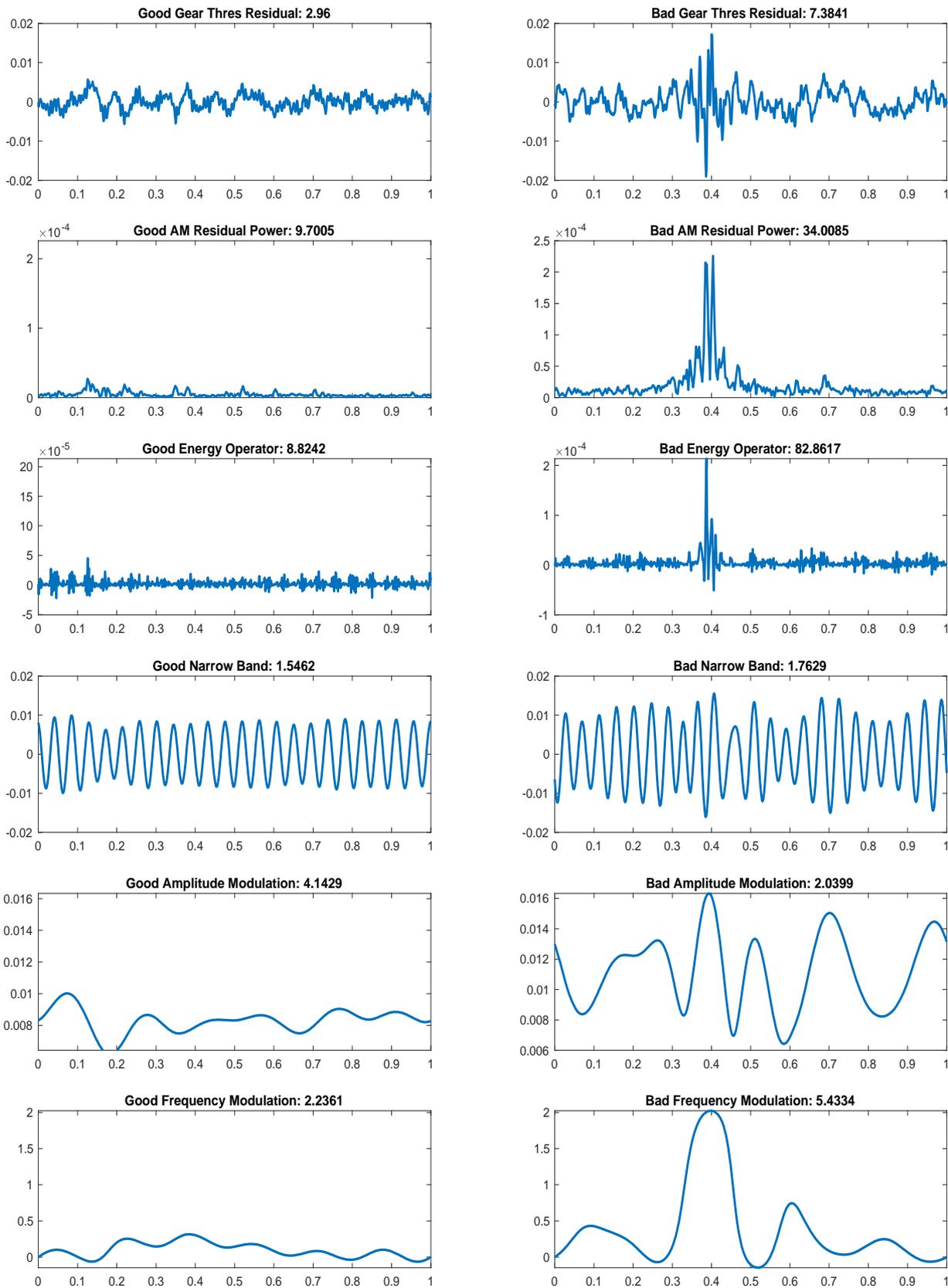


Figure 3 – Examples of graphs for good and bad gear signals and their parameters

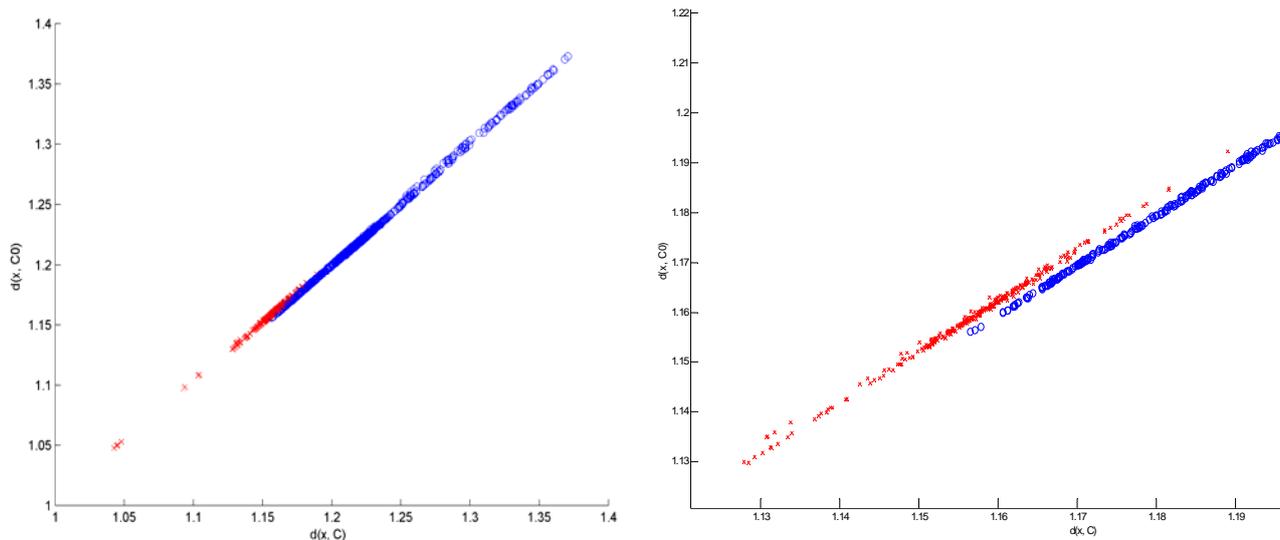


Figure 4 – Dataset in the $d(x, C^+)$ and $d(x, C^0)$ coordinate system

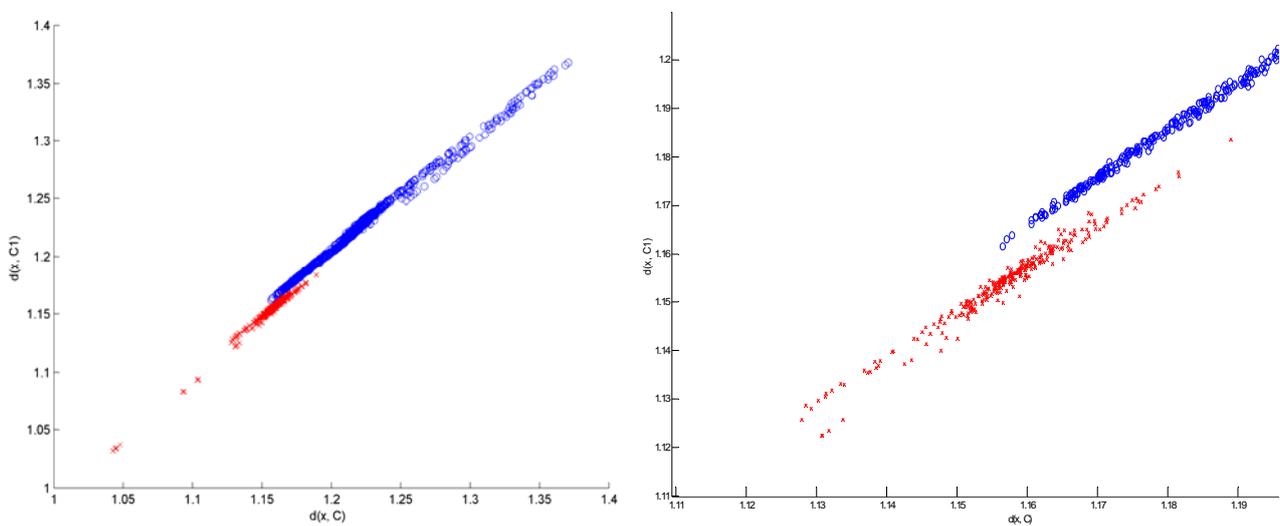


Figure 5 – Dataset in the $d(x, C^+)$ and $d(x, C^1)$ coordinate system

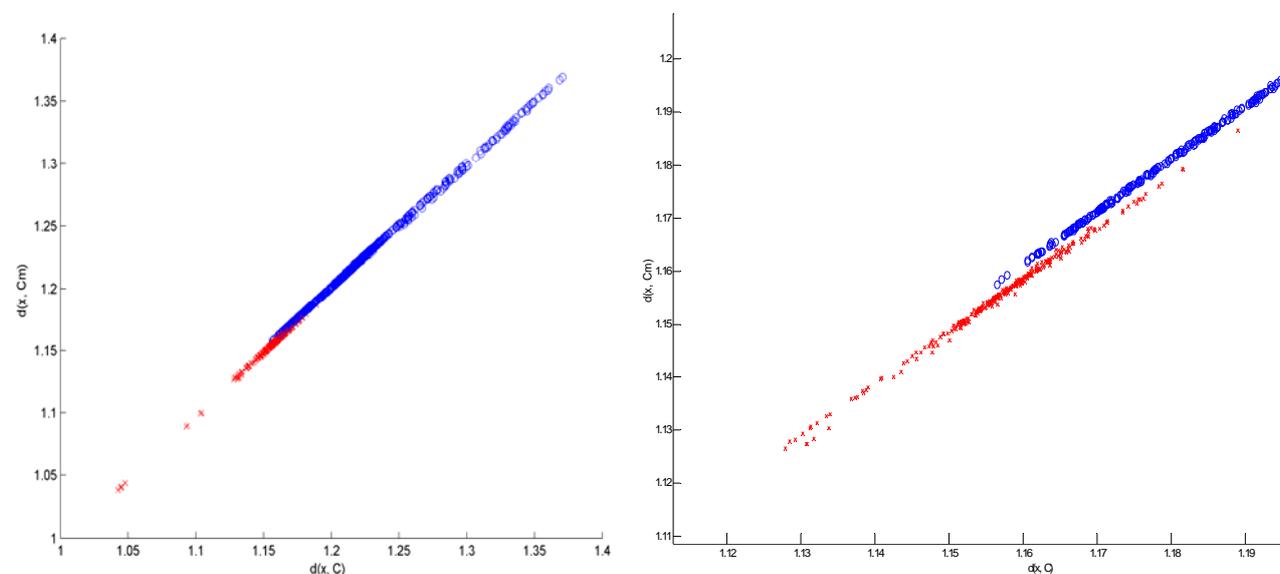


Figure 6 – Dataset in the $d(x, C^+)$ and $d(x, C^{avg})$ coordinate system

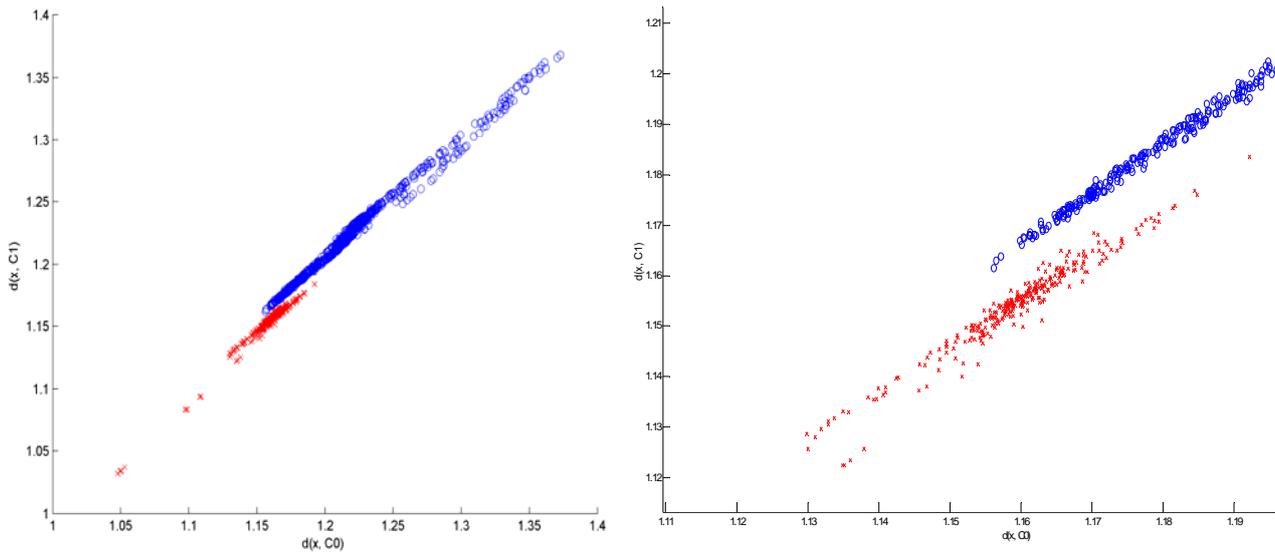


Figure 7 – Dataset in the $d(x, C^0)$ and $d(x, C^1)$ coordinate system

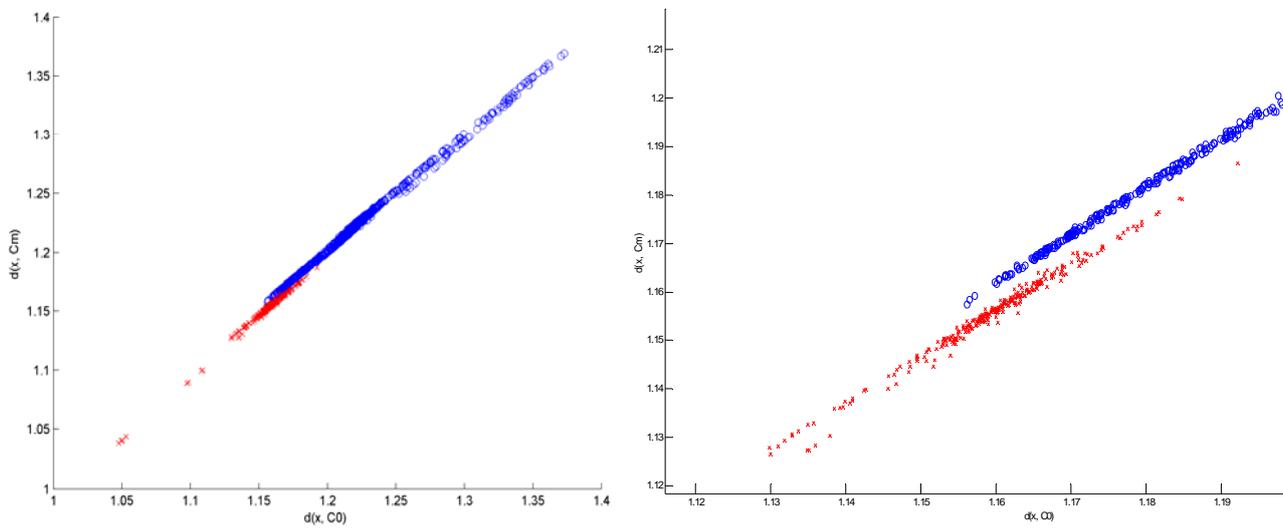


Figure 8 – Dataset in the $d(x, C^0)$ and $d(x, C^{avg})$ coordinate system

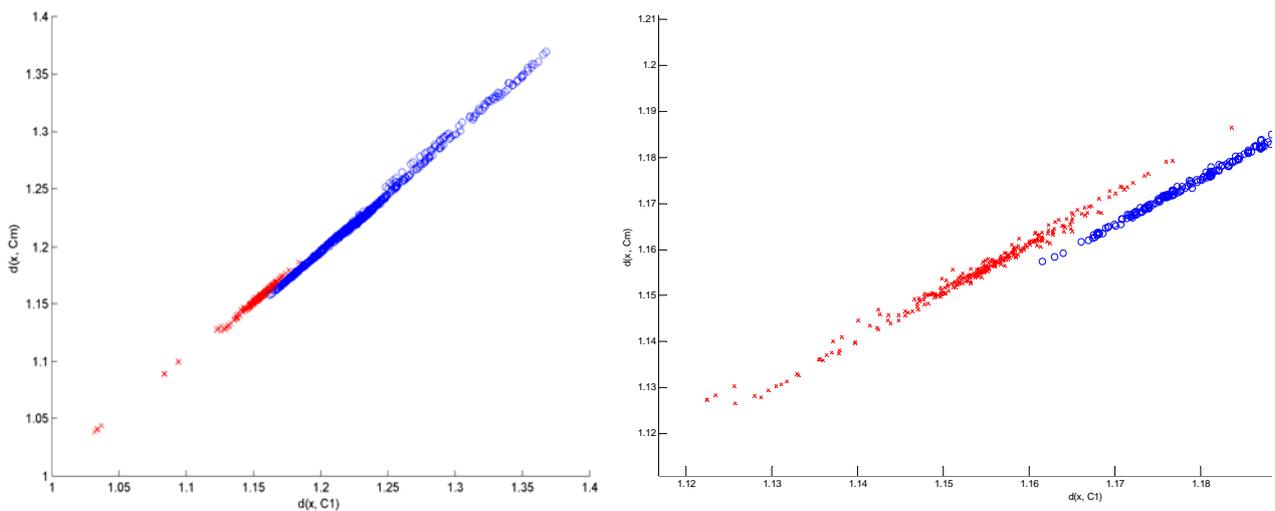


Figure 9 – Dataset in the $d(x, C^1)$ and $d(x, C^{avg})$ coordinate system

6 DISCUSSION

The results of conducted experiments allow to conclude that the proposed method provides a significant reduction in the data dimensionality (in particular, for the considered problem of constructing a DM for helicopter gear diagnosis, it reduces the data dimensionality due to the compression of features by 46876 times).

Additionally, it may be of interest to study a possible combination of the proposed method with methods for sample forming using metrics of the value of instances. As the results of the conducted experiments for randomly selected instances in a two-dimensional system of artificial features obtained on the basis of the proposed method showed a significant reduction of the sample for individual tasks may allow to provide acceptable accuracy. And taking into account individual estimates of the instance significance will allow, even for small samples, to ensure the topological representativeness of the formed sample in relation to the original sample.

CONCLUSIONS

The urgent problem of data-driven diagnostic model constructing for decision-making automation in health and usage monitoring process is considered in the paper.

The scientific novelty of obtained results is that a method is proposed for the mapping of multidimensional data into a two-dimensional space preserving local properties of class separation, allowing for the visualization of multidimensional data and the production of simple diagnostic models for the automatic classification of diagnostic objects. A method for synthesizing diagnostic models based on a two-layer feed-forward neural network is also proposed, which allows obtaining models in a non-iterative mode.

The practical significance of obtained results is that a sample of observations of the state of helicopter gears was obtained, which can be used to compare data-driven diagnostic methods and data processing methods that solve the problems of data dimensionality reduction. Mathematical support has been developed that allows displaying a sample from a multidimensional to a two-dimensional space, which makes it possible to visualize data and reduces the dimensionality of the data. Diagnostic models have been obtained that allow automating the decision-making process on whether the diagnosed object (helicopter gear) belongs to one of two classes of states.

The prospects for further research are to compare methods for constructing data-driven models, as well as methods for reducing the dimensionality of data based on the proposed sample. Additionally, it may be of interest to study a possible combination of the proposed method with methods for sample forming using metrics of the value of instances.

ACKNOWLEDGEMENTS

The Dataset used in experiments were provided by the GPMS Inc, USA. The work is supported by the National University “Zaporizhzhia Polytechnic” and supported by the international projects of WORK4CE: Cross-domain

© Subbotin S. A., Bechhoefer E., 2025
DOI 10.15588/1607-3274-2025-1-11

competences for healthy and safe work in the 21st century (Project Reference: 619034-EPP-1-2020-1-UAEPPKA2-CBHE-JP) of the ERASMUS+ programme of the European Union and ViMaCs: Virtual Master Cooperation Data Science Project-ID: 57513461 (Phase 1), 57602060 (Phase 2) of the DAAD.

REFERENCES

1. Land J. E. HUMS – The benefits – Past, present and future, *2001 IEEE Aerospace Conference, 10–17 March 2001, Big Sky : proceedings*. Los Alamitos, IEEE, 2001, Vol. 6, pp. 3083–3094. DOI: 10.1109/AERO.2001.931326
2. Bechhoefer E., Butterworth B. A comprehensive analysis of the performance of gear fault detection algorithms [Electronic resource], *Proceedings of the Annual Conference of the PHM Society*, 2019, Vol. 11, № 1:). Mode of Access: <https://papers.phmsociety.org/index.php/phmconf/article/view/823>
3. Giurgiutiu V., Cuc A., Goodman P. Review of vibration-based helicopters health and usage monitoring methods [Electronic resource], *55th Meeting of the Society for Machinery Failure Prevention Technology, Virginia Beach, 2–5 April 2001 : proceedings*, pp. 1–10. Mode of Access: <https://apps.dtic.mil/sti/tr/pdf/ADP013516.pdf>
4. Draper A., Gourlay J. The operational benefits of health and usage monitoring systems in UK military helicopters [Electronic resource], *HUMS 2003 Conference*, pp. 71–79. Mode of Access: <https://humsconference.com.au/Papers2003/HUMSp408.pdf>
5. Simani S., Fantuzzi C., Patton R.J. Model-based fault diagnosis in dynamic systems using identification techniques. London, Springer, 2013, 282 p. DOI: 10.1007/978-1-4471-3829-7
6. Wen G., Chen X., Lei Z., Huang X. New generation artificial intelligence-driven diagnosis and maintenance techniques. Advanced machine learning models, methods and applications. Singapore, Springer Nature, 2024, 349 p.
7. Engelbrecht A. Computational intelligence: an introduction. Sidney, John Wiley & Sons, 2007, 597 p. DOI: 10.1002/9780470512517
8. Kruse S., Plattner H. Efficient Discovery of Approximate Dependencies / S. Kruse, // *Proceedings of the VLDB Endowment*. – 2018. – Vol. 11. – No. 7. – P. 759–772. DOI: 10.14778/3192965.3192968
9. Subbotin S. A., Oliinyk A. A. The dimensionality reduction methods based on computational intelligence in problems of object classification and diagnosis, *Advances in Intelligent Systems and Computing*, 2017, Vol. 543, pp. 11–19. DOI: https://doi.org/10.1007/978-3-319-48923-0_2
10. Subbotin S.A. Experimental investigation and analysis of information quality indices correlation for diagnostic neuromodels, *Radio Electronics, Computer Science, Control*, 2011, № 1, pp. 104–110. DOI: 10.15588/1607-3274-2011-1-19
11. Subbotin S. The quality indicators of decision tree and forest based models, *Computer Modeling and Intelligent*



- Systems (CMIS-2020) : The Third International Workshop on Zaporizhzhia, Ukraine, April 27-May 1, 2020 / ed.: S. Subbotin : proceedings, Aachen, CEUR-WS, 2020, pp. 718–743. (CEUR-WS.org, vol. 2608). Access mode: <http://ceur-ws.org/Vol-2608/paper55.pdf>*
12. Webb A. R. Statistical pattern recognition. Chichester, Wiley, 2003, 514 p.
 13. Berezsky O., Zarichnyi M. Metric methods in computer vision and pattern recognition, *Advances in Intelligent Systems and Computing*, 2020, Vol. 1293. Cham, Springer, 2020, pp. 188–209. DOI: 10.1007/978-3-030-63270-0_13
 14. Subbotin S. The neuro-fuzzy network synthesis and simplification on precedents in problems of diagnosis and pattern recognition, *Optical Memory and Neural Networks (Information Optics)*, 2013, Vol. 22, № 2, pp. 97–103. DOI: 10.3103/s1060992x13020082
 15. Kazlas P. T., Monsen P. T., LeBlanc M. J. Neural network-based helicopter gearbox health monitoring system, *IEEE-SP Workshop on Neural Networks for Signal Processing, Linthicum Heights, 6–9 September 1993 : proceedings*. Los Alamitos: IEEE, 1993, pp. 431–440. DOI: 10.1109/NNSP.1993.471845
 16. d'Amato C., Bryl V., Serafini. L. Data-driven logical reasoning [Electronic resource]. Mode of Access: <https://ceur-ws.org/Vol-900/paper5.pdf>
 17. Lamperti G., Zanella M. Rule-based diagnosis, *Diagnosis of Active Systems*, Vol 741, pp. 193–233. DOI: https://doi.org/10.1007/978-94-017-0257-7_7
 18. Pétrowski A., Ben-Hamida S. Evolutionary algorithms. Chichester, Wiley, 2017, 256 p.
 19. Eds: A. Kumar, P. Rathore, R. Agrawal, Diaz V. Swarm Intelligence Optimization. Algorithms and Applications. Chichester, Wiley, 2021, 384 p.
 20. Rao S. Engineering optimization. Theory and practice. Chichester, Wiley, 2019, 832 p.
 21. Lee J. A., Verleysen M. Nonlinear dimensionality reduction. New York, Springer, 2007, 309 p. DOI: 10.1007/978-0-387-39351-3
 22. Jensen R., Shen Q. Computational intelligence and feature selection: rough and fuzzy approaches. Hoboken: John Wiley & Sons, 2008, 339 p. DOI: 10.1002/9780470377888
 23. Guyon I., Elisseeff A. An introduction to variable and feature selection, *Journal of machine learning research*, 2003. Vol. 3, pp. 1157–1182.
 24. Eds: Motoda H., Liu H. Feature extraction, construction and selection. A data mining perspective. New York, Springer, 2012, 410 p.
 25. Zheng A., Casari A. Feature Engineering for Machine Learning: Principles and Techniques for Data Scientists. Sebastopol, CA: O'Reilly, 2018, 215 p.
 26. Indyk P., Motwani R., Raghavan P., Vempala S. Locality-preserving hashing in multidimensional spaces, *Twenty-ninth annual ACM symposium on theory of computing (STOC '97) : proceedings*. New York, ACM, 1997, pp. 618–625. DOI:10.1145/258533.258656.
 27. Subbotin, S. A. Evaluation of informativity and selection of instances based on hashing, *Radio Electronics, Computer Science, Control*, 2020, № 3, pp. 129–137. DOI: 10.15588/1607-3274-2020-3-12
 28. Subbotin S. A. Methods of sampling based on exhaustive and evolutionary search, *Automatic Control and Computer Sciences*, 2013, Vol. 47, № 3, pp. 113–121. DOI: 10.3103/s0146411613030073
 29. Jankowski N., Grochowski M. Comparison of instance selection algorithms I. Algorithms survey, *Artificial Intelligence and Soft Computing : 7th International Conference ICAISC-2004, Zakopane, 7–11 June, 2004 : proceedings*. – Berlin : Springer, 2004, pp. 598–603. (Lecture Notes in Computer Science, Vol. 3070). DOI: 10.1007/978-3-540-24844-6_90
 30. Tan L., Jiang J. Digital signal processing. fundamentals and applications. London, Elsevier/Academic Press, 2018, 920 p.
 31. Lymariiev I., Subbotin S., Oliinyk A., Drokin I. Diagnostic signal nonstationarity reduction to predict the helicopter transmission state on the basis of intelligent information technologies, *The Second International Workshop on Computer Modeling and Intelligent Systems (CMIS-2019), Zaporizhzhia, 15–19 of April 2019 : proceedings*. Mode of Access: <https://ceur-ws.org/Vol-2353/paper40.pdf>
 32. Lymariiev I. O., Subbotin S. A., Oliinyk A. A., Drokin I. V. Methods of large-scale signals transformation for diagnosis in neural network models, *Radio Electronics, Computer Science, Control*, 2018, Vol. 4, pp. 63–71. DOI: 10.15588/1607-3274-2018-4-11
 33. Gao R. X., Yan R.. Wavelets. Theory and applications for manufacturing. New York, Springer, 2010. – 224 p. DOI: 10.1007/978-1-4419-1545-0
 34. Scitovski R., Sabo K., Martínez-Álvarez F., Ungar Š. Cluster analysis and applications. Cham, Springer, 2021, 271 p. DOI: 10.1007/978-3-030-74552-3
 35. Hurter C. Image-based visualization. Interactive multi-dimensional data exploration. Cham, Springer, 2022, 111 p.
 36. Haykin S. Neural networks and learning machines. Hoboken, Prentice Hall, 2009, 906 p.
 37. Rumelhart D., Hinton G., Williams R. Learning representations by back-propagating errors, *Nature*, 1986, Vol. 323, pp. 533–536. DOI:10.1038/323533a0

Received 09.12.2024.
Accepted 06.02.2025.

ПОБУДОВА ДІАГНОСТИЧНОЇ МОДЕЛІ, КЕРОВАНОЇ ДАНИМИ, ДЛЯ МОНІТОРИНГУ СПРАВНОСТІ ТА ВИКОРИСТАННЯ СПОРЯДЖЕННЯ ГЕЛІКОПТЕРІВ

Субботін С. О. – д-р техн. наук, професор, завідувач кафедри програмних засобів Національного університету «Запорізька політехніка», Запоріжжя, Україна.

Бекгофер Е. – доктор філософії, генеральний директор та засновник GPMS Inc., Сполучені Штати Америки.

АНОТАЦІЯ

Актуальність. Сучасні технічні об'єкти (зокрема транспортні засоби) є надзвичайно складними та висувають великі вимоги до надійності. Це потребує автоматизації моніторингу стану та діагностування несправностей об'єктів та їх складових. Прогнозне обслуговування підвищує експлуатаційну готовність технічних об'єктів. Об'єктом дослідження є процес моніторингу справності та використання технічних об'єктів. Предметом дослідження є методи обчислювального інтелекту для побудови керованої даними моделі та відповідні завдання опрацювання даних для системи моніторингу працездатності та використання.

Мета. Мета роботи – сформулювати задачі обробки даних, сформувати набір даних для побудови керованої даними моделі та побудувати простий метод автоматичної побудови діагностичних моделей на прикладі системи моніторингу стану та використання гелікоптерів.

Метод. Запропоновано метод для відображення багатовимірних даних у двовимірний простір із збереженням локальних властивостей поділу класів, що дозволяє візуалізувати багатовимірні дані та створювати прості діагностичні моделі для автоматичної класифікації об'єктів діагностування. Запропонований метод дозволяє отримати високоточну діагностичну модель з малими навчальними вибірками за умови збереження частоти класів у вибірках. Запропоновано також метод синтезу діагностичних моделей на основі двошарової нейронної мережі прямого поширення, що дозволяє отримувати моделі в неутративному режимі.

Результати. Отримано вибірку спостережень стану механізмів вертольота, яку можна використовувати для порівняння методів діагностування, керованої даними, та методів опрацювання даних, які вирішують задачі скорочення розмірності даних. Розроблено програмне забезпечення, яке дозволяє відображати вибірку з багатовимірного простору в двовимірний, що дає змогу візуалізувати дані та зменшує розмірність даних. Отримано діагностичні моделі, які дозволяють автоматизувати процес прийняття рішення про належність діагностованого об'єкта (спорядження вертольота) до одного з двох класів станів.

Висновки. Результати проведених експериментів дозволяють зробити висновок, що запропонований метод забезпечує суттєве зменшення розмірності даних (зокрема, для розглянутої задачі побудови моделі діагностування вертолітного обладнання зменшує розмірність даних за рахунок стиснення ознак у 46876 разів). Оскільки результати проведених експериментів для випадково вибраних екземплярів у двовимірній системі штучних ознак, отриманих на основі запропонованого методу, показали значне скорочення вибірки для окремих завдань, це може дозволити забезпечити прийнятну точність. А врахування індивідуальних оцінок значущості екземплярів дозволить навіть для малих вибірок забезпечити топологічну репрезентативність сформованої вибірки по відношенню до вихідної вибірки.

Перспективи подальших досліджень полягають у порівнянні методів побудови моделей, керованих даними, а також методів зменшення розмірності даних на основі запропонованої вибірки. Крім того, може становити інтерес дослідження можливого поєднання запропонованого методу з методами формування вибірки з використанням метрик значення екземплярів.

КЛЮЧОВІ СЛОВА: діагностування на основі даних, система моніторингу працездатності та використання, зменшення розмірності даних, класифікація.

ЛІТЕРАТУРА

1. Land J. E. HUMS – The benefits – Past, present and future / J. E. Land // 2001 IEEE Aerospace Conference, 10–17 March 2001, Big Sky : proceedings. Los Alamitos, IEEE, 2001. – Vol. 6. – P. 3083–3094. DOI: 10.1109/AERO.2001.931326
2. Bechhoefer E. A comprehensive analysis of the performance of gear fault detection algorithms [Electronic resource] / E. Bechhoefer, B. Butterworth // Proceedings of the Annual Conference of the PHM Society. – 2019. – Vol. 11, № 1). – Mode of Access: <https://papers.phmsociety.org/index.php/phmconf/article/view/823>
3. Giurgiutiu V. Review of vibration-based helicopters health and usage monitoring methods [Electronic resource] / V. Giurgiutiu, A. Cuc, P. Goodman // 55th Meeting of the Society for Machinery Failure Prevention Technology, Virginia Beach, 2–5 April 2001 : proceedings. – P. 1–10. – Mode of Access: <https://apps.dtic.mil/sti/tr/pdf/ADP013516.pdf>
4. Draper A. The operational benefits of health and usage monitoring systems in UK military helicopters [Electronic resource] / A. Draper, J. Gourlay // HUMS 2003 Conference. – P. 71–79. – Mode of Access: <https://humsconference.com.au/Papers2003/HUMSp408.pdf>
5. Simani S. Model-based fault diagnosis in dynamic systems using identification techniques / S. Simani, C. Fantuzzi, R.J. Patton. – London : Springer, 2013. – 282 p. DOI: 10.1007/978-1-4471-3829-7
6. New generation artificial intelligence-driven diagnosis and maintenance techniques. Advanced machine learning models, methods and applications / [G. Wen, X. Chen, Z. Lei, X. Huang]. – Singapore : Springer Nature, 2024. – 349 p.
7. Engelbrecht A. Computational intelligence: an introduction / A. Engelbrecht. – Sidney : John Wiley & Sons, 2007. – 597 p. DOI: 10.1002/9780470512517
8. Kruse S. Efficient Discovery of Approximate Dependencies / S. Kruse, H. Plattner // Proceedings of the VLDB Endowment. – 2018. – Vol. 11, No. 7. – P. 759–772. DOI: 10.14778/3192965.3192968

9. Subbotin S. A. The dimensionality reduction methods based on computational intelligence in problems of object classification and diagnosis / S. A. Subbotin, A. A. Oliinyk // *Advances in Intelligent Systems and Computing*. – 2017. – Vol. 543. – P. 11–19. DOI: https://doi.org/10.1007/978-3-319-48923-0_2
10. Subbotin S. A. Experimental investigation and analysis of information quality indices correlation for diagnostic neuromodels // *Radio Electronics, Computer Science, Control*. – 2011. – № 1. – P. 104–110. DOI: 10.15588/1607-3274-2011-1-19
11. Subbotin S. The quality indicators of decision tree and forest based models / S. Subbotin // *Computer Modeling and Intelligent Systems (CMIS-2020)* : The Third International Workshop on Zaporizhzhia, Ukraine, April 27-May 1, 2020 / ed.: S. Subbotin : proceedings. – Aachen : CEUR-WS, 2020. – P. 718–743. – (CEUR-WS.org, vol. 2608). – Access mode: <http://ceur-ws.org/Vol-2608/paper55.pdf>
12. Webb A. R. Statistical pattern recognition / A. R. Webb. – Chichester : Wiley, 2003. – 514 p.
13. Berezsky O. Metric methods in computer vision and pattern recognition / O. Berezsky, M. Zarichnyi // *Advances in Intelligent Systems and Computing*. – 2020. – Vol. 1293. – Cham : Springer, 2020. – P. 188–209. DOI: 10.1007/978-3-030-63270-0_13
14. Subbotin S. The neuro-fuzzy network synthesis and simplification on precedents in problems of diagnosis and pattern recognition / S. Subbotin // *Optical Memory and Neural Networks (Information Optics)*. – 2013. – Vol. 22, № 2. – P. 97–103. DOI: 10.3103/s1060992x13020082
15. Kazlas P. T. Neural network-based helicopter gearbox health monitoring system / P. T. Kazlas, P. T. Monsen, M. J. LeBlanc // *IEEE-SP Workshop on Neural Networks for Signal Processing, Lanthicum Heights, 6–9 September 1993* : proceedings. – Los Alamitos : IEEE, 1993. – P. 431–440. DOI: 10.1109/NNSP.1993.471845
16. d'Amato C. Data-driven logical reasoning [Electronic resource] / C. d'Amato, V. Bryl, L. Serafini. – Mode of Access: <https://ceur-ws.org/Vol-900/paper5.pdf>
17. Lamperti G. Rule-based diagnosis / G. Lamperti, M. Zanella // *Diagnosis of Active Systems*. – Vol 741. – P. 193–233. DOI: https://doi.org/10.1007/978-94-017-0257-7_7
18. Pétrowski A. Evolutionary algorithms / A. Pétrowski, S. Ben-Hamida. – Chichester : Wiley, 2017. – 256 p.
19. *Swarm Intelligence Optimization. Algorithms and Applications* / [eds: A. Kumar, P. Rathore, R. Agrawal, V. Diaz]. – Chichester: Wiley, 2021. – 384 p.
20. Rao S. *Engineering optimization. Theory and practice* / S. Rao. – Chichester : Wiley, 2019. – 832 p.
21. Lee J. A. *Nonlinear dimensionality reduction* / J. A. Lee, M. Verleysen. – New York : Springer, 2007. – 309 p. DOI: 10.1007/978-0-387-39351-3
22. Jensen R. *Computational intelligence and feature selection: rough and fuzzy approaches* / R. Jensen, Q. Shen. – Hoboken: John Wiley & Sons, 2008. – 339 p. DOI: 10.1002/9780470377888
23. Guyon I. An introduction to variable and feature selection / I. Guyon, A. Elisseeff // *Journal of machine learning research*. – 2003. – Vol. 3. – P. 1157–1182.
24. *Feature extraction, construction and selection. A data mining perspective* / eds: H. Motoda, H. Liu. – New York : Springer, 2012. – 410 p.
25. Zheng A. *Feature Engineering for Machine Learning: Principles and Techniques for Data Scientists* / A. Zheng, A. Casari. – Sebastopol, CA: O'Reilly, 2018. – 215 p.
26. *Locality-preserving hashing in multidimensional spaces* / [P. Indyk, R. Motwani, P. Raghavan, S. Vempala] // *Twenty-ninth annual ACM symposium on theory of computing (STOC '97)* : proceedings. New York : ACM, 1997. – P. 618–625. DOI:10.1145/258533.258656.
27. Subbotin S. A. Evaluation of informativity and selection of instances based on hashing / S. A. Subbotin // *Radio Electronics, Computer Science, Control*. – 2020. – № 3. – P. 129–137. DOI: 10.15588/1607-3274-2020-3-12
28. Subbotin S. A. Methods of sampling based on exhaustive and evolutionary search / S. A. Subbotin // *Automatic Control and Computer Sciences*. – 2013. – Vol. 47, № 3. – P. 113–121. DOI: 10.3103/s0146411613030073
29. Jankowski N. Comparison of instance selection algorithms I. Algorithms survey / N. Jankowski, M. Grochowski // *Artificial Intelligence and Soft Computing* : 7th International Conference ICAISC-2004, Zakopane, 7–11 June, 2004 : proceedings. – Berlin : Springer, 2004. – P. 598–603. – (Lecture Notes in Computer Science, Vol. 3070). DOI: 10.1007/978-3-540-24844-6_90
30. Tan L. *Digital signal processing. fundamentals and applications* / L. Tan, J. Jiang. – London : Elsevier/Academic Press, 2018. – 920 p.
31. Diagnostic signal nonstationarity reduction to predict the helicopter transmission state on the basis of intelligent information technologies / [I. Lymariiev, S. Subbotin, A. Oliinyk, I. Drokin] // *The Second International Workshop on Computer Modeling and Intelligent Systems (CMIS-2019), Zaporizhzhia, 15–19 of April 2019* : proceedings. – Mode of Access: <https://ceur-ws.org/Vol-2353/paper40.pdf>
32. Methods of large-scale signals transformation for diagnosis in neural network models / [I. O. Lymariiev, S. A. Subbotin, A. A. Oliinyk, I. V. Drokin] // *Radio Electronics, Computer Science, Control*. – 2018. – Vol. 4. – P. 63–71. DOI: 10.15588/1607-3274-2018-4-11
33. Gao R. X. *Wavelets. Theory and applications for manufacturing* / R. X. Gao, R. Yan. – New York: Springer, 2010. – 224 p. DOI: 10.1007/978-1-4419-1545-0
34. *Cluster analysis and applications* / [R. Scitovski, K. Sabo, F. Martínez-Álvarez, S. Ungar]. – Cham : Springer, 2021. – 271 p. DOI: 10.1007/978-3-030-74552-3
35. Hurter C. *Image-based visualization. Interactive multidimensional data exploration* / C. Hurter. – Cham : Springer, 2022. – 111 p.
36. Haykin S. *Neural networks and learning machines* / S. Haykin. – Hoboken : Prentice Hall, 2009. – 906 p.
37. Rumelhart, D. Learning representations by back-propagating errors / D. Rumelhart, G. Hinton, R. Williams // *Nature*. – 1986. – Vol. 323. – P. 533–536. DOI:10.1038/323533a0

METHOD OF FORMING MULTIFACTOR PORTRAITS OF THE SUBJECTS SUPPORTING SOFTWARE COMPLEXES, USING A MULTILAYER PERCEPTRON

Pukach A. I. – PhD, Assistant at Department of Automated Control Systems, Institute of Computer Science and Information Technologies, Lviv Polytechnic National University, Lviv, Ukraine.

Teslyuk V. M. – Dr. Sc., Professor, Head of the Department of Automated Control Systems, Institute of Computer Science and Information Technologies, Lviv Polytechnic National University, Lviv, Ukraine.

ABSTRACT

Context. The problem of identification and determination of personalized comprehensive indicators of presence each of the impact factors in the processes of personal subjectivization of the researched supported object's perception by the relevant subjects interacting with it and making influence on its support, is being considered in this research. The process of forming multifactor portraits of subjects supporting software complexes, using a multilayer perceptron, is an object of study. While methods and means of forming such multifactor portraits of subjects supporting software complexes is the subject of study respectively.

Objective. The goal of the work is the creation of a method of forming multifactor portraits of subjects supporting software complexes, using a multilayer perceptron.

Method. A method of forming multifactor portraits of subjects supporting software complexes is proposed, using artificial neural networks of the multilayer perceptron type, which provides possibility to form appropriate personalized multifactor portraits of subjects which, directly or indirectly, interact with the object of support (which can represent both the supported software complex itself as well as the processes associated with its complex support activities).

Results. The results of functioning of the developed method are the corresponding models of multifactor portraits of subjects supporting software complexes, which later are used to solve a cluster of scientific and applied problems of software complexes' support automation, in particular, the problem of identification and determination of personalized comprehensive indicators of presence each of the impact factors (from appropriate pre-agreed and declared set of impact factors) in the processes of personal subjectivization of the researched supported object's perception by the relevant subjects interacting (directly, or indirectly) with it and making influence on its support. As an example, of practical application and approbation of the developed method, the results of resolving the applied practical task of automated search and selection of a maximal relevant candidate (from among the members of the support team of the supported software complex) for best solving of a stack of specialized client's requests (related to the support of this software complex), are given.

Conclusions. The developed method provides possibility to resolve the scientific and applied problem of identification and determination of personalized comprehensive indicators of presence each of the impact factors (from appropriate pre-agreed and declared set of impact factors) in the processes of personal subjectivization of the researched supported object's perception by the relevant subjects interacting (directly, or indirectly) with it and making influence on its support. In addition, the developed method provides possibility for creating appropriate models of multifactor portraits of subjects supporting software complexes, which makes it possible to use them in solving problems, tasks, or issues related to the automation of search and selection of subjects supporting software complexes, which (subjects) meet the given criteria both in the context of subjectivization processes of personal perception of the support objects (e.g. supported software complexes themselves, or processes directly related to their support), as well as in the context of compatibility in interaction with client's users of these supported software products (as those users, in fact, are also subjects of interaction with the same researched supported object).

KEYWORDS: automation, software complex, support, impact factor, multifactor portrait, neural networks, multilayer perceptron.

ABBREVIATIONS

DevOps – Development and Operations;
TC – test case;
MP – multilayer perceptron;
AI – artificial intelligence;
ANN – artificial neural networks.

NOMENCLATURE

$Fmp[1..p][1..q]$ is a matrix function of converting the input data (incoming to the neurons of the input layer of the encapsulated pre-trained artificial neural network of the multilayer perceptron type) into the corresponding results at the output (i.e., to the neurons of the output layer) of the same ANN MP;

$H[1..q]$ is a hidden layers of neurons of an encapsulated pre-trained artificial neural network of the multilayer perceptron type;

$H[1][y(1)]$ is a specific neuron $y(1)$ of the first hidden layer $H[1]$ of an encapsulated pre-trained artificial neural network of the multilayer perceptron type, activated during simulation of the current test case;

$H[q][y(q)]$ is a specific neuron $y(q)$ of the last hidden layer $H[q]$ of an encapsulated pre-trained artificial neural network of the multilayer perceptron type, activated during simulation of the current test case;

$I[1..n]$ is a neurons of the input layer of an encapsulated pre-trained artificial neural network of the multilayer perceptron type;

$I[x]$ is a specific x -th neuron of the input layer of an encapsulated pre-trained artificial neural network of the multilayer perceptron type, activated during simulation of the current test case;

$Imp[1..n]$ is an input layer of neurons of an encapsulated pre-trained artificial neural network of the multilayer perceptron type;

$O[1..m]$ is a neurons of output layer of an encapsulated pre-trained artificial neural network of the multilayer perceptron type;

$O[z]$ is a specific z -th neuron of output layer of an encapsulated pre-trained artificial neural network of the multilayer perceptron type, activated during simulation of the current test case;

$Omp[1..m]$ is a output layer of neurons of an encapsulated pre-trained artificial neural network of the multilayer perceptron type;

$SubjMFP[s]Inst[c]$ is an instant slice of the multifactor portrait of the current researched subject $Subj[s]$, obtained based on the simulation results of the current test case $TC[c]$;

$SubjMFP[s]$ is an in-process-of-formation multifactor portrait of the current researched subject $Subj[s]$, obtained based on the results of fully completed simulation of all test cases available for this subject;

$SubjPort[x]$ is a generalized representation of a multifactor portrait of some/any researched subject $Subj[x]$;

$TC[1..s][i(1),...,i(n)]$ is a matrix of test cases (from the set $[i(1),...,i(n)]$) for modeling/simulation processes (as well as obtaining relevant results) regarding the subjectivization of perception of the supported object by the relevant current researched subject (from the set $TC[1..s]$) interacting with this object.

INTRODUCTION

The current state of the software product support industry requires researches for new solutions and opportunities to provide, first of all, comprehensive and automated support. The need for comprehensive support of software products is dictated by the conditions of the entire information technologies field modern market – in particular: high requirements for the competitiveness of each software product, because clients, users, and investors are interested in the most complete and comprehensive support for all those software complexes which they use, or invest in. Meanwhile, support automation is driven by the extremely high intensity and dynamics of the information technology market, where speed and time are one of the greatest values and, again, a competitive advantage.

Thus, the topic (and the problematic, at the same time) of automation in the context of comprehensive activities related to software products' support continues to be an extremely relevant and non-trivial basic and formative component of the corresponding relevant scientific and applied problem, which includes a whole cluster of derived scientific and applied problems, one of which is the

need to investigate the processes of subjectivization of the perception of researched support object (that can be represented by the supported software complex itself, as well as by processes related to its support) by relevant subjects which interact with the first (the supported object) directly or indirectly and make influence on its support. Therefore, to solve this task, as well as a corresponding stack of relevant and/or derived tasks, the method, proposed and presented in this work, has been developed. This method provides the possibility of building/constructing appropriate models of multifactor portraits of investigated subjects supporting dedicated software complexes, using artificial neural networks, in particular, a multilayer perceptron. However, the multilayer perceptron itself, like any other artificial neural networks, does not contain ready-made solutions for solving the given problem/task, therefore, it is used as a complementary tool in the development of the method represented in this research. At the same time, the use of a multilayer perceptron makes it possible to additionally apply all its advantages, the main of which is its extremely wide range and potential of possibilities, with (at the same time) a comparatively simple and clear implementation.

The object of study is the process of forming multifactor portraits of subjects supporting software complexes, using a multilayer perceptron. **The subject of study** are methods and means of forming multifactor portraits of subjects supporting software complexes, using a multilayer perceptron.

The purpose of the work consists in the development of a method of forming multifactor portraits of the subjects supporting software complexes, using a multilayer perceptron, which (the developed method) provides possibility to investigate the processes of personal subjectivization of the perception of the researched support object by relevant subjects who interact with the first (directly or indirectly) and affect its support.

1 PROBLEM STATEMENT

Let's consider the formalization of the researched problem/task of forming a multifactor portrait(s) of the subject(s) supporting investigated software complex(es), in the form of a multi-parameter multifunctional dependence task.

In such case, the input variable of this task is the trained multilayer perceptron, encapsulated into the corresponding subjective perception model of the support object (e.g. supported software complex, or processes of its support) – $Omp[1..m]=Fmp[1..p][1..q](Imp[1..n])$, as well as a set of test cases (TC) (for each of the subjects of interaction with the object of support) for modeling/simulation the processes of subjectivization of the perception of a given support object by each individual subject interacting with it – $TC[1..s]=[i(1),...,i(n)]$.

The output variables of this task are the chains of neurons, activated in result of the simulation of each separate test case $TC[1..s][1..n]$ (for each separate researched subject from declared set of subjects $[1..s]$).

Let a set of test cases, representing the processes of subjectivization of the perception of support object by the relevant subjects interacting with it, is given. After performing simulation of current test case by the existing pre-trained multilayer perceptron artificial neural network, we will get the corresponding chain of neurons activated on each of the layers of this multilayer perceptron: starting from the input layer of neurons $I[1..n]$, continuing with the neurons of all hidden layers of this MP – $H[1..q]$, and finalizing with neurons of the output layer of this MP – $O[1..m]$.

$$I[x] \rightarrow H[I][y(1)] \rightarrow \dots \rightarrow H[q][y(q)] \rightarrow O[z], \quad (1)$$

The main necessary criterion of the researched task is the finiteness of the sets of: researched subjects, modeling test cases, as well as predefined and declared impact factors, which is caused by the possibility of operating (in this specific kind of problem) only by the finite sets of these determined parameters and their values.

Limitation(s) of the task:

1. The data values of absolutely all test cases (for absolutely all investigated subjects) must be presented in the format of real numbers and (obviously) in a normalized form of representation (which means: in the range of values between 0.0 and 1.0): $TC[1..s][1..n] \in [0..1]$.

Expression (1) provides possibility to represent the resulting partial (instant, or “one-moment”) interpretation of the researched support object by current considered/investigated subject.

However, this format of subjective interpretation of the researched support object represents only a partial (for only one specific test case) and at the same time combined (in the context of defined impact factors) influence of all the impact factors present in this interpretation in general, without any decomposition of this subjective interpretation partially by each (specifically) of the declared impact factors.

Thus, raises corresponding scientific and applied task of identifying and determining personalized complex indicators of presence each of the impact factors in the processes of personal subjectivization of the researched supported object’s perception by the relevant subjects interacting with it and making influence on its support. Or in other words, this task can be interpreted as: the task of forming a multifactor portraits of the researched subjects, participating in support of dedicated software complexes. Thus, precisely to solve this scientific and applied task – the method of forming multifactor portraits of the subjects supporting software complexes, using a multilayer perceptron, has been developed and presented in this research.

The purpose of the paper is to highlight the developed method of forming multifactor portraits of the subjects supporting software complexes, using a multilayer perceptron artificial neural networks, as well as the corresponding developed algorithm for building/constructing relevant models, which, taken together, provides all needed possibilities for solving researched scientific and

applied task of identifying and determining personalized complex indicators of presence each of the impact factors in the processes of personal subjectivization of the researched supported object’s perception by the relevant subjects interacting with it.

2 REVIEW OF THE LITERATURE

The analysis of relevant researches and materials was carried out in two key directions: in the direction of studying the processes of subjective perception of the objects, as well as in the direction of software complexes’ support automation (in particular, using AI and machine learning approaches). Resulted information on each direction is presented below.

For example, the authors of work [1] investigate the problem of distortion of the subject’s perception of the researched objects in situations with presence of a “noise” (or various stimuli, which could be also represented by certain considered external impact factors) inside the main visual channel, through which these subjects receive all the necessary information about the researched object. So, investigation, performed in this work, actually, confirmed the influence of personal (internal) impact factors (in particular, specifically the “expectation” factor) of the subjects on the perception of the researched object, because, according to the results presented in this work: the perception of reasonably expected objects was enhanced (heightened) relatively to unexpected objects, when the visual data were unreliable (which means, these data were more influenced by present external impact factors), while this effect was changing in favor of unexpected objects when the signal was more reliable (which means, it was less exposed to external impact factors).

At the same time, authors of works [2] and [3] study in detail the problem of subjectivization of user’s perception of software products’ graphical interfaces. Since the correct approach to supporting the graphical interface of the developed software products requires taking into account and consideration the subjective perception of users and the principles of objects grouping (such as Gestalt laws), because users usually group simple objects (graphic elements and shapes) into connected visually dominant objects. In addition, according to the results of those authors’ research, measuring the characteristics of visually emphasized objects of the graphical interface (of the supported software product), displayed on the screen, is a promising way for evaluation the quality of the user interface, although, on the other hand, it brings additional problems related to the ambiguity of object’s recognition caused by the subjective perception of different users. That is why the authors analyzed the applicability of selected object-oriented metrics for quality assessment, in particular, the toolbars (as a key element of the graphical user interface) and the ability to distinguish well-designed samples, focusing mainly and precisely on the subjective perception of users.

While, another work [4] is already devoted to the development of a specialized algorithm for automating the processing of client requests (for support services) based

on their prioritization and the formation of an appropriate queue, to ensure their fair and balanced processing. Also, in the work [5], the authors, based on a review of the relevant specialized literature sources, conducted a generalized analysis of the current state of affairs in the field of client requests' (for support services) processing automation in order to identify certain trends or tendencies, according to which, for example: the creation of automated incident management tools is a major topic in this field, followed by escalating inquiries and predicting (forecasting) customers' sentiments, as well as various appropriate algorithms for a very wide variety of classifications in this field. In addition, the author of the work [6] presented research aimed at creation of an intelligent user "LAVA" – a solution that provides context-dependent support for users of a supported software product, taking into account the interaction with these users, and is based on machine learning and searching in a dedicated two-level distributed knowledge repository (e.g. knowledge base). At the same time, works [7] and [8] study in detail the problems of automating the process of classifying customer requests received in dedicated systems of support services, because it is an extremely time-consuming process in case it is performed manually, while the use of machine learning gives really good results when solving tasks of such type (e.g. classification tasks). At the same time, work [9] examines the problems of processing client's requests (received in support services) in the context of using machine learning algorithms (namely in the part of natural language processing), which are fundamental for optimizing the processing of requests in a wide variety of customer support services and systems. Also, it is worth noting the works [10], [11] and [12], which reveal, in fact, the basic issues of both the fundamental theoretical principles and principles of functioning, as well as the practical application, of an artificial neural networks and a multilayer perceptron, which provides possibilities for better understanding and realizing their full potential, both in the context of possible use in the field of software complexes support, and as a whole.

3 MATERIALS AND METHODS

The main functional purpose of the developed method of forming multifactor portraits of the subjects supporting software complexes, with the use of a multilayer perceptron, is to ensure the possibility of identification and determination personalized complex indicators of presence each of the impact factors (within a relevant pre-agreed and declared set of impact factors) in the processes of personal subjectivization of the researched supported object's perception by the relevant subjects interacting with it (directly or indirectly) and making influence on its support.

The main idea of the developed method, which provides possibility to achieve the declared goal and task of this research, is to use the results of another researches [13], [14] and [15] in order to ensure the possibility of solving the declared task of identification and determination personalized complex indicators of presence each of

the impact factors in the processes of personal subjectivization of the researched supported object's perception by the relevant subjects interacting with it. In particular, researches [13] and [14] provides possibilities to identify the probabilities of belonging of the hidden layers' neurons (of the encapsulated trained multilayer perceptron artificial neural network) to the corresponding declared impact factors. Meanwhile, research [15] provides the possibility of modeling these processes using R-system and Python environments.

After all, having the specific numerical values of probability indicators of belonging each of the hidden layers' neurons (of the encapsulated trained MP ANN) to the appropriate relevant impact factors, it becomes possible to further step-by-step construction of a multifactor portrait of the corresponding researched subject. Since each chain of activated neurons obtained using expression (1) is, in fact, an instant slice of the researched subject's multifactor portrait:

$$SubjMFP[s]Inst[c] = [\sum[c](F[1]), \dots, \sum[c](F[f])]. \quad (2)$$

Thus, after fully completing the simulation for all available test cases belonging to one particular specific researched subject, we will obtain a corresponding set of various (perhaps, even partially repeated) chains of neurons activated during the simulation of each of these test cases (exactly for this one separate specific researched subject):

$$SubjMFP[s] = [\sum[1..n](F[1])/n, \dots, \sum[1..n](F[f])/n]. \quad (3)$$

Expression (2) provides the possibility of interpreting the researched task by usage of relevant results obtained by simulation of previously trained multilayer perceptron artificial neural network, where:

– $\sum[c](F[1])$ – the sum of the belonging of all activated neurons (in simulation results of the current test case $[c]$) to the first one impact factor $F[1]$ from a pre-declared set of defined impact factors;

– $\sum[c](F[f])$ – the sum of the belonging of all activated neurons (in simulation results of the current test case $[c]$) to the last one impact factor $F[f]$ from a pre-declared set of defined impact factors.

Meanwhile, expression (3), in turn, provides the possibility of interpretation the researched task of forming multifactor portraits of the subjects supporting the investigated object (the supported software complex, or processes of its support), where:

– $\sum[1..n](F[1])$ – the sum of the belongings of all activated neurons (in obtained simulation results of all test cases $[1..n]$ for current researched subject) to the first one impact factor $F[1]$ from a pre-declared set of defined impact factors;

– $\sum[1..n](F[f])$ – the sum of the belongings of all activated neurons (in obtained simulation results of all test cases $[1..n]$ for current researched subject) to the last one impact factor $F[f]$ from a pre-declared set of defined impact factors;

– n – the number (amount) of test cases for the current researched subject.

In addition, expression (3) also provides additional normalization of the obtained resulting data.

Thus, expression (3) actually represents the corresponding relevant mathematical model of the multi-factor

portrait of the subjects supporting investigated software complexes.

Below, in the Figure 1, a corresponding generalized block diagram of the developed algorithm of forming a multifactor portrait of the researched subject interacting with the investigated supported software complex is represented.

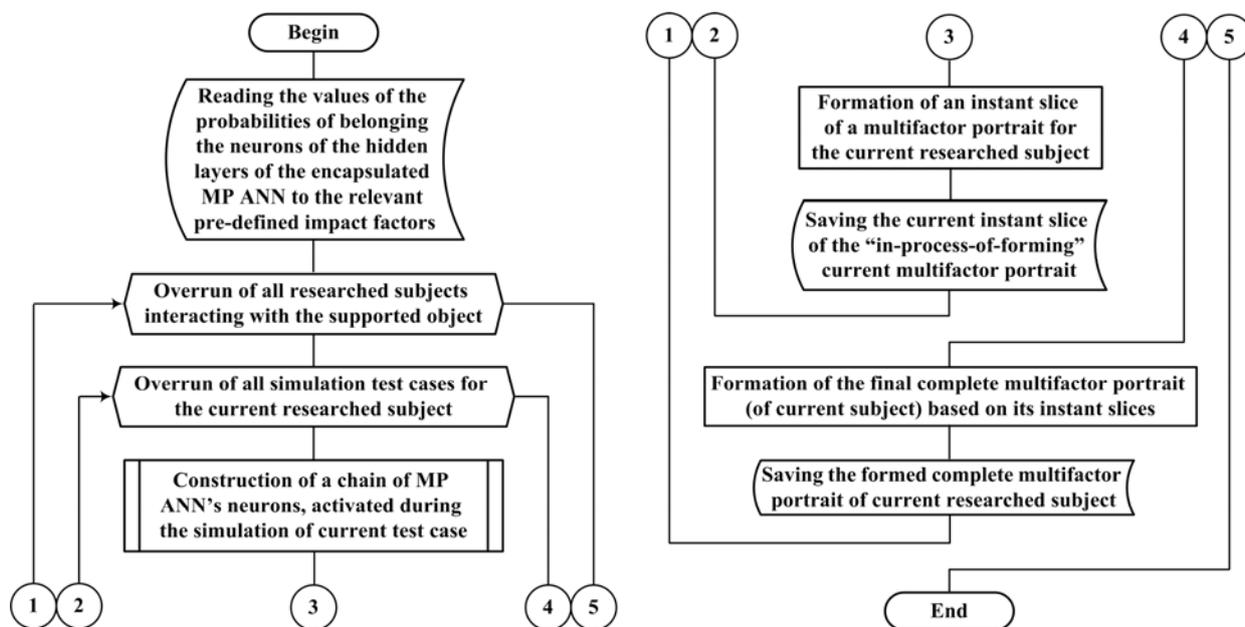


Figure 1 – Block diagram of the developed algorithm of forming a multifactor portrait(s) of the researched subject(s) interacting with the investigated supported software complex

The developed algorithm, block diagram of which is presented in Figure 1, as well as the corresponding mathematical model represented by expression (3), actually, create the needed basis of the developed method of forming multifactor portraits of the subjects supporting software complexes, with usage of a multilayer perceptron.

While the corresponding: mathematical model (represented in research [13]), the model of decomposed isolation dominance (represented in research [14]), as well as the appropriate information model (represented in research [15]) form up a complementary, but, at the same time, absolutely integral and irreplaceable component part of this method.

So, taken together, all these abovementioned constituent elements form a holistic method which provide possibility to resolve the declared scientific and applied task, namely: the task of identification and determination of personalized comprehensive indicators of presence each of the impact factors in the processes of personal subjectivization of the researched supported object’s perception by the relevant subjects interacting with it and making influence on its support, is being considered in this research.

4 EXPERIMENTS

The experiment consists in a step-by-step execution of all the stages described in this research, namely: correct

© Pukach A. I., Teslyuk V. M., 2025
 DOI 10.15588/1607-3274-2025-1-12

formulation of the researched task with all the necessary input data, as well as mandatory compliance with the declared restrictions; preparation and encapsulation of corresponding artificial neural network of a multilayer perceptron type, training this model, and obtaining all necessary results of the specific numerical values of probabilities of the hidden layers’ neurons’ belonging to the corresponding pre-declared impact factors; preparation of test cases for modeling the subjectivization processes of the perception of support object by all researched subjects (separately by each of these subjects); step-by-step execution of the developed algorithm of forming a multifactor portrait(s) of each of the studied/researched subjects by forming the corresponding instant slices of these (“in-process-of-constructing”) multifactor portrait(s), as well as the final full and complete multifactor portrait(s) of the subject(s) based on the relevant developed and proposed mathematical model.

5 RESULTS

The main results of functioning of the developed method – are formed/constructed multifactor portraits of the subjects supporting investigated object (e.g. the supported software complex, or processes of its support). In fact, each these obtained in such a way multifactor portraits (for each particular researched subject) represents a set (or array) of personalized comprehensive indicators of presence each of the pre-declared impact factors in proc-



esses of personal subjectivization of perception of the investigated support object by the corresponding relevant researched subject(s).

Thus, any multifactor portrait of any researched subject of interaction with an investigated supported software complex can be presented using the following appropriate expression (4):

$$SubjPort[x]=[Presence(F[1]), \dots, Presence(F[f])] \quad (4)$$

So, in accordance with the proposed generalized representation of a multifactor portrait(s) of the researched subject(s), formalized using expression (4), any multifactor portrait could be, in fact, represented by a corresponding set (array) consisting of elements, where:

– $Presence(F[1])$ – indicator of presence of the first (from a relevant set of all the pre-declared impact factors) impact factor $F[1]$ in current considered multifactor portrait $SubjPort[x]$;

– $Presence(F[f])$ – indicator of presence of the last (from a relevant set of all the pre-declared impact factors) impact factor $F[f]$ in current considered multifactor portrait $SubjPort[x]$.

Also, all these obtained resulting data of indicators of presence each of the impact factors in the researched multifactor portrait are already pre-normalized by using an expression (3).

Visually, any researched multifactor portrait is best perceived in the format of any standard diagram or histogram. In particular, below, in Figure 2, an example of visual interpretation of the obtained multifactor portraits of the researched subjects is presented, precisely with the help of a standard set of tools for constructing diagrams and/or histograms.

Also, as additional part of the research, carried out in this paper, the developed method has been tested on the example of resolving the specific applied practical task of

automated search and selection of the candidate (from among the members of corresponding relevant support team of the supported software complex) which would be the best in resolving the stack of specialized client requests/task (related to supporting activities for this software complex).

The solution of the given applied practical task is reduced to the application of the developed method for purpose of forming appropriate multifactor portraits of the researched subjects: members of the support team of given software complex, followed by their comparison with a certain “reference” multifactor portrait of some “ideal” candidate, the most appropriate, and maximally suitable and/or adapted for solving the given stack of specialized client requests related to the support of this software complex.

Therefore, the member of the support team (of considered software complex), whose multifactor portrait is maximally comparable to the reference one – will be the most suitable candidate for delegating him the appropriate stack of specialized client requests regarding the support of this software complex.

Below, in Figure 3, appropriate personal multifactor portraits of the research subjects (of the applied practical task being solved) are presented, using the developed method, as well as the results of their comparison with the reference multifactor portrait of the “ideal” candidate, both with the selected result of this comparison and the corresponding maximal suitable candidate from among the researched subjects (which, looking ahead, appeared to be “Subject 12”).

In addition, below, in Figure 4, the results of solving this applied practical problem are presented, visualized in a convenient form of representation with the help of a comparative histogram.

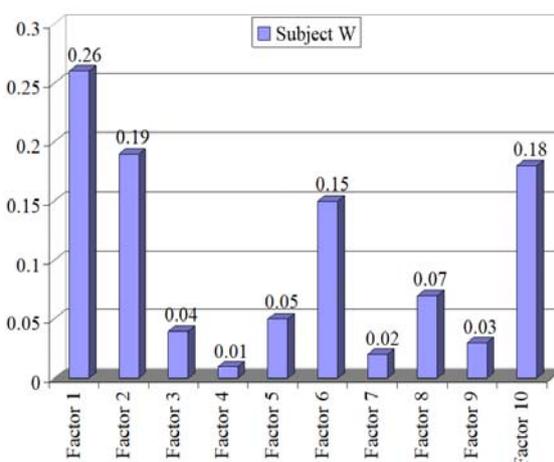
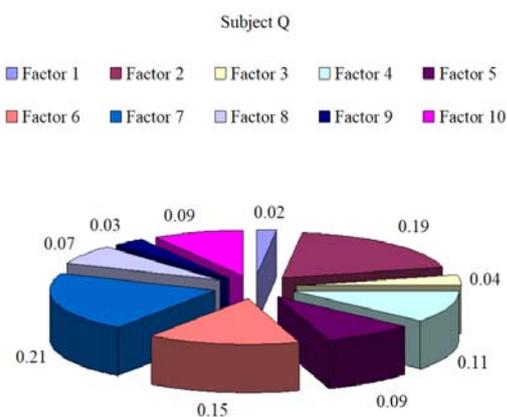


Figure 2 – Visualization examples of formed multifactor portraits of the researched subject using a standard set of tools for constructing diagrams and/or histograms

Portraits	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5	Factor 6	Factor 7	Factor 8	Factor 9	Factor 10	
Subject 1	0.045	0.049	0.116	0.098	0.134	0.179	0.175	0.067	0.028	0.109	
Subject 2	0.037	0.092	0.187	0.121	0.082	0.076	0.164	0.045	0.078	0.118	
Subject 3	0.114	0.033	0.108	0.090	0.084	0.035	0.084	0.091	0.299	0.062	
Subject 4	0.126	0.055	0.095	0.102	0.087	0.079	0.117	0.124	0.107	0.108	
Subject 5	0.034	0.177	0.090	0.149	0.146	0.088	0.065	0.056	0.102	0.093	
Subject 6	0.046	0.113	0.049	0.096	0.152	0.076	0.064	0.123	0.164	0.117	
Subject 7	0.095	0.060	0.162	0.228	0.073	0.068	0.035	0.086	0.106	0.087	
Subject 8	0.084	0.033	0.143	0.026	0.045	0.104	0.301	0.123	0.104	0.037	
Subject 9	0.048	0.132	0.229	0.042	0.162	0.068	0.037	0.067	0.152	0.063	
Subject 10	0.126	0.172	0.076	0.036	0.084	0.140	0.038	0.152	0.083	0.093	
Subject 11	0.045	0.101	0.165	0.095	0.141	0.119	0.084	0.095	0.095	0.060	
Subject 12	0.087	0.067	0.099	0.183	0.122	0.095	0.068	0.044	0.128	0.107	
Subject 13	0.058	0.107	0.138	0.124	0.109	0.076	0.094	0.088	0.117	0.089	
Subject 14	0.054	0.065	0.083	0.071	0.072	0.097	0.127	0.208	0.105	0.118	
Subject 15	0.131	0.063	0.203	0.121	0.083	0.115	0.046	0.074	0.086	0.078	
Reference	0.089	0.058	0.107	0.202	0.064	0.092	0.106	0.049	0.086	0.147	
ComparisonABS	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5	Factor 6	Factor 7	Factor 8	Factor 9	Factor 10	Sum
Subj.1 vs Refer.	0.044	0.009	0.009	0.104	0.07	0.087	0.069	0.018	0.058	0.038	0.506
Subj.2 vs Refer.	0.052	0.034	0.08	0.081	0.018	0.016	0.058	0.004	0.008	0.029	0.380
Subj.3 vs Refer.	0.025	0.025	0.001	0.112	0.02	0.057	0.022	0.042	0.213	0.085	0.602
Subj.4 vs Refer.	0.037	0.003	0.012	0.1	0.023	0.013	0.011	0.075	0.021	0.039	0.334
Subj.5 vs Refer.	0.055	0.119	0.017	0.053	0.082	0.004	0.041	0.007	0.016	0.054	0.448
Subj.6 vs Refer.	0.043	0.055	0.058	0.106	0.088	0.016	0.042	0.074	0.078	0.03	0.590
Subj.7 vs Refer.	0.006	0.002	0.055	0.026	0.009	0.024	0.071	0.037	0.02	0.06	0.310
Subj.8 vs Refer.	0.005	0.025	0.036	0.176	0.019	0.012	0.195	0.074	0.018	0.11	0.670
Subj.9 vs Refer.	0.041	0.074	0.122	0.16	0.098	0.024	0.069	0.018	0.066	0.084	0.756
Subj.10 vs Refer.	0.037	0.114	0.031	0.166	0.02	0.048	0.068	0.103	0.003	0.054	0.644
Subj.11 vs Refer.	0.044	0.043	0.058	0.107	0.077	0.027	0.022	0.046	0.009	0.087	0.520
Subj.12 vs Refer.	0.002	0.009	0.008	0.019	0.058	0.003	0.038	0.005	0.042	0.04	0.224
Subj.13 vs Refer.	0.031	0.049	0.031	0.078	0.045	0.016	0.012	0.039	0.031	0.058	0.390
Subj.14 vs Refer.	0.035	0.007	0.024	0.131	0.008	0.005	0.021	0.159	0.019	0.029	0.438
Subj.15 vs Refer.	0.042	0.005	0.096	0.081	0.019	0.023	0.06	0.025	0	0.069	0.420

Figure 3 – Formed personal multifactor portraits of the researched subjects, as well as the results of their comparison with the corresponding reference multifactor portrait of the ideal candidate

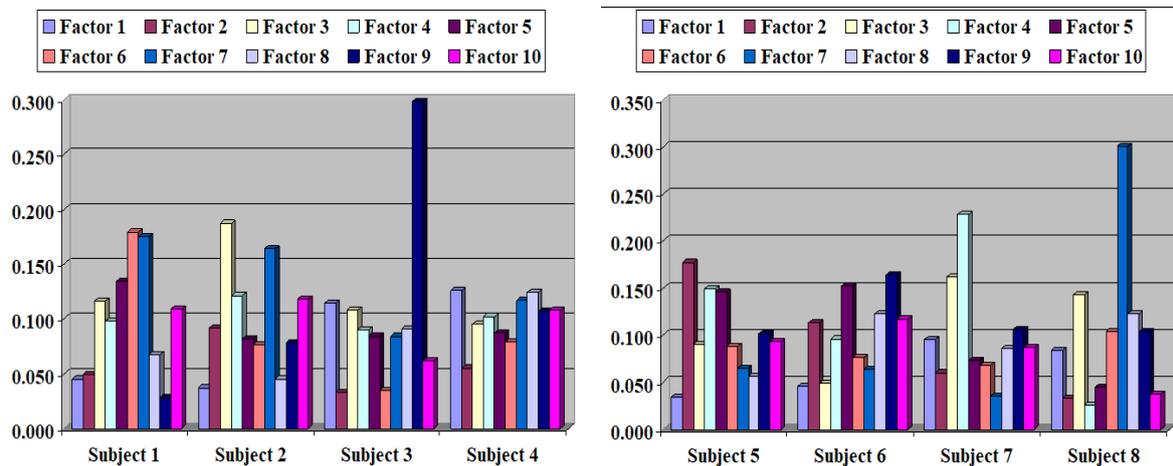


Figure 4 – The results of solving an applied practical task, obtained using the developed method, and visualized in a convenient form of representation by the usage of a comparative histogram

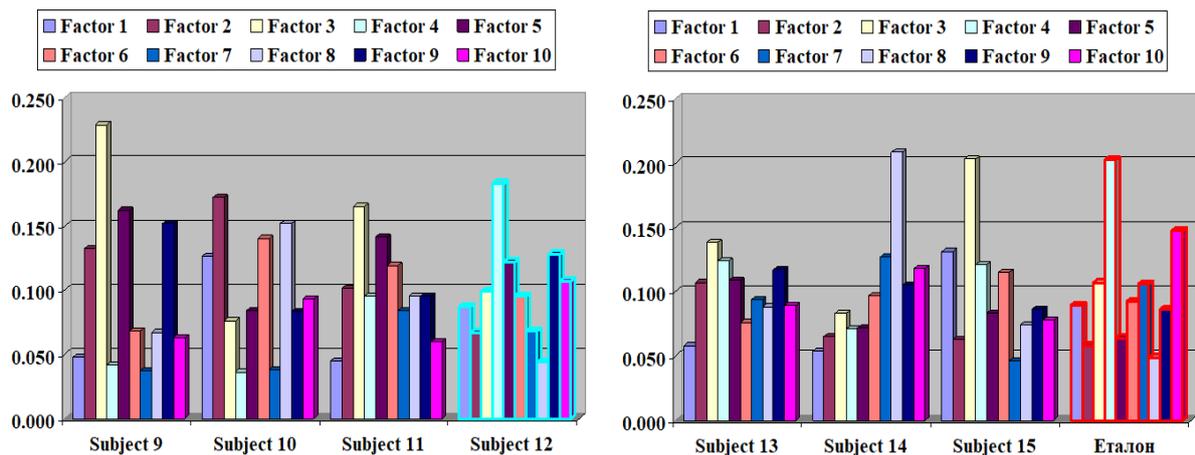


Figure 4 – The results of solving an applied practical task, obtained using the developed method, and visualized in a convenient form of representation by the usage of a comparative histogram (Continuation)

Thus, as an example of approbation and practical application of the developed method, the results of solving the applied practical task of automated search and selection of the candidate (from among the members of corresponding relevant support team of the supported software complex) which would be the best in resolving the stack of specialized client requests/task (related to supporting activities for this software complex), have been presented in this research.

In general, the obtained results testify to the suitability and effectiveness of the developed method for solving the declared scientific and applied task of identification and determination of personalized comprehensive indicators of presence each of the impact factors (from a pre-agreed and declared set of impact factors) in the processes of personal subjectivization of the researched supported object’s perception by the relevant subjects interacting with it (directly or indirectly) and making influence on its support.

6 DISCUSSION

In the scope of research [16], a comprehensive study has been carried out aimed mainly at identifying and analyzing the most modern and actual methods, algorithms and sensor technologies in existing studies of the intellectual interaction of a person with a computer (e.g. more known ask a “human computer interaction” – HCI), the topic of trends in the area of this interaction, as well as potential directions for future researches in this area. In particular, it is also going about the corresponding intelligent user interfaces of the various used software products and complexes. Research in both these directions is mainly focused on intelligent recognition of emotions, gestures and facial expressions of users, involving a sensor technologies (such as camera, portable sensors / transducers, trackers, gyroscopes, etc.). At the same time, researchers and developers most often use the intellectual solutions they need – based on the methods, algorithms and sensor technologies of artificial intelligence, machine learning, artificial neural networks, including a multilayer perceptron.

At the same time, another research [17] presents an overview of research focused on a universal usability, plasticity of user interface design, and facilitating the development of an interfaces with universal usability. The basics of adaptive or intelligent user interfaces (of various software products and complexes) are presented, focused on three main areas: artificial intelligence, user modeling, and intelligent human-computer interaction. According to the results of this research, the interface of the developed and supported software product or complex should take into account the individual characteristics of users, based rather on the actual behavior of these users, then on their feedback about the supported software product. Also, the need for a more complete understanding of the human psychological model of the researched subjects (users) is emphasized, which requires a high-quality interdisciplinary approach and cooperation between different fields’ researchers.

Additionally, the research [18] also investigates the interaction of subjects (users) with modern intelligent technological systems, which (interaction) goes far beyond just the concept of convenience of supported software products usage, extending to a person's emotions before, during and after using a system (e.g. software product / complex), and cannot be determined only by examining the main attributes of usability, such as the effectiveness and satisfaction of the actors (i.e. users) interacting with that system. Therefore, such research requires more intelligent and complex approaches, including artificial intelligence technologies as well as considering and taking into account the approach of a general intelligent computing environment.

Thus, the existing provided researches confirm the actuality, relevance and necessity of studying and taking into account the peculiarities of subjects interacting with developed and supported software products, complexes, or systems. At the same time, they study various options for taking into account these features of the subjects that interact (and influence accordingly) the developed and supported software products, systems and complexes. However, in none of the researched cases, the influence of

relevant factors, which, precisely, make impact onto these subjects themselves, is being considered and appropriately taken into account.

Meanwhile, the method of forming multifactor portraits of subjects supporting software complexes, developed and presented in current research, provides possibilities for carrying out the necessary identification and determination of personalized comprehensive indicators of presence each of the impact factors (from a pre-agreed and declared set of impact factors) in the processes of personal subjectivization of the researched supported object's (e.g. supported software product, complex, systems, as well as the processes of its support) perception by the relevant subjects interacting with it.

Thus, at the output of its functional operation, the developed method ensures obtaining a relevant suitable model(s) of the multifactor portrait(s) of the researched subject(s), which, later, makes it possible to take it into account when researching and resolving various related scientific and applied tasks of a more complex scientific and applied problem of software complexes support automation.

As a further practical application of the developed method, we can see the solution of a stack of relevant applied practical tasks, among which, for example, are such as:

- development of models of various subjects interacting with the researched and supported software complex(es);
- adjustment of personal characteristics, skills, and features of the researched subjects in order to correct their multifactor portrait(s) in the necessary vector of requirements;
- development of personal instructions for subjects in order to improve their characteristics, including, in particular, professional ones;
- research on the compatibility of various (interacting each other) subjects based on their multifactor portraits;
- and others.

Thus, taking into account a wide range of applied tasks, the expediency of further research in this direction is fully justified.

In addition, the developed method can also be used in other areas of science and practice, in particular, for example, in such fields as psychology and sociology, where the formation of various multifactor portraits is an extremely important stage in the investigation of any researched subjects.

CONCLUSIONS

The method of forming multifactor portraits of subjects supporting software complexes, using a multilayer perceptron, is developed. The main scientific and applied task, resolved by the developed method, is the task of identification and determination of personalized comprehensive indicators of presence each of the impact factors (from a pre-agreed and declared set of impact factors) in the processes of personal subjectivization of the researched supported object's (e.g. supported software

© Pukach A. I., Teslyuk V. M., 2025
DOI 10.15588/1607-3274-2025-1-12

complex, as well as the processes of its support) perception by the relevant subjects interacting with it (both directly or indirectly) and making influence on its support. The input data of the developed method of forming multifactor portraits of the software complex support subjects – is the relevant trained multilayer perceptron ANN, encapsulated into a corresponding model of subjective perception of the investigated support object (the supported software complex, or processes of its support). as well as a set of test cases (for each of the researched subjects interacting with the object of support) for modeling the processes of subjectivization of the supported object's perception by each separate researched subject interacting with this object.

The corresponding developed algorithm of forming a multifactor portrait(s) of researched subject(s) (interacting with the investigated supported software complex) is presented, as well as a corresponding mathematical model of multifactor portrait(s) of these subject(s) of interaction, which form the basis of the developed method. While the corresponding: mathematical model (represented in research [13]), the model of decomposed isolation dominance (represented in research [14]), as well as the appropriate information model (represented in relevant research [15]) – make up a complementary, but absolutely integral comprehensive and irreplaceable component of the developed method. So in aggregate, all these constituent elements form a complete method, which provide possibility to solve the declared scientific and applied task of this research.

The basic forms of representation of the multifactor portrait(s)' models of subjects (interacting with the supported software complexes) are presented, which includes: an appropriate developed mathematical form of representation, a visual form of representation using a standard set of tools for constructing diagrams and/or histograms; as well as the given examples of simulation results.

The developed method provides the possibility of building a relevant multifactor portraits' models of the researched subjects interacting with the investigated supported software complexes, as well as any other objects (or processes) of software products' complex support.

The scientific novelty consists in the development of a method of forming multifactor portraits of the subjects supporting program complexes, which provides possibilities for studying both, in general: the impact factors performing influence onto the investigated software complexes and their support, as well as the processes of perception's subjectivization of the investigated object of support by the relevant researched subjects who interact with the first (directly or indirectly), and affect its support.

The practical significance consists in the development of an appropriate specialized dedicated algorithm of forming a multifactor portrait(s) of the researched subject(s) interacting with the investigated supported software complex(es), as well as a corresponding mathemati-

cal model of the multifactor portraits of these researched subjects.

Prospects for further research consist in the development of appropriate additional specialized algorithmic and programming tools for modeling multifactor portraits of the researched subjects interacting with the investigated supported software complexes and products, as well as in the further application of the developed models in the fields of studying both the influence of impact factors as well as the relevant processes of perception's subjectivization of the investigated support object by the relevant subjects who interact with it, both in the direction of software complexes' support automation, and in other related (or possible potentially related) areas of scientific and applied research.

ACKNOWLEDGEMENTS

Work is proactive. The research was carried out as part of the scientific activity of the authors outside of working hours in their main positions.

REFERENCES

1. Rossel P., Peyrin C., Kauffmann L. Subjective perception of objects depends on the interaction between the validity of context-based expectations and signal reliability [Electronic resource], *Vision Research*, 2023, Vol. 206, article ID 108191. Mode of access: <https://doi.org/10.1016/j.visres.2023.108191> (date of access: 30.09.2024). Title from screen.
2. Hynek J., Hruška T. Segmentation of Dashboard Screen Images: Preparation of Inputs for Object-based Metrics of UI Quality [Electronic resource], *In Proceedings of the 14th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications*, 2019, pp. 199–207. Mode of access: <https://doi.org/10.5220/0007312301990207> (date of access: 30.09.2024). – Title from screen.
3. Hynek J., Hruška T. Application of Object-Based Metrics for Recognition of Well-Designed Dashboards [Electronic resource], *International Journal of Human-Computer Interaction*, 2018, Vol. 35, No. 13, pp. 1203–1215. Mode of access: <https://doi.org/10.1080/10447318.2018.1518004> (date of access: 30.09.2024). Title from screen.
4. Clarin J. A. Priority-Based Scheduling Algorithm for Help Desk Support System [Electronic resource], *International Journal of Intelligent Systems and Applications in Engineering*, 2023, Vol. 11, No. 4, pp. 299–307. Mode of access: https://www.researchgate.net/profile/Jeffrey-Clarin/publication/374535831-International_Journal_of_INTELLIGENT_SYSTEMS_AND_APPLICATIONS_IN_ENGINEERING_Priority-Based_Scheduling_Algorithm_for_Help_Desk_Support_System/links/65226ad2fc5c2a0c3bc0afe2/International-Journal-of-INTELLIGENT-SYSTEMS-AND-APPLICATIONS-IN-ENGINEERING-Priority-Based-Scheduling-Algorithm-for-Help-Desk-Support-System.pdf (date of access: 30.09.2024). – Title from screen.
5. Fuchs S., Drieschner C., Wittges H. Improving Support Ticket Systems Using Machine Learning: A Literature Review [Electronic resource], *In Proceedings of the 55-th Hawaii International Conference on System Sciences*, 2022, pp. 1893–1902. Mode of access: <https://doi.org/10.24251/HICSS.2022.238> (date of access: 30.09.2024). – Title from screen.
6. Baltic J. Structuring and Controlling the Knowledge for the Software User Support [Electronic resource], *Modern Computing*, 2021, Vol. 9, No. 2, pp. 195–209. Mode of access: <https://doi.org/10.22364/bjmc.2021.9.2.04> (date of access: 30.09.2024). Title from screen.
7. Dlodlo D., Sibanda K. Automated Ticket Classification for Information Technology Helpdesks using Machine Learning [Electronic resource], *2nd Zimbabwe Conference of Information and Communication Technologies (ZCICT)*, 2023, pp. 1–7. Mode of access: <https://doi.org/10.1109/ZCICT59466.2023.10528578> (date of access: 30.09.2024). Title from screen.
8. Gandla P. K. K., Verma R. K., Panigrahi C. R. et al. Ticket Classification Using Machine Learning [Electronic resource], *In Proceedings of the 7th International Conference on Advance Computing and Intelligent Engineering, Lecture Notes in Networks and Systems*, 2024, Vol. 1, pp. 487–501. Mode of access: https://doi.org/10.1007/978-981-99-5015-7_41 (date of access: 30.09.2024). Title from screen.
9. Zangari A., Marcuzzo M., Schiavinato M. et al. Ticket automation: An insight into current research with applications to multilevel classification scenarios [Electronic resource], *Expert Systems with Applications*, 2023, Volume 225, article ID 119984. Mode of access: <https://doi.org/10.1016/j.eswa.2023.119984> (date of access: 30.09.2024). – Title from screen.
10. Chakraverty S., Mall S. Artificial Neural Networks for Engineers and Scientists: Solving Ordinary Differential Equations (1st ed.) [Electronic resource], CRC Press, 2017, 168 p. Mode of access: <https://doi.org/10.1201/9781315155265> (date of access: 30.09.2024). Title from screen.
11. Samarasinghe S. Neural Networks for Applied Sciences and Engineering: From Fundamentals to Complex Pattern Recognition (1st ed.) [Electronic resource], *Auerbach Publications*, 2006, 570 p. Mode of access: <https://doi.org/10.1201/9780849333750> (date of access: 30.09.2024). Title from screen.
12. Vang-Mata R. Multilayer Perceptrons: Theory and Applications [Electronic resource]. Hauppauge, New York, Nova Science Publishers, 2020, 153 p. Mode of access: <https://novapublishers.com/shop/multilayer-perceptrons-theory-and-applications/> (date of access: 30.09.2024). Title from screen.
13. Pukach A. I., Teslyuk V. M. Mathematical model for analysis of influencing factors on software complexes support [Electronic resource], *Printing and Publishing*, 2024, 1(87), pp. 75–85. Mode of access: <https://doi.org/10.32403/0554-4866-2024-1-87-75-85> (date of access: 01.10.2024). Title from screen.
14. Pukach A. I., Teslyuk V. M. Model of decomposed insulating dominance for the analysis of influencing factors of software complexes support automation [Electronic resource], *Scientific Bulletin of UNFU*, 2024, 34(5). pp. 170–179. Mode of access: <https://doi.org/10.36930/40340521> (date of access: 01.10.2024). Title from screen.
15. Pukach A. I., Teslyuk V. M. Information model for automation of software complexes support influencing factors analysis with usage of the R-system and Python environments [Electronic resource], *Bulletin of Lviv State University of Life Safety*, 2024, 29, pp. 54–64. Mode of access: <https://doi.org/10.32447/20784643.29.2024.06> (date of access: 01.10.2024). Title from screen.

16. Šumak B., Brdnic S., Pušnik M. Sensors and Artificial Intelligence Methods and Algorithms for Human-Computer Intelligent Interaction: A Systematic Mapping Study [Electronic resource], *Sensors*, 2022, 22(1), 20, 40 p. Mode of access: <https://doi.org/10.3390/s22010020> (date of access: 01.10.2024). Title from screen.
17. Miraz M. H., Ali M., Excell P. S. Adaptive user interfaces and universal usability through plasticity of User Interface Design [Electronic resource], *Computer Science Review*, 2021, Vol. 40, article ID 100363. Mode of access: <https://doi.org/10.1016/j.cosrev.2021.100363> (date of access: 01.10.2024). Title from screen.
18. Ntoa S., Margetis G., Antona M. et al. User Experience Evaluation in Intelligent Environments: A Comprehensive Framework [Electronic resource], *Technologies*, 2021, 9(2), 41, 36 p. Mode of access: <https://doi.org/10.3390/technologies9020041> (date of access: 01.10.2024). – Title from screen.

Received 22.10.2024.
Accepted 14.12.2024.

УДК 004.8

МЕТОД ФОРМУВАННЯ МУЛЬТИФАКТОРНИХ ПОРТРЕТІВ СУБ'ЄКТІВ ПІДТРИМКИ ПРОГРАМНИХ КОМПЛЕКСІВ, З ВИКОРИСТАННЯМ БАГАТОШАРОВОГО ПЕРЦЕПТРОНА

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, проф., завідувач кафедри Автоматизованих Систем Управління Інституту Комп'ютерних Наук та Інформаційних Технологій Національного Університету «Львівська Політехніка», Львів, Україна.

АНОТАЦІЯ

Актуальність. Розглянуто проблему ідентифікації та визначення персоналізованих комплексних показників наявності кожного з факторів впливу в процесах особистісної суб'єктивізації сприйняття досліджуваного об'єкта підтримки відповідними суб'єктами, які з ним взаємодіють і впливають на його підтримку. Об'єктом дослідження є процес формування мультифакторних портретів суб'єктів підтримки програмних комплексів, з використанням багатопланового перцептрона. Предметом дослідження є методи та засоби формування мультифакторних портретів суб'єктів підтримки програмних комплексів.

Мета роботи – розроблення методу формування мультифакторних портретів суб'єктів підтримки програмних комплексів, з використанням багатопланового перцептрона.

Метод. Розроблено метод формування мультифакторних портретів суб'єктів підтримки програмних комплексів, з використанням штучних нейронних мереж типу багатопланового перцептрона, що дає змогу формувати відповідні персональні мультифакторні портрети суб'єктів, котрі, напряму або опосередковано, взаємодіють з об'єктом підтримки, в якості якого може виступати як сам підтримуваний програмний комплекс, так і процеси, пов'язані з його комплексною підтримкою.

Результати. Результатами роботи розробленого методу є відповідні моделі мультифакторних портретів суб'єктів підтримки програмних комплексів, які, в подальшому, використано для розв'язання кластеру науково-прикладних задач автоматизації підтримки програмних комплексів, зокрема, задачу ідентифікації та визначення персоналізованих комплексних показників присутності кожного із факторів впливу (наперед узгодженої та задекларованої множини факторів впливу) в процесах персональної суб'єктивізації сприйняття досліджуваного об'єкта підтримки відповідними суб'єктами, котрі взаємодіють з першим (напряму або опосередковано), та впливають на його підтримку. В якості прикладу практичного застосування розробленого методу, наведено результати розв'язання прикладної практичної задачі автоматизованого пошуку та підбору кандидата (з числа членів команди підтримки програми) для розв'язання стеку спеціалізованих клієнтських запитів (щодо підтримки цього програмного комплексу).

Висновки. Розроблений метод вирішує поставлену задачу ідентифікації та визначення персоналізованих комплексних показників присутності кожного із факторів впливу (наперед узгодженої та задекларованої множини факторів впливу) в процесах персональної суб'єктивізації сприйняття досліджуваного об'єкта підтримки відповідними суб'єктами, котрі взаємодіють з першим (напряму або опосередковано), та впливають на його підтримку. В доповнення, розроблений метод забезпечує можливість побудови відповідних моделей мультифакторних портретів суб'єктів підтримки програмних комплексів, що дає змогу використовувати їх при розв'язанні будь-яких задач, пов'язаних з автоматизацією пошуку та підбору суб'єктів підтримки програмних комплексів, що відповідають заданим критеріям як в контексті процесів суб'єктивізації персонального сприйняття об'єктів підтримки (підтримуваних програмних комплексів, чи процесів їх підтримки), так і в контексті сумісності у взаємодії з клієнтами та/або користувачами підтримуваних програмних продуктів (оскільки ці користувачі, фактично, теж являються суб'єктами взаємодії з тим же досліджуваним об'єктом підтримки).

КЛЮЧОВІ СЛОВА: автоматизація, програмний комплекс, підтримка, фактор впливу, мультифакторний портрет, нейронні мережі, багатоплановий перцептрон.

ЛІТЕРАТУРА

- Rossel P. Subjective perception of objects depends on the interaction between the validity of context-based expectations and signal reliability [Electronic resource] / P. Rossel, C. Peyrin, L. Kauffmann // *Vision Research*. – 2023. – Vol. 206. – article ID 108191. – Mode of access: <https://doi.org/10.1016/j.visres.2023.108191> (date of access: 30.09.2024). – Title from screen.
- Hynek J. Segmentation of Dashboard Screen Images: Preparation of Inputs for Object-based Metrics of UI Quality [Electronic resource] / J. Hynek, T. Hruška // In Proceedings of the 14th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications. – 2019. – P. 199–207. – Mode of access: <https://doi.org/10.5220/0007312301990207> (date of access: 30.09.2024). – Title from screen.

3. Hynek J. Application of Object-Based Metrics for Recognition of Well-Designed Dashboards [Electronic resource] / J. Hynek, T. Hruška // International Journal of Human-Computer Interaction. – 2018. – Vol. 35, No. 13. – P. 1203–1215. – Mode of access: <https://doi.org/10.1080/10447318.2018.1518004> (date of access: 30.09.2024). – Title from screen.
4. Clarin J. A. Priority-Based Scheduling Algorithm for Help Desk Support System [Electronic resource] / J. A. Clarin // International Journal of Intelligent Systems and Applications in Engineering. – 2023. – Vol. 11, No. 4. – P. 299–307. – Mode of access: https://www.researchgate.net/profile/Jeffrey-Clarin/publication/374535831_International_Journal_of_INTELLIGENT_SYSTEMS_AND_APPLICATIONS_IN_ENGINERING_Priority-Based_Scheduling_Algorithm_for_Help_Desk_Support_System/links/65226ad2fc5c2a0c3bc0afe2/International-Journal-of-INTELLIGENT-SYSTEMS-AND-APPLICATIONS-IN-ENGINEERING-Priority-Based-Scheduling-Algorithm-for-Help-Desk-Support-System.pdf (date of access: 30.09.2024). – Title from screen.
5. Fuchs S. Improving Support Ticket Systems Using Machine Learning: A Literature Review [Electronic resource] / S. Fuchs, C. Drieschner, H. Wittges // In Proceedings of the 55-th Hawaii International Conference on System Sciences. – 2022. – P. 1893–1902. – Mode of access: <https://doi.org/10.24251/HICSS.2022.238> (date of access: 30.09.2024). – Title from screen.
6. Baltic J. Structuring and Controlling the Knowledge for the Software User Support [Electronic resource] / J. Baltic // Modern Computing. – 2021. – Vol. 9, No. 2. – P. 195–209. – Mode of access: <https://doi.org/10.22364/bjmc.2021.9.2.04> (date of access: 30.09.2024). – Title from screen.
7. Dlodlo D. Automated Ticket Classification for Information Technology Helpdesks using Machine Learning [Electronic resource] / D. Dlodlo, K. Sibanda // 2nd Zimbabwe Conference of Information and Communication Technologies (ZCICT). – 2023. – P. 1–7. – Mode of access: <https://doi.org/10.1109/ZCICT59466.2023.10528578> (date of access: 30.09.2024). – Title from screen.
8. Ticket Classification Using Machine Learning [Electronic resource] / [P. K. K. Gandla, R. K. Verma, C. R. Panigrahi et al.] // In Proceedings of the 7th International Conference on Advance Computing and Intelligent Engineering, Lecture Notes in Networks and Systems. – 2024. – Vol. 1. – P. 487–501. – Mode of access: https://doi.org/10.1007/978-981-99-5015-7_41 (date of access: 30.09.2024). – Title from screen.
9. Ticket automation: An insight into current research with applications to multilevel classification scenarios [Electronic resource] / [A. Zangari, M. Marcuzzo, M. Schiavinato et al.] // Expert Systems with Applications. – 2023. – Volume 225. – article ID 119984. – Mode of access: <https://doi.org/10.1016/j.eswa.2023.119984> (date of access: 30.09.2024). – Title from screen.
10. Chakraverty S. Artificial Neural Networks for Engineers and Scientists: Solving Ordinary Differential Equations (1st ed.) [Electronic resource] / S. Chakraverty, S. Mall. – CRC Press, 2017. – 168 p. – Mode of access: <https://doi.org/10.1201/9781315155265> (date of access: 30.09.2024). – Title from screen.
11. Samarasinghe S. Neural Networks for Applied Sciences and Engineering: From Fundamentals to Complex Pattern Recognition (1st ed.) [Electronic resource] / S. Samarasinghe. – Auerbach Publications, 2006. – 570 p. – Mode of access: <https://doi.org/10.1201/9780849333750> (date of access: 30.09.2024). – Title from screen.
12. Vang-Mata R. Multilayer Perceptrons: Theory and Applications [Electronic resource] / R. Vang-Mata Hauppauge. – New York : Nova Science Publishers, 2020 p. – 153 p. – Mode of access: <https://novapublishers.com/shop/multilayer-perceptrons-theory-and-applications/> (date of access: 30.09.2024). – Title from screen.
13. Пукач А.І. Математична модель аналізу факторів впливу підтримки програмних комплексів [Електронний ресурс] / А. І. Пукач, В. М. Теслюк // Поліграфія і видавнича справа. – 2024. – 1(87). – С. 75–85. – Режим доступу: <https://doi.org/10.32403/0554-4866-2024-1-87-75-85> (дата звернення: 01.10.2024). – Назва з екрана.
14. Пукач А. І. Модель декомпонованої ізоляційної домінації для аналізу чинників впливу на автоматизацію підтримки програмних комплексів [Електронний ресурс] / А. І. Пукач, В. М. Теслюк // Науковий вісник НЛТУ України. – 2024. – 34(5). – С. 170–179. – Режим доступу: <https://doi.org/10.36930/40340521> (дата звернення: 01.10.2024). – Назва з екрана.
15. Пукач А. І. Інформаційна модель аналізу факторів впливу автоматизації підтримки програмних комплексів з використанням середовищ R та Python [Електронний ресурс] / А. І. Пукач, В. М. Теслюк // Вісник Львівського державного університету безпеки життєдіяльності. – 2024. – 29. – С. 54–64. – Режим доступу: <https://doi.org/10.32447/20784643.29.2024.06> (дата звернення: 01.10.2024). – Назва з екрана.
16. Šumak B. Sensors and Artificial Intelligence Methods and Algorithms for Human-Computer Intelligent Interaction: A Systematic Mapping Study [Electronic resource] / B. Šumak, S. Brdник, M. Pušnik // Sensors. – 2022. – 22(1), 20. – 40 p. – Mode of access: <https://doi.org/10.3390/s22010020> (date of access: 01.10.2024). – Title from screen.
17. Miraz M. H. Adaptive user interfaces and universal usability through plasticity of User Interface Design [Electronic resource] / M. H. Miraz, M. Ali, P. S. Excell // Computer Science Review. – 2021. – Vol. 40. – article ID 100363. – Mode of access: <https://doi.org/10.1016/j.cosrev.2021.100363> (date of access: 01.10.2024). – Title from screen.
18. User Experience Evaluation in Intelligent Environments: A Comprehensive Framework [Electronic resource] / [S. Ntoa, G. Margetis, M. Antona et al.] // Technologies. – 2021. – 9(2), 41. – 36 p. – Mode of access: <https://doi.org/10.3390/technologies9020041> (date of access: 01.10.2024). – Title from screen.

METHOD OF PREVENTING FAILURES OF ROTATING MACHINES BY VIBRATION ANALYSIS USING MACHINE LEARNING TECHNIQUES

Zalutska O. O. – Assistant of the Department of Computer Sciences, Khmelnytskyi National University, Khmelnytskyi, Ukraine.

Hladun O. V. – Student of the Department of Computer Sciences, Khmelnytskyi National University, Khmelnytskyi, Ukraine.

Mazurets O. V. – PhD, Associate Professor, Department of Computer Science, Khmelnytskyi National University, Khmelnytskyi, Ukraine.

ABSTRACT

Context. The problem of determining transitional conditions that precede the shift from an operating state to a non-operating state based on data obtained from the sensors of rotating machine elements is being solved. The object of the study is the process of detecting faults and states that indicate an approach to breakdown in rotating machine elements based on data obtained from sensors. The subject of the study is the application of k -means and the elbow method algorithms for clustering and convolutional neural networks for classifying sensor data and detecting near-failure states of machine elements.

Objective. The purpose of the work is to create a method for processing sensor data from rotating machines using convolutional neural networks to accurately detect conditions close to failure in rotating machine elements, which will increase the efficiency of maintenance and prevent equipment failures.

Method. The proposed method of preventing failures of rotating machines by vibration analysis using machine learning techniques using a combination of clustering and deep learning methods. At the first stage, the sensor data undergoes preprocessing, including normalization, dimensionality reduction, and noise removal, after which the K -means algorithm is applied. To determine the optimal number of clusters, the Elbow method is used, which provides an effective grouping of the states of rotating machine elements, identifying states close to the transition to fault. A CNN model has also been developed that classifies clusters, allowing for the accurate separation of nominal, fault, and transitional conditions. The combination of clustering methods with the CNN model improves the accuracy of detecting potential faults and enables timely response, which is critical for preventing accidents and ensuring the stability of equipment operation.

Results. A method of preventing failures of rotating machines by vibration analysis using machine learning techniques and a relevant software package have been developed. The implemented method allows us to identify not only normal and emergency states but also to distinguish a third class – transitional, close to breakdown. The quality of clustering for the three classes is confirmed by the value of the silhouette coefficient of 0.506, which indicates the proper separation of the clusters, and the Davis-Boldin index of 0.796, which demonstrates a high level of internal cluster coherence. Additionally, CNN was trained to achieve 99% accuracy for classifying this class, which makes the method highly efficient and distinguishes it from existing solutions.

Conclusions. A method of preventing failures of rotating machines by vibration analysis using machine learning techniques was developed, the allocation of the third class – transitional, indicating a state close to breakdown – was proposed, and its effectiveness was confirmed. The practical significance of the results lies in the creation of a neural network model for classifying the state of rotating elements and the development of a web application for interacting with these models.

KEYWORDS: rotation machines, element failure, transitional conditions, clustering, classification, CNN.

ABBREVIATIONS

AI is an artificial intelligence;
CAM is a computer aided manufacturing;
CNN is a convolution neural network;
DL is a deep learning;
LSTM is a long short-term memory;
ML is a machine learning;
NN is a neural network;
RGB is a red, green and blue color model;
RNN is a recurrent neural network;
SVC is a scalable video coding;
SVM is a support vector machine.

NOMENCLATURE

γ is an input sensory data converted into numerical representations;
 x is a horizontal pixel coordinate in the resulting image;
 y is a vertical pixel coordinate in the resulting image;

b_j is a bias for the neuron of the j -th dense layer;
 f is an activation function;
 g is a convolution kernel that moves through the data to extract features;
 $P(i,j)$ is a value after the MaxPooling operation at position (i,j) in the original matrix;
 w_{ij} is a weight between the neuron of the i -th layer and the j -th neuron of the dense layer;
 $X(i+m, j+n)$ is a initial matrix of values that are passed to CNN for processing;
 y_i is a output of the j -th neuron on the dense layer.

INTRODUCTION

Modern machine diagnostics systems use data from numerous sensors that measure key operating parameters in real time, such as vibration, temperature, pressure, and voltage. The problem lies not only in the sheer volume of data, but also in its complexity, which makes it difficult to apply classical analysis methods. Artificial intelligence

solves these problems by detecting hidden patterns and anomalies in multidimensional data, which increases the accuracy of fault prediction [1].

The use of artificial intelligence tools in the diagnostics of machine elements allows timely detection and prevention of malfunctions and prediction of possible failures, which increases the reliability of maintenance, reduces repair costs and equipment downtime [2].

The development of tools for diagnosing the technical condition of machines should be aimed not only at detecting faulty or completely damaged components, but also at preventing possible failures. Ensuring early detection of critical changes in the condition of parts allows for timely action to prevent their failure during operation. This helps to reduce maintenance costs, increase machine efficiency and improve operational safety, which is critical in many areas where the stability of technical equipment is of key importance.

The object of study process of detecting malfunctions and states that indicate an approximation of a failure in the elements of rotating machines based on data obtained from sensors.

The subject of study application of clustering algorithms and convolutional neural networks to classify sensor data and detect near-failure states of machine elements.

The purpose of the work is the development of a method for processing data from rotating machine sensors using convolutional neural networks to accurately detect states close to failure in rotating machine elements, which will increase the efficiency of maintenance and prevent equipment failures.

1 PROBLEM STATEMENT

Let $X = \{x_1, x_2, \dots, x_n\}$ be a set of sensor data collected from rotating machine elements, where each $x_i \in \mathbb{R}^d$ represents a vector of d sensor records at a given time. The dataset X is unlabeled and consists of a set of text files x_n , where each file contains sensor data for a certain period of time and a key file $K = \{k_1, k_2, \dots, k_n\}$, where $k_i \in \{0, 1\}$ is a binary label indicating whether the rotating element is in good condition ($k_i = 0$) or faulty ($k_i = 1$). The dataset X , containing sensor records is analyzed with the assumption that there is an intermediate class between fault and nominal, which means that the received sensor data on the element state does not belong to the nominal cluster and is approaching the fault condition.

The purpose of the study is to implement a method for identifying a transitional class of component condition based on indicators from sensors that read the condition of equipment. The study focuses on the possibility of identifying this class using clustering methods, such as the k -means method, and classification using convolutional neural networks.

This method allows identifying not only nominal and faulty parts, but also those that are on the verge of failure,

which makes it possible to identify parts with an increased risk of failure in advance and organize timely maintenance.

2 REVIEW OF THE LITERATURE

Recent studies show that traditional diagnostic methods that involve manual analysis of sensor data are giving way to automated approaches that use machine learning and, in particular, deep neural networks [3], [4]. Convolutional neural networks have shown significant potential in processing complex, multidimensional data [5] from industrial sensors due to their ability to detect complex relationships between signals and identify anomalies that may indicate malfunctions.

In article [6], the authors review modern machine learning methods used to monitoring and predicting faults in glass industrial rotating machines. The focus is on the use of sensor data, which is important for accurate fault diagnosis. Both traditional methods, such as regression, decision trees, and SVMs, and more modern deep learning approaches, such as neural networks, CNNs, and RNNs, have been studied.

Deep neural networks have shown significant advantages in detecting complex patterns in sensor data, which has increased the accuracy of fault diagnosis to 93–97%. This is significantly higher than traditional methods, which have an accuracy of about 80–85%. RNNs [7] were particularly effective, demonstrating up to 90% accuracy when working with sequential data, such as vibrations and temperature measurements.

The study also showed that the use of NNs significantly reduced the processing time of large amounts of sensor data, which is especially important for systems with high performance requirements. In real-life examples, in particular when working with motors and rotary machines, the use of neural networks reduced the number of failures by 20–25% compared to traditional methods.

The authors of [8] also consider modern DL approaches for fault detection and prediction in industrial systems. The main focus is on how deep learning outperforms traditional diagnostic methods capable of working with large amounts of complex sensor data. Among the main methods discussed are CNNs used to process vibration signals and images, which facilitates early detection of faults.

The authors of [9] and [10] consider the use of machine learning methods to predict industrial equipment failures based on time series of sensor data. The researchers emphasize the importance of predicting the specific moment when equipment goes from a nominal to a faulty condition, which reduces the risk of unplanned machine downtime. The paper applies various machine learning models, such as RNN and LSTM, to capture and learn from temporal patterns that indicate the approach of a machine element failure. The results show that these models are able to effectively identify failure patterns in continuous time series, providing earlier and more accurate failure prediction.

The authors of [11] also proposed a three-stage fault prediction method for rotating equipment. It uses a combination of CNN and LSTM to detect the degradation period and fault type, and (Bi)-LSTM and SVC to predict the trend and identify specific faults. The method has been successfully tested on the IMS dataset[12].

To solve such problems, authors often use the method of converting numerical data into graphs, as this allows to use CNNs to analyze them, making it possible to achieve comparable classification results to traditional machine learning algorithms such as XGBoost.

A study [13] was also conducted on the application of image-based methods for diagnosing machine faults using data from 6DOF IMU. Three methods are proposed: converting time data into a gray image, RGB image, and RGB image with X, Y, Z axes. All methods show high accuracy in classifying different operating states. The gray image provides faster training, while RGB methods offer additional analysis capabilities. The study also examines the interpretability of models using Grad-CAM [14].

Paper [15] compares methods of converting tabular data into images for use with convolutional neural networks, showing that even a basic CNN can achieve results similar to the XGBoost algorithm optimized by traditional methods.

Recent studies substantiate the effectiveness of artificial intelligence techniques, particularly neural networks, for monitoring and diagnosing the condition of rotating machinery. CNN and LSTM networks have demonstrated robust performance in detecting faults by analyzing sensor data, highlighting their potential for early fault identification. These architectures excel in extracting complex patterns and temporal dependencies within sensor readings, enabling them to recognize subtle indicators of equipment deterioration.

However, the critical issue of accurately pointing the transition from nominal operational to failure conditions remains underexplored. Current models predominantly focus on distinguishing between nominal and faulty conditions after critical issue become evident. Addressing this gap requires more advanced predictive modeling that can identify faults well before they become critical, thus providing a buffer for preventive measures.

Future research should therefore prioritize the development of sophisticated, multi-faceted models that incorporate predictive capabilities. By effectively forecasting potential failures before they manifest, such models could significantly reduce unplanned downtimes, cut maintenance costs, and enhance the operational safety and reliability of rotating machinery systems.

3 MATERIALS AND METHODS

To solve the problem of preventing failures of rotating machines by vibration analysis using machine learning techniques, it is necessary to implement an appropriate method (Fig. 1). The process begins with obtaining and downloading the input data, namely the “zeroShot” dataset [16], which includes 1158 files, each of which con-

© Zalutskya O. O., Hladun O. V., Mazurets O. V., 2025
DOI 10.15588/1607-3274-2025-1-13

tains 93752 records with sensor readings. Then, the files contained in the dataset are converted into graph images and undergo a preprocessing stage, which includes resizing to 256x256 pixels, normalizing pixel values by dividing by 255, and eliminating noise using a median filter. The cluster features are also prepared for further analysis by converting the cluster labels to the one-hot encoding format [17] to ensure correct input into the model.

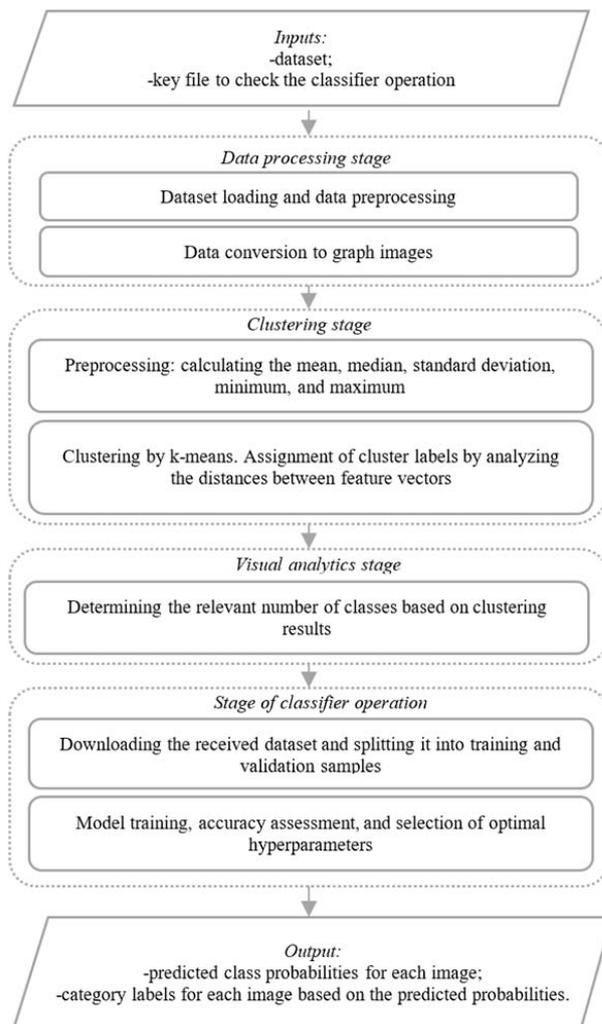


Figure 1 – A method of preventing failures of rotating machines by vibration analysis using machine learning techniques

For clustering, several important steps were taken to prepare the data and apply machine learning methods. First, basic statistical characteristics were calculated for each data file, including the mean, median, standard deviation, minimum and maximum values. These characteristics were used as features for further clustering. To ensure the uniformity of the scale of the features, the data was standardized using the StandardScaler method [18].

To determine the optimal number of clusters, it is necessary to apply the Elbow method, which helps to find the point where a further increase in the number of clusters

does not significantly reduce the inertia, the sum of squared distances to the centroids.

To solve the clustering problem, the k-means [19] and Elbow Method [20] methods were used. The k-means algorithm is used to divide data into clusters, where each cluster is characterized by the average value of the coordinates of all points in the cluster. The basic principle of k-means is as follows [21]:

- 1) first, k cluster centers are randomly selected;
- 2) each data point is assigned to the closest cluster center;
- 3) after that, the cluster centers are recalculated based on the average value of all points belonging to each cluster;
- 4) the process is repeated until the centers stop changing, i.e., the algorithm converges.

The Elbow Method is used to determine the optimal number of clusters in the clustering process. The principle of this method is to analyze the inertia – the sum of distances between points and their respective centroids – for different values of the number of clusters. The inertia graph usually shows a breaking point or “elbow” after which further increase in the number of clusters does not lead to a significant decrease in inertia [22]. This point is interpreted as the optimal number of clusters, which ensures a balance between the compactness of the clusters and their number.

To evaluate the quality of clustering, metrics such as the average value of the silhouette score, inertia, Davies-Bouldin Index, adjusted rand Index, and normalized mutual information were calculated.

Fig. 2 shows the clustering process as part of a method to prevent failures of rotating machines based on vibration analysis by using machine learning techniques.

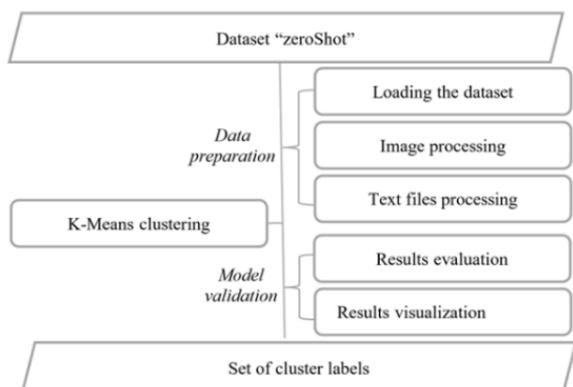


Figure 2 – The process of clustering, one of the stages of the method

The visualization of the clustering results was presented using two-dimensional scatter plot, silhouette plot, and parallel coordinates to reveal the internal structure of the clusters.

The obtained clustering results are stored for further use in the classification process. After that, the class labels preparation stage is performed, where the initial class labels are loaded from a file and converted to a one-hot

encoding format for use in the classification model. Next, the obtained dataset is divided into training and test samples in the ratio of 80% to 20%. After the data is divided, the classifier is trained on the training set.

The final step is to predict classes for the test images. The obtained results are stored with the model for further use, which allows to reproduce the classification or clustering of new images.

To effectively classify the states of rotating machines based on vibration analysis, it is necessary to develop a neural network classifier. Fig. 3 shows a neural network architecture that receives an image as an input, represented as a multidimensional tensor, where each dimension corresponds to the width, height, and number of channels.

The neural network architecture for binary classification consists of three Convolutional Convolution2D layers and MaxPooling2D sub-sampling layers that help to extract local features, reduce dimensionality, and reduce computational costs. The Flatten layer transforms the data into a flat structure, after which two fully connected Dense layers provide the final classification with an output that determines the probability of belonging to one of the classes.

The neural network architecture for three-class classification is built in a similar way, but includes an additional input layer, InputLayer, which is combined with a multidimensional tensor by concatenation to take into account additional characteristics, allowing the model to recognize more classes. In the course of building the architecture and training the neural networks, we used the pillow and OpenCV libraries for basic image processing and computer vision, as well as TensorFlow with Keras for deep learning, which allows to build and train neural networks.

The first layer of the model is the Conv2D convolutional layer, which applies several filters to detect local features such as edges and textures. Convolution is a basic operation in CNNs that is used to extract features from input data, such as sensor values [23]. The convolution for two-dimensional data is represented by formula (1) [23]:

$$(\gamma \cdot g)(x, y) = \sum_i \sum_j f(i, j) \cdot g(\gamma - i, y - j). \quad (1)$$

The next layer of the neural network is MaxPooling2D, which reduces the spatial dimensions of the tensor by extracting the most important features, which improves efficiency and reduces the number of parameters to calculate. This operation selects the maximum value in each array and is calculated using the formula [23] (2):

$$P(i, j) = \max_{0 \leq m < p, 0 \leq n < p} (X(i + m, j + n)). \quad (2)$$

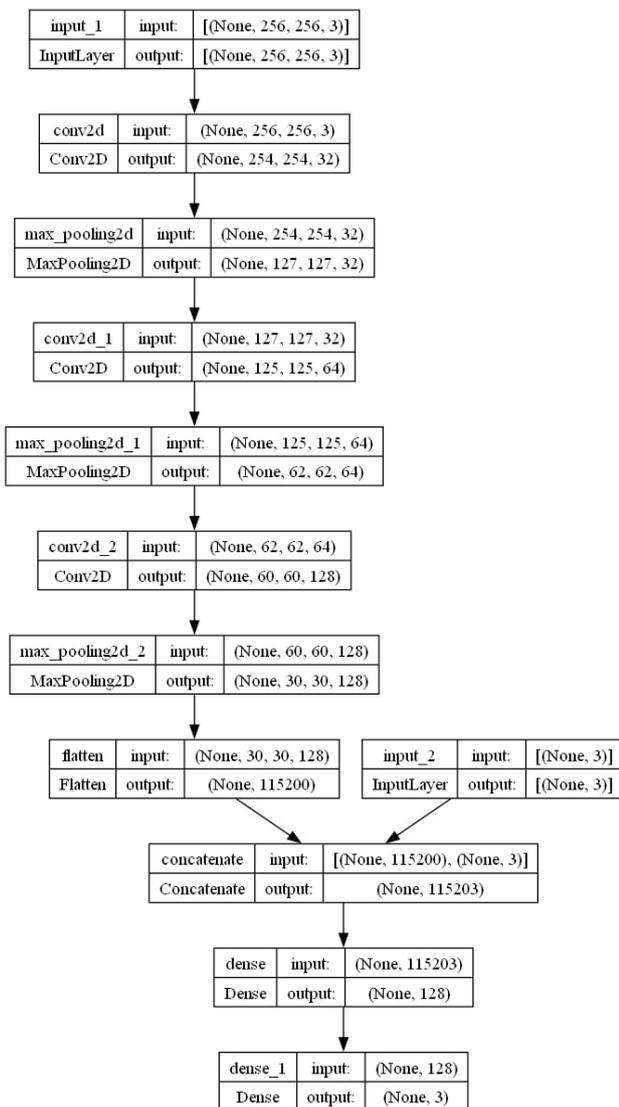


Figure 3 – Architecture of the developed convolutional neural network

After MaxPooling2D, a convolutional layer is implemented to enhance the feature analysis and help detect more complex structures. The pooling layer is again applied to reduce the size. The third convolutional layer continues to highlight complex image features, which prepares the data for the transition to the final layers. One more pooling layer completes this process and resizes the original tensor so that the data can be transferred to dense layers. The Flatten layer transforms the multidimensional tensor into a flat vector, which is necessary for further use in the fully connected Dense layers. A dense layer is a fully connected layer of a neural network, where each neuron from one layer is connected to all the neurons of the next layer. The first dense layer performs nonlinear transformations and combines the extracted features for better classification or prediction. The last dense layer completes the network, giving the final result, the probability of an image belonging to a certain class. The op-

eration of this layer is based on matrix multiplication, which is performed according to formula (3) [24]:

$$y_i = f\left(\sum_{i=1}^n w_{ij}x_i + b_i\right). \quad (3)$$

The method of preventing failures of rotating machines based on vibration analysis by using machine learning techniques was proposed. The basis of the proposed method is a combination of classical machine learning methods, such as *K*-means and the elbow method for clustering, and a CNN neural network model for classifying the conditions of rotating machines.

4 EXPERIMENTS

According to the described method, a software package was implemented, including two machine learning models for clustering into two and three classes, two neural network models for classification into two and three classes, and a website for the practical use of the obtained models. The scheme of the system's program modules is shown in Fig. 4.

The dataset used for the study consists of sensor data collected from various mechanical rotating systems covering a wide range of operational parameters. These parameters, presented in text form, were converted into numerical values suitable for analysis.

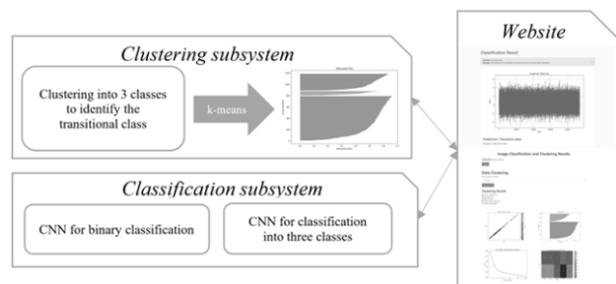


Figure 4 – Scheme of operation of the system program modules

The “zeroShot” dataset consists of 1158 files, each of which contains 93752 sensor records, as well as a file with labels for evaluating the model's performance

For clustering, no additional graphical interface was created: all processing and clustering results were displayed in the console, and the matplotlib and seaborn libraries were used for graphical representation of the data. The following libraries were used: numpy for working with multidimensional arrays, which optimizes the computations and matrix operations required to process large data sets; scikit-learn for access to machine learning algorithms, including clustering; matplotlib and seaborn for visualizing the results.

To perform the “Classifier operation” stage, the CNN neural network models were trained with the parameters shown in Table 1.

Table 1 – A set of hyperparameters for classification

Hyperparameter name	2 classes classification		3 classes classification	
Model's version	V1	V2	V3	V4
<i>Optimization hyperparameters</i>				
optimizer	adam			
loss	binary_crossentropy		categorical_crossentropy	
<i>Training hyperparameters</i>				
metrics	accuracy			
batch_size	32		32	
epochs	6	3	10	5

The architecture of the neural network for binary classification includes three layers Conv2D and MaxPooling2D for feature extraction and dimensionality reduction, Flatten for conversion to a flat structure, and two layers Dense for classification. The architecture for three-class classification is similar, but an InputLayer with concatenation is added to take into account additional characteristics. Pillow and OpenCV were used for image processing, and TensorFlow with Keras was used to build and train neural networks.

A website based on the Flask framework [25] was also implemented to interact with AI models and display analysis results. An example of the website is shown in Fig. 5. The program codes implemented in the study were uploaded to the GitHub cloud platform to ensure accessibility and the possibility of their further analysis, verification, and reuse. [26].

The defined stages of the experiment and the methods used for data preparation and processing provide a comprehensive analysis and classification of the data. Further, the results of clustering and classification will be evaluated in terms of their accuracy, reliability, and ability to identify transient states.

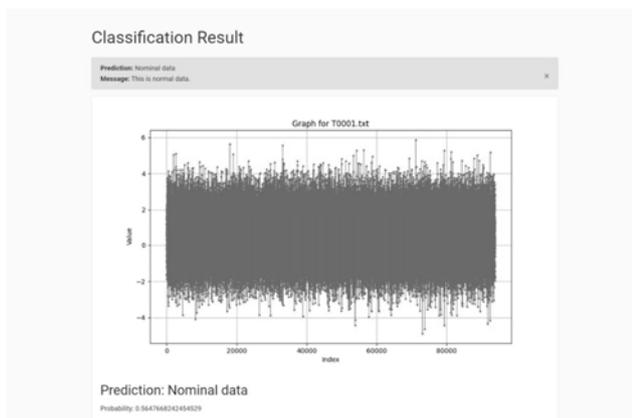


Figure 5 – Web application interface

These results will form the basis for building a prognostic model and identifying opportunities to improve the machine maintenance process, which has the potential to increase equipment stability and safety.

5 RESULTS

According to the outlined plan of the experiment, the following clustering results were obtained (Fig. 6), the dis-

tribution curve of the optimal number of Elbow method classes by the inertia parameter is presented, which decreases sharply when moving from 1 to 3 clusters, indicating that the optimal number of clusters is in the range of 2–3. This result shows that dividing the data into 3 clusters can be reasonable and will provide the best balance between clustering accuracy and model complexity.

The obtained indicators demonstrated the high accuracy of the model. For the nominal class, the model achieved precision of 0.9987, recall of 0.9885, and F1-score of 0.9936. The fault class showed precision of 0.9704, recall of 0.9966, and F1-score of 0.9833. For the transitional class, the precision is 0.9942, recall is 0.9896, and F1-score is 0.9958. The overall accuracy of the model was 0.9914. The macro- and weighted average F1-score are 0.9923 and 0.9914, respectively, which indicates high classification performance.

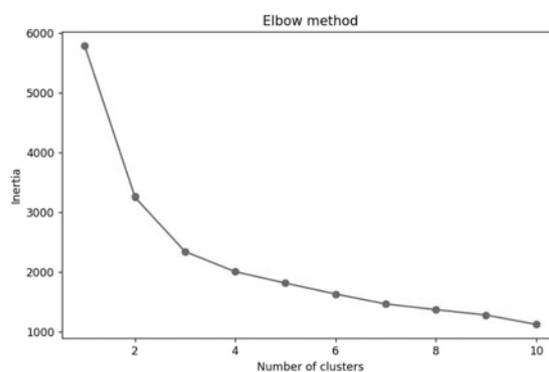


Figure 6 – Obtained clustering results

Also, according to the experimental conditions, a neural network multiclass classifier was created to distinguish three states of rotating machine elements: nominal, fault, and transitional. The confusion matrix shown in Fig. 7.

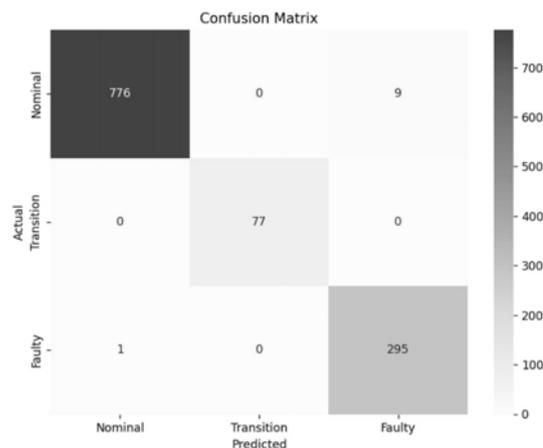


Figure 7 – Confusion matrix for multiclass model

Table 2 shows the results for binary and multiclass classification. The main focus is on the multiclass model v4, which demonstrates better classification results due to the improvement of its architecture.

Table 2 – The results of classification by different models

	precision	recall	F1-score	support	required time, sec		precision	recall	F1-score	support	required time, sec
<i>Binary classification (v1)</i>						<i>Binary classification (v2)</i>					
Nominal data	0.9718	0.6511	0.7806	293	43	Nominal data	0.9763	0.9829	0.9796	293	91
Fault data	0.9032	0.9992	0.9485	865		Fault data	0.9942	0.9919	0.9931	865	
accuracy			0.9117	1158		accuracy			0.9896	1158	
macro avg	0.9375	0.8251	0.8646	1158		macro avg	0.9852	0.9874	0.9863	1158	
weighted avg	0.9197	0.9111	0.9048	1158		weighted avg	0.9897	0.9896	0.9896	1158	
<i>Multi-class classification (v3)</i>						<i>Multi-class classification (v4)</i>					
Nominal data	0.9718	0.6092	0.7493	785	102	Nominal data	0.9987	0.9885	0.9936	785	206
Fault data	0.1901	0.8192	0.3084	77		Fault data	0.9704	0.9966	0.9833	296	
Transitional class	0.9145	0.9972	0.9527	296		Transitional class	0.9942	0.9896	0.9958	77	
accuracy			0.7217	1158		accuracy			0.9914	1158	
macro avg	0.6921	0.8085	0.6701	1158		macro avg	0.9897	0.9951	0.9923	1158	
weighted avg	0.9046	0.7228	0.7714	1158		weighted avg	0.9916	0.9914	0.9914	1158	

Based on the table, we can see significant improvements in the results between the v3 and v4 versions of the models for the multi-class classification task. In particular, for each of the three classes (“Nominal”, “Fault”, “Transitional”), the v4 version has higher precision, recall, and f1-score values compared to v3. This indicates the improved ability of the v4 model to correctly identify

each class, especially for “Fault” and “Transitional”. The overall precision score is also significantly higher in v4 (0.9914 vs. 0.7217 in v3), which demonstrates an overall improvement in model performance for all classes.

In the silhouette plots of Fig. 8a and Fig. 8b show the clustering quality assessment for two and three clusters.

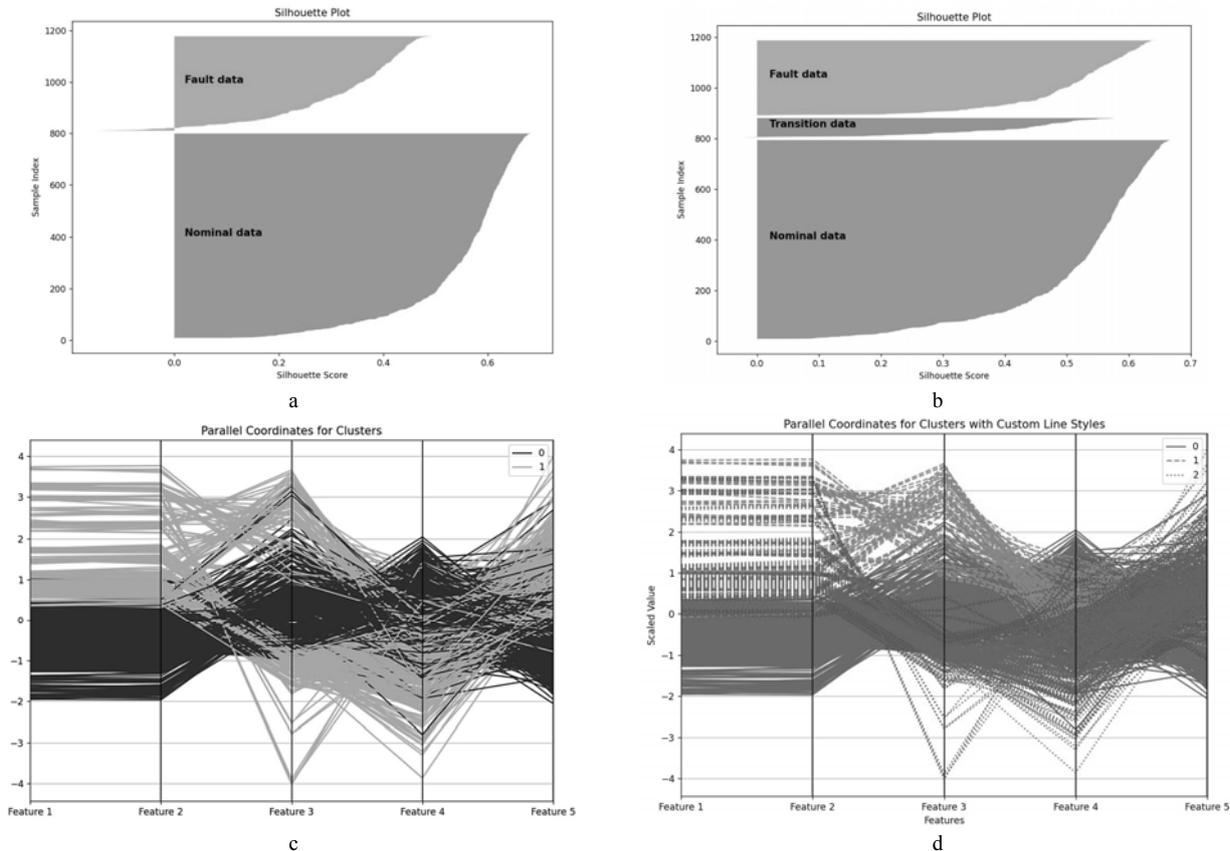


Figure 8 – Visualization of clustering results:

a – evaluation of the quality of clustering into two classes using the silhouette method; b – evaluation of the quality of clustering into three classes using the silhouette method; c – visualization of cluster division through parallel coordinates of features into two clusters; d – visualization of cluster division through parallel coordinates of features into three clusters

For the two clusters (Fig. 8a), both clusters have positive silhouette coefficient values, indicating a clear separation between them, but the overall average silhouette score is limited, which may indicate a lack of detail in the internal structure of the data. For the three clusters (Fig. 8b), the silhouette score is also positive for each cluster, and the overall average silhouette score is higher, indicating a better distribution of the data. Fig. 9 shows a graphical representation of the clustering by two classes.

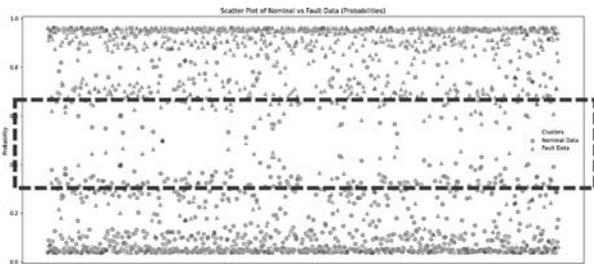


Figure 9 – Graphical representation of clustering by two classes

The above experiments were performed on the following hardware: Intel core i5-11400H processor, RAM: 16 GB. Software: Windows 11 Home, Python 3.9.0 programming language, JavaScript – ECMAScript 2023 (ES14), Visual Studio Code editor, TensorFlow/Keras, Scikit-learn, OpenCV, Matplotlib, Seaborn, Numpy, Pillow, Flask libraries were used.

6 DISCUSSION

According to the Elbow Method graph (Fig. 7), which shows the correlation between the number of clusters and the inertia value, the indicators decrease rapidly when moving from 1 to 3 clusters, after which the decrease becomes less noticeable. Such a sharp decline in the first steps usually indicates a zone of optimal distribution, where the addition of new clusters slightly improves the distribution but significantly increases the complexity of the model.

This result indicates that the optimal number of clusters may be in the range of 2–3, since further increasing the number of clusters does not significantly improve the inertia. The choice of 3 clusters is justified because this division provides an effective reduction of internal variance, which is important for more accurate modeling of the data structure, without excessive model complexity and excessive division of data into small groups.

According to Fig. 10, the third class is proposed to be the transitional class, which indicates a condition of a machine element that does not belong to either the nominal or fault classes. Based on the graphical analysis, it can be concluded that at a uniform distance from both main clusters, the model demonstrates a reduced ability to accurately determine the belonging to a particular cluster. This is due to the fact that the central points are equidistant from both clusters, combining the features of each of them. In this regard, it is advisable to expand the clustering model to three classes. Thus, in addition to the main

classes (nominal and fault), a third transition class is added, which accumulates the common characteristics of both clusters, allowing the model to more accurately reflect the data structure.

Thus, the division into three clusters not only allows for a better representation of the data structure, but also provides a more reasonable representation of the internal features of the dataset. According to the visualizations of parallel coordinates for two- and three-class clustering, clustering into three classes is appropriate. Adding a third cluster (labeled #1 in the graph) in Fig. 9d reveals additional structural differences, especially on the features “Feature 3” and “Feature 4”, which indicates the presence of unique features that were not visible in the two-class clustering. At the same time, there is an overlap of lines on certain features, which may indicate imperfect separation of the clusters. Nevertheless, the third cluster helps to segment the data more accurately, taking into account less pronounced differences.

Therefore, the division into 3 clusters supports a balanced decision between adequate clustering accuracy and overall model performance. This result indicates that the optimal number of clusters may be in the range of 2–3, since further increasing the number of clusters does not bring significant improvement in inertia. The choice of 3 clusters is justified because this division provides an effective reduction of internal variance, which is important for more accurate modeling of the data structure, without excessive model complexity and excessive breakdown of data into small groups. Thus, the division into 3 clusters maintains a balanced decision between appropriate clustering accuracy and overall model efficiency.

The clustering results show that the three-level classification is optimal, providing a clear separation and display of the data structure. Analysis using the elbow method and the silhouette coefficient confirmed the effectiveness of this method. In addition, a neural network model was created to recognize the third, transitional class, which increased the overall accuracy and flexibility of the model, providing better consideration of complex variations in the data when analyzing the condition of rotating machine elements.

CONCLUSIONS

The method of preventing failures of rotating machines based on vibration analysis by using machine learning techniques was implemented. The main feature of the proposed method is the ability not only to classify conditions as normal or faulty, but also to identify an intermediate condition characterized by an increased probability of element failure. The use of clustering has made it possible to achieve an accuracy of over 80% in identifying this third class, which makes it possible to predict probable failures at early stages.

In addition, the method is based on the use of a convolutional neural network that has been trained with an accuracy of 99% to classify states, including a new “transitional” class. This ensures high efficiency and reliability of the classification, which allows not only to increase the

safety and reliability of equipment operation, but also to minimize the costs associated with emergency conditions and repairs.

The scientific novelty of obtained results is that the method of preventing failures of rotating machines based on vibration analysis by using machine learning techniques is firstly proposed. In the resulting dataset, it is proposed to distinguish a new, third class of transitional, which indicates a transient, close-to-failure state of a rotating element, and the effectiveness of introducing this class is proved.

The practical significance of the obtained results is the creation of application software, specifically neural network models for classifying the condition of rotating machine elements and the implementation of a corresponding web application for interacting with the obtained models.

Prospects for further research may be aimed at optimizing algorithms and neural network architecture to reduce training time, which will improve the efficiency of models with large amounts of data and different classes. Furthermore, it is also possible to develop methods for adaptive training of models based on new data, which will increase the efficiency of their application in real-world conditions.

ACKNOWLEDGEMENTS

This research was conducted as part of the All-Ukrainian Competition-Hackathon of Scientific Papers by Young Scientists in the Field of Intelligent Information Technologies in 2024.

REFERENCES

- Havugimana I., Liu B., Liu F., Zhang J., Li B., Wan P. Review of artificial intelligent algorithms for engine performance, control, and diagnosis, *Energies*, 2023, Vol. 16, No 3, P. 1206. DOI: 10.3390/en16031206.
- Du C., Li W., Rong Y., Li F., Yu F., Zeng X. Research on the application of artificial intelligence method in automobile engine fault diagnosis, *Engineering Research Express*, 2021, Vol. 3, No 2, p. 026002. DOI: 10.1088/2631-8695/ac01ad.
- Gong C.-S. A., Su C.-H. S., Chen Y.-H., Guu D.-Y. How to Implement Automotive Fault Diagnosis Using Artificial Intelligence Scheme, *Micromachines*, 2022, Vol. 13, P. 1380. DOI: 10.3390/mi13091380.
- Matania O., Bachar L., Bechhoefer E., Bortman J. Signal processing for the condition-based maintenance of rotating machines via vibration analysis: A tutorial, *Sensors*, 2024, Vol. 24, No 2, P. 454. DOI: 10.3390/s2402045.
- Krak I., Zalutskaya O., Molchanova M., Mazurets O., Manziuk E., Barmak O. Method for neural network detecting propaganda techniques by markers with visual analytic, *CEUR Workshop Proceedings*, 2024, Vol. 3790, pp. 158–170. Mode of access: <https://ceur-ws.org/Vol-3790/paper14.pdf>.
- Singh V., Gangsar P., Porwal R., Atulkar A. Artificial intelligence application in fault diagnostics of rotating industrial machines: a state-of-the-art review, *Journal of Intelligent Manufacturing*, 2021. DOI: 10.1007/s10845-021-01861-5.
- Krak I., Zalutskaya O., Molchanova M., Mazurets O., Bahrii R., Sobko O., Barmak O. Abusive Speech Detection Method for Ukrainian Language Used Recurrent Neural Network, *CEUR Workshop Proceedings*, 2023, Vol. 3387, pp. 16–28. DOI: 10.31110/COLINS/2024-3/002.
- Qiu S., Cui X., Ping Z., Shan N., Li Z., Bao X., Xu X. Deep learning techniques in intelligent fault diagnosis and prognosis for industrial systems: a review, *Sensors*, 2023, Vol. 23, No 3, P. 1305. DOI: 10.3390/s23031305.
- Vago N. O. P., Forbicini F., Fraternali P. Predicting machine failures from multivariate time series: an industrial case study, *Machines*, 2024, Vol. 12, No 6, P. 357. DOI: 10.3390/machines12060357.
- Lee Y., Park C., Kim N., Ahn J., Jeong J. LSTM-autoencoder based anomaly detection using vibration data of wind turbines, *Sensors*, 2024, Vol. 24, No 9, P. 2833. DOI: 10.3390/s24092833.
- Peng H., Li H., Zhang Y., Wang S., Gu K., Ren M. Multi-sensor vibration signal based three-stage fault prediction for rotating mechanical equipment, *Entropy*, 2022, Vol. 24, No 2, P. 164. DOI: 10.3390/e24020164.
- IMS Bearings. Mode of access: https://data.nasa.gov/Raw-Data/IMS-Bearings/brfb-gzcv/about_data.
- Luczak D. Machine fault diagnosis through vibration analysis: time series conversion to grayscale and RGB images for recognition via convolutional neural networks, *Energies*, 2024, Vol. 17, No 9, P. 1998. DOI: 10.3390/en17091998.
- Selvaraju R. R., Cogswell M., Das A., Vedantam R., Parikh D., Batra D. Grad-CAM: visual explanations from deep networks via gradient-based localization, *International Journal of Computer Vision*, 2019, Vol. 128, No 2, pp. 336–359. DOI: 10.1007/s11263-019-01228-7.
- Medeiros N. L., Silva N. S. R., Endo P. T. A comparative analysis of converters of tabular data into image for the classification of arboviruses using convolutional neural networks, *PLOS ONE*, 2023, Vol. 18, No 12, P. e0295598. DOI: 10.1371/journal.pone.0295598.
- Department of Software of the National University – Zaporizhzhya Polytechnic. Mode of access: https://pz.zp.ua/files/vkiit/zeroShot_ykiit.rar.
- Yu L., Zhou R., Chen R., Lai K. K. Missing data preprocessing in credit classification: one-hot encoding or imputation?, *Emerging Markets Finance and Trade*, 2020, pp. 1–11. DOI: 10.1080/1540496x.2020.1825935.
- Setiawan H., Firmada A., Khair U. Enhancing the accuracy of diabetes prediction using feedforward neural networks: strategies for improved recall and generalization, *Brilliance: Research of Artificial Intelligence*, 2024, Vol. 4, No 1, pp. 201–207. DOI: 10.47709/brilliance.v4i1.3888.
- Miraftebzadeh S. M., et al. K-means and alternative clustering methods in modern power systems, *IEEE Access*, 2023, P. 1. DOI: 10.1109/access.2023.3327640.
- Sutomo F., Muaafii D. A., Al Rasyid D. N., Kurniawan Y. I., Afuan L., Cahyono T., et al. Optimization of the k-nearest neighbors algorithm using the elbow method on stroke prediction, *Jurnal Teknik Informatika (Jutif)*, 2023, Vol. 4, No 1, pp. 125–130. DOI: 10.52436/1.jutif.2023.4.1.839.
- Ashari I. F., Nugroho E. D., Baraku R., Yanda I. N., Liwardana R. Analysis of elbow, silhouette, Davies-Bouldin, Calinski-Harabasz, and Rand-Index evaluation on k-means algorithm for classifying flood-affected areas in Jakarta, *Journal of Applied Informatics and Computing*, 2023, Vol. 7, No. 1, pp. 89–97. DOI: 10.30871/jaic.v7i1.4947.

22. Humaira H., Hestry R., Rasyidah R. Determining the appropriate cluster number using elbow method for k-means algorithm, *Proceedings of the 2nd Workshop on Multidisciplinary and Applications (WMA)*, Padang, Indonesia, Jan. 24–25, 2018. DOI: 10.4108/eai.24-1-2018.2292388.
 23. Pang Y., Sun M., Jiang X., Li X. Convolution in convolution for network in network, *IEEE Transactions on Neural Networks and Learning Systems*, 2017, Vol. 29, No. 5, pp. 1587–1597. DOI: 10.1109/TNNLS.2017.2676130
 24. Zafar A., Aamir M., Mohd Nawi N., Arshad A., Riaz S., Alruban A., et al. A comparison of pooling methods for convolutional neural networks, *Applied Sciences*, 2022, Vol. 12, No. 17, p. 8643. DOI: 10.3390/app12178643
 25. Welcome to Flask – Flask Documentation (3.1.x). Mode of access: <https://flask.palletsprojects.com/en/stable>.
 26. GitHub – AlexFRUZ/Corrupted_File_Detection. Mode of access: https://github.com/AlexFRUZ/Corrupted_File_Detection.
- Received 18.12.2024.
Accepted 23.01.2024.

УДК 004.8

МЕТОД ПОПЕРЕДЖЕННЯ АВАРІЙНИХ СТАНІВ ОБЕРТОВИХ МАШИН ЗА АНАЛІЗОМ ВІБРАЦІЙ ЗАСОБАМИ МАШИННОГО НАВЧАННЯ

Залуцька О. О. – асистент кафедри комп'ютерних наук Хмельницького національного університету, Хмельницький, Україна.

Гладун О. В. – студент бакалаврського рівня освіти кафедри комп'ютерних наук Хмельницького національного університету, Хмельницький, Україна.

Мазурець О. В. – канд. техн. наук, доцент кафедри комп'ютерних наук Хмельницького національного університету, Хмельницький, Україна.

АНОТАЦІЯ

Актуальність. Вирішується проблема визначення перехідних станів, що передують переходу з робочого стану у неробочий за отриманими даними з датчиків обертових елементів машин. Об'єктом дослідження є процес виявлення несправностей та станів, що свідчать про наближення до поломки у елементах обертових машин на основі даних, отриманих з сенсорів. Предметом дослідження є застосування алгоритмів *k-means* та методу Elbow для кластеризації та згорткових нейронних мереж для класифікації даних з сенсорів та виявлення близьких до поломки станів елементів машини.

Мета роботи. Метою роботи є створення методу обробки сенсорних даних обертових машин з використанням згорткових нейронних мереж для точного виявлення станів, близьких до відмови, в елементах обертових машин, що дозволить підвищити ефективність технічного обслуговування та запобігти відмовам обладнання.

Метод. Запропонований метод запобігання відмовам обертових машин базується на аналізі сигналів вібрації з використанням комбінації методів кластеризації та глибокого навчання. На першому етапі дані з датчиків проходять попередню обробку, що включає нормалізацію, зменшення розмірності та видалення шумів, після чого застосовується алгоритм *k-середніх*. Для визначення оптимальної кількості кластерів використовується метод Elbow, який забезпечує ефективне групування станів обертових елементів машини, виявляючи стани, близькі до переходу в несправність. Також була розроблена модель CNN, яка класифікує кластери, дозволяючи точно розділити номінальні, несправні та перехідні стани. Поєднання методів кластеризації з CNN-моделлю підвищує точність виявлення потенційних несправностей і дозволяє своєчасно реагувати на них, що є критично важливим для запобігання аваріям і забезпечення стабільності роботи обладнання.

Результати. Створено метод попередження аварійних станів обертових машин за аналізом вібрацій засобами машинного навчання та відповідний комплекс програмного забезпечення. Реалізований метод дозволяє ідентифікувати не лише нормальні й аварійні стани, але й виділяти третій клас – близький до поломки. Якість кластеризації для трьох класів підтверджується значенням коефіцієнта силуету 0,506, що свідчить про належну відокремленість кластерів, та індексом Девіса-Болдіна 0,796, що демонструє високий рівень внутрішньої когерентності кластерів. Додатково було натреновано CNN, яка досягає 99% точності для класифікації цього класу, що робить метод високоефективним і вирізняє його серед існуючих рішень.

Висновки. Було розроблено метод попередження аварійних станів обертових машин за аналізом вібрацій засобами машинного навчання, запропоновано виокремлення третього класу – перехідного, що вказує на стан, близький до поломки, і підтверджено його ефективність. Практичне значення результатів полягає у створенні нейромережових моделей для класифікації стану обертових елементів та розробці вебзастосунку для взаємодії з цими моделями.

КЛЮЧОВІ СЛОВА: обертові елементи машини, відмова елемента, перехідні стани, кластеризація, класифікація, CNN.

ЛІТЕРАТУРА

1. Review of Artificial Intelligent Algorithms for Engine Performance, Control, and Diagnosis / [I. Havugimana, B. Liu, F. Liu et al.] // *Energies*. – 2023. – Vol. 16, No. 3. – P. 1206. DOI: 10.3390/en16031206
2. Research on the application of artificial intelligence method in automobile engine fault diagnosis / [C. Du, W. Li, Y. Rong et al.] // *Engineering Research Express*. – 2021. – Vol. 3, no. 2. – P. 026002. DOI: 10.1088/2631-8695/ac01ad
3. How to implement automotive fault diagnosis using artificial intelligence scheme / [C.-S. Gong, C.-H. S. Su, Y.-H. Chen, D.-Y. Guu] // *Micromachines*. – 2022. – Vol. 13. – P. 1380. DOI: 10.3390/mi13091380.
4. Signal Processing for the Condition-Based Maintenance of Rotating Machines via Vibration Analysis: A Tutorial / [O. Matania, L. Bachar, E. Bechhofer, J. Bortman] // *Sensors*. – 2024. – Vol. 24, No. 2. – P. 454. DOI: 10.3390/s24020454.
5. Method for neural network detecting propaganda techniques by markers with visual analytic / [I. Krak, O. Zalutskaya, M. Molchanova et al.] // *CEUR Workshop Proceedings*. –

2024. – Vol. 3790. – P. 158–170. Mode of access: <https://ceur-ws.org/Vol-3790/paper14.pdf>.
6. Artificial Intelligence Application in Fault Diagnostics of Rotating Industrial Machines: A State-of-the-Art Review / [V. Singh, P. Gangsar, R. Porwal, A. Atulkar] // *Journal of Intelligent Manufacturing*. – 2021. DOI: 10.1007/s10845-021-01861-5.
 7. Abusive Speech Detection Method for Ukrainian Language Used Recurrent Neural Network / [I. Krak, O. Zalutska, M. Molchanova et al.] // *COLINS*. – 2024. – Vol. 3. – P. 16–28.
 8. Deep Learning Techniques in Intelligent Fault Diagnosis and Prognosis for Industrial Systems: A Review / [S. Qiu, X. Cui, Z. Ping et al.] // *Sensors*. – 2023. – Vol. 23, No. 3. – P. 1305. DOI: 10.3390/s23031305.
 9. Pinciroli Vago N. O. Predicting Machine Failures from Multivariate Time Series: An Industrial Case Study / Nicolò Oreste Pinciroli Vago, Francesca Forbicini, Piero Fraternali // *Machines*. – 2024. – Vol. 12, no. 6. – P. 357. DOI: 10.3390/machines12060357
 10. LSTM-Autoencoder Based Anomaly Detection Using Vibration Data of Wind Turbines / [Y. Lee, C. Park, N. Kim et al.] // *Sensors*. – 2024. – Vol. 24, No. 9. – P. 2833. DOI: 10.3390/s24092833.
 11. Multi-Sensor Vibration Signal Based Three-Stage Fault Prediction for Rotating Mechanical Equipment / [H. Peng, H. Li, Y. Zhang et al.] // *Entropy*. – 2022. – Vol. 24, No. 2. – P. 164. DOI: 10.3390/e24020164.
 12. IMS Bearings // NASA Open Data Portal. – Mode of access: https://data.nasa.gov/Raw-Data/IMS-Bearings/brfb-gzcv/about_data
 13. Łuczak D. Machine Fault Diagnosis through Vibration Analysis: Time Series Conversion to Grayscale and RGB Images for Recognition via Convolutional Neural Networks / D. Łuczak // *Energies*. – 2024. – Vol. 17, No. 9. – P. 1998. DOI: 10.3390/en17091998
 14. Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization / [R. R. Selvaraju, M. Cogswell, A. Das et al.] // *International Journal of Computer Vision*. – 2019. – Vol. 128, no. 2. – P. 336–359. DOI: 10.1007/s11263-019-01228-7.
 15. Medeiros Neto L. A comparative analysis of converters of tabular data into image for the classification of Arboviruses using Convolutional Neural Networks / L. Medeiros Neto, S. R. da Silva Neto, P. T. Endo // *PLOS ONE*. – 2023. – Vol. 18, No. 12. – P. e0295598. DOI: 10.1371/journal.pone.0295598.
 16. Department of Software of the National University – Zaporizhzhya Polytechnic // Downloading the dataset “zeroShot” – Mode of access: https://pz.zp.ua/files/vkiit/zeroShot_vkiit.rar.
 17. Missing Data Preprocessing in Credit Classification: One-Hot Encoding or Imputation? / [L. Yu, R. Zhou, R. Chen, K. K. Lai] // *Emerging Markets Finance and Trade*. – 2020. – P. 1–11. DOI: 10.1080/1540496x.2020.1825935.
 18. Setiawan H. Enhancing the Accuracy of Diabetes Prediction Using Feedforward Neural Networks: Strategies for Improved Recall and Generalization / H. Setiawan, A. Firnanda, U. Khair // *Brilliance: Research of Artificial Intelligence*. – 2024. – Vol. 4, No. 1. – P. 201–207. DOI: 10.47709/brilliance.v4i1.3888.
 19. K-means and Alternative Clustering Methods in Modern Power Systems / [S. M. Miraftabzadeh et al.] // *IEEE Access*. – 2023. – P. 1. DOI: 10.1109/access.2023.3327640.
 20. Optimization of the k-nearest neighbors algorithm using the elbow method on stroke prediction / [F. Sutomo, D. A. Muaafii, D. N. Al Rasyid et al.] // *Jurnal Teknik Informatika (Jutif)*. – 2023. – Vol. 4, No. 1. – P. 125–130. – DOI: 10.52436/1.jutif.2023.4.1.839.
 21. Analysis of Elbow, Silhouette, Davies-Bouldin, Calinski-Harabasz, and Rand-Index Evaluation on K-Means Algorithm for Classifying Flood-Affected Areas in Jakarta / [I. F. Ashari, E. D. Nugroho, R. Baraku et al.] // *Journal of Applied Informatics and Computing*. – 2023. – Vol. 7, No. 1. – P. 89–97. DOI: 10.30871/jaic.v7i1.4947.
 22. Humaira H. Determining the Appropriate Cluster Number Using Elbow Method for K-Means Algorithm / [H. Humaira, R. Hestry, R. Rasyidah] // *Proceedings of the 2nd Workshop on Multidisciplinary and Applications (WMA) 2018, 24–25 January 2018, Padang, Indonesia*. – 2020. DOI: 10.4108/eai.24-1-2018.2292388.
 23. Convolution in convolution for network in network / [Y. Pang, M. Sun, X. Jiang, X. Li] // *IEEE Transactions on Neural Networks and Learning Systems*. – 2017. – Vol. 29, No. 5. – P. 1587–1597. DOI: 10.1109/TNNLS.2017.2676130
 24. A comparison of pooling methods for convolutional neural networks / [A. Zafar, M. Aamir, N. Mohd Nawi et al.] // *Applied Sciences*. – 2022. – Vol. 12, No. 17. – P. 8643. DOI: 10.3390/app12178643
 25. Welcome to Flask – Flask Documentation (3.1.x) // Welcome to Flask – Flask Documentation (3.1.x). – Mode of access: <https://flask.palletsprojects.com/en/stable>.
 26. GitHub – AlexFRUZ/Corrupted_File_Detection // GitHub. – Mode of access: https://github.com/AlexFRUZ/Corrupted_File_Detection.

DEEPAKE AUDIO DETECTION USING YOLOV8 WITH MEL-SPECTROGRAM ANALYSIS: A CROSS-DATASET EVALUATION

Zbezhkhovska U. R. – PhD, Leading Researcher of Scientific and Methodical Department for Quality Assurance of Educational Activities and Higher Education, Ivan Kozhedub Kharkiv National Air Force University, Kharkiv, Ukraine.

ABSTRACT

Context. The problem of detecting deepfake audio has become increasingly critical with the rapid advancement of voice synthesis technologies and their potential for misuse. Traditional audio processing methods face significant challenges in distinguishing sophisticated deepfakes, particularly when tested across different types of audio manipulations and datasets. The object of study is developing a deepfake audio detection model that leverages mel-spectrograms as input to computer vision techniques, focusing on improving cross-dataset generalization capabilities.

Objective. The goal of the work is to improve the generalization capabilities of deepfake audio detection models by employing mel-spectrograms and leveraging computer vision techniques. This is achieved by adapting YOLOv8, a state-of-the-art object detection model, for audio analysis and investigating the effectiveness of different mel-spectrogram representations across diverse datasets.

Method. A novel approach is proposed using YOLOv8 for deepfake audio detection through the analysis of two types of mel-spectrograms: traditional and concatenated representations formed from SincConv filters. The method transforms audio signals into visual representations that can be processed by computer vision algorithms, enabling the detection of subtle patterns indicative of synthetic speech. The proposed approach includes several key components: BCE loss optimization for binary classification, SGD with momentum (0.937) for efficient training, and comprehensive data augmentation techniques including random flips, translations, and HSV color augmentations. The SincConv filters cover a frequency range from 0 Hz to 8000 Hz, with a step size of approximately 533.33 Hz per filter, providing detailed frequency analysis capabilities. The effectiveness is evaluated using the EER metric across multiple datasets: ASVspoof 2021 LA (25,380 genuine and 121,461 spoofed utterances) for training, and ASVspoof 2021 DF, Fake-or-Real (111,000 real and 87,000 synthetic utterances), In-the-Wild (17.2 hours fake, 20.7 hours real), and WaveFake (117,985 fake files) datasets for testing cross-dataset generalization.

Results. The experiments demonstrate varying effectiveness of different mel-spectrogram representations across datasets. Concatenated mel-spectrograms showed superior performance on diverse, real-world datasets (In-the-Wild: 34.55% EER, Fake-or-Real: 35.3% EER), while simple mel-spectrograms performed better on more homogeneous datasets (ASVspoof DF: 28.99% EER, WaveFake: 34.55% EER). Feature map visualizations reveal that the model's attention patterns differ significantly between input types, with concatenated spectrograms showing more distributed focus across relevant regions for complex datasets. The training process, conducted over 50 epochs with a learning rate of 0.01 and warm-up strategy, demonstrated stable convergence and consistent performance across multiple runs.

Conclusions. The experimental results confirm the viability of using YOLOv8 for deepfake audio detection and demonstrate that the effectiveness of mel-spectrogram representations depends significantly on dataset characteristics. The findings suggest that input representation should be selected based on the specific properties of the target audio data, with concatenated spectrograms being more suitable for diverse, real-world scenarios and simple spectrograms for more controlled, homogeneous datasets. The study provides a foundation for future research in adaptive representation selection and model optimization for deepfake audio detection.

KEYWORDS: deepfake detection, YOLOv8, mel-spectrogram, generalization capabilities.

ABBREVIATIONS

CNN is a convolutional neural network;
YOLO is a You Only Look Once model;
LA is a logical access;
DF is a deepfake;
PAN is a path aggregation network;
FPN is a feature pyramid network;
SPP is a spatial pyramid pooling;
BCE is a binary cross-entropy loss;
SGD is a stochastic gradient descent;
EER is a equal error rate metric;
STFT is a short-time Fourier transform;
FAR is a false acceptance rate;
FRR is a false rejection rate;
FP is a false positive;
TN is a true negative;
FN is a false negative;
TP is a true positive;
TTS is a text-to-speech;

VC is a vocoder;
HSV is a Hue, Saturation, and Value.

NOMENCLATURE

f is a frequency of a function;
 v_t is a velocity term accumulating past gradients for momentum;
 μ is a momentum coefficient for gradient updates;
 θ_t is a model parameter at training step t ;
 η is a learning rate parameter controlling step size;
 $\nabla_{\theta} L(\theta_t, x^{(i)}, y^{(i)})$ is a gradient of the loss function L with respect to the parameters θ , computed for a single training sample $x^{(i)}, y^{(i)}$;
 m is a mel scale value;
 y is a ground truth label (0 for real audio, 1 for deepfake);
 \hat{y} is a predicted probability of the sample being a deepfake;

X is a set of audio signals $\{x_1, x_2, \dots, x_N\}$;
 Y is a set of corresponding labels $\{y_1, y_2, \dots, y_N\}$.

INTRODUCTION

Detecting deepfake audio has become increasingly critical as the technology to create synthetic and altered speech has evolved. Deepfake audio can convincingly imitate human voices, often with the intention to deceive or manipulate, posing significant risks in areas such as security, media integrity, and public trust. Traditional audio processing methods face challenges in distinguishing deepfakes, particularly when tested across different types of audio manipulations and datasets.

One promising approach to address these challenges involves converting audio signals into visual representations, such as mel-spectrograms, which capture the sound's time-frequency features [2]. By transforming audio into images, computer vision models, which are highly effective at image recognition tasks, can be applied to detect patterns indicative of deepfake audio. CNN and other computer vision architectures can then analyze these spectrograms to detect anomalies or characteristics that differentiate genuine audio from deepfake audio. This method provides a novel and powerful approach to improve the accuracy of deepfake audio detection.

The object of study is developing a deepfake audio detection model that leverages mel-spectrograms as input to computer vision techniques.

Building such models requires significant computational resources, as training a network on large datasets of audio data is time-intensive. A major challenge in deepfake detection is ensuring that the model generalizes well, meaning it performs effectively not only on the dataset it was trained on but also on entirely new and unseen datasets. Many models perform well within their training environment but falter when encountering novel types of deepfakes, generalizing a key objective for practical deployment.

The subject of study is using mel-spectrograms in combination with computer vision models to enhance deepfake audio detection, focusing on improving the model's generalization across diverse datasets.

Current approaches to deepfake audio detection face significant challenges in generalization. Many models perform well on their training datasets but struggle when encountering new types of deepfakes or audio from different sources [2–27]. This limitation is particularly problematic given the rapid evolution of deepfake technologies. Mel-spectrograms offer a promising solution by transforming audio data into a visual format that can be analyzed using advanced computer vision techniques. These techniques have shown remarkable speed and accuracy in various image recognition tasks, but their full potential in deepfake audio detection via mel-spectrograms remains to be explored [28–29]. By investigating this approach, there is an opportunity to address the critical challenge of cross-dataset generalization in deepfake audio

detection, potentially leading to more robust and versatile detection systems.

The purpose of the work is to improve the generalization capabilities of deepfake audio detection models by employing mel-spectrograms and leveraging computer vision techniques. By training the model on one dataset and testing it on others, this study aims to develop a more robust detection system that is effective across different types of deepfake audio.

1 PROBLEM STATEMENT

Suppose we are given an audio dataset represented as a set of instances $\langle X, Y \rangle$, where $X = \{x_1, x_2, \dots, x_N\}$ is the set of audio signals, and $Y = \{y_1, y_2, \dots, y_N\}$ represents the corresponding labels, where $y_i = 1$ for real audio and $y_i = 0$ for fake audio. For each audio signal x_i , we convert it into a mel-spectrogram representation $S(x_i)$, such that the problem of deepfake detection can be transformed into an image classification problem using the mel-spectrograms.

Given this set of mel-spectrograms $\langle S(X), Y \rangle$, the task is to train a computer vision model $F(\theta, S(x_i))$, where θ represents the set of model parameters, to predict whether an audio sample is real or fake. The objective is to minimize a loss function $L(F(\theta, S(x_i)), y_i) \rightarrow \text{opt}$, where opt represents the optimal performance in terms of classification accuracy.

In addition, the problem of cross-dataset generalization is of primary interest. Specifically, for a model trained on a dataset $\langle S(X_{\text{train}}), Y_{\text{train}} \rangle$, we aim to evaluate its performance on a distinct test set $\langle S(X_{\text{test}}), Y_{\text{test}} \rangle$, where $X_{\text{test}} \neq X_{\text{train}}$ and the distribution of deepfake techniques may differ. The challenge is ensuring that the trained model generalizes well across diverse datasets, maintaining high accuracy on unseen data and addressing the limitations of dataset-specific detection methods.

2 REVIEW OF THE LITERATURE

The rapid advancement of artificial intelligence has led to the proliferation of deepfake audio, posing significant challenges to audio authenticity and security. Deepfake audio detection methods can be broadly categorized into pipeline approaches and end-to-end detectors [2–5]. Pipeline approaches involve a two-step process of feature extraction and classification. At the same time, end-to-end detectors aim to learn the detection task in a single step using deep neural networks.

Feature extraction techniques are crucial in capturing discriminative characteristics present in fake audio. These include short-term and long-term spectral features, prosodic features, and features derived from deep learning [6–8]. Short-term spectral features like Short-Time Fourier Transform effectively detect abrupt changes in audio signals [9], while prosodic features help uncover irregularities in speech pitch, intonation, and rhythm [10–11].

Recent advancements have incorporated self-supervised learning models like Wav2Vec, Wav2Vec2 XLS-R, and Hubert into the feature extraction process [12–15]. These models learn discriminative features from

raw audio without explicit labeling, potentially enhancing detection efficacy. However, a key challenge is ensuring these features generalize well across different types of deepfake attacks and audio datasets.

Traditional classifiers such as Support Vector Machines, Gaussian Mixture Models, and Logistic Regression have been employed in deepfake audio detection [16–18]. While these methods offer simplicity and efficiency, they often struggle to capture the intricate patterns introduced by sophisticated deepfake audio generation techniques, limiting their effectiveness against evolving attacks.

Deep learning approaches have shown significant promise in detecting subtle manipulations within audio data. CNN, particularly Light CNN, has performed excellently in deepfake audio classification tasks [19]. Residual Networks (ResNet) and its variants have also achieved promising results [20–21]. However, these models often require large amounts of training data and may not generalize well to unseen attack types. More advanced architectures like Res2Net [22], RawNet2 [23], and Squeeze-and-Excitation Networks [24] have been proposed to capture finer-grained audio features. Graph Neural Networks, such as RawGAT-ST, have improved performance in detecting a broad spectrum of spoofing attacks [25]. While these models offer enhanced feature learning capabilities, they often come at the cost of increased computational complexity and reduced interpretability.

A critical challenge in deepfake audio detection is the model's ability to generalize across different datasets and attack types. In [26], the authors observed that while their SincNet-based model performed well on known attacks, it struggled with attacks significantly different from those in the training set. This highlights the importance of diverse training data and robust evaluation protocols to ensure models can detect a wide range of deepfake techniques.

Transformer-based models like Rawformer have demonstrated improved performance and generalization across different datasets. The SE-Rawformer demonstrated good generalization, performing well on both ASVspoof 2019 LA and ASVspoof 2021 LA datasets [27]. However, the rapid evolution of deepfake technologies means that models must continuously adapt to new attack vectors, posing ongoing challenges for generalization.

While most research has focused on audio-specific architectures, the potential application of YOLOv8 to deepfake audio detection via mel-spectrogram transformation presents an interesting avenue for exploration. YOLOv8's efficiency in processing images could potentially translate to fast analysis of mel-spectrograms, enabling real-time deepfake audio detection. Its localization capabilities could be leveraged to identify specific segments of audio that have been manipulated. However, adapting YOLO from image detection to audio analysis may present challenges in capturing temporal dependencies and audio-specific features.

In the realm of deepfake video detection, YOLO-based approaches have shown promising results. The authors in [28] proposed a YOLO-CRNN-based deepfake detection approach that combines YOLO-Face for face detection with EfficientNet-B5 and Bi-LSTM for spatial-temporal feature

extraction, achieving 89.38% accuracy and outperforming state-of-the-art methods on the CelebDF-FaceForensics++ (c23) dataset. Similarly, in [29] developed a YOLO-based framework for detecting manipulated faces in videos, demonstrating good generalization across different datasets. These successes in video deepfake detection suggest potential for adapting YOLO-based approaches to the audio domain, although careful consideration of the unique challenges in audio processing will be necessary.

3 MATERIALS AND METHODS

This study employs YOLOv8 [30], a state-of-the-art object detection model, for the task of deepfake audio detection. YOLOv8, known for its efficiency and accuracy in image recognition tasks, has been adapted to process mel-spectrograms derived from audio signals. The YOLO family of models has been at the forefront of real-time object detection, and YOLOv8 represents the latest iteration with significant improvements in both speed and accuracy.

YOLOv8 introduces several key enhancements over its predecessors, utilizing a new backbone network, CSPDarknet53 [31], which employ a cross-stage partial network to better balance accuracy and computational cost. The backbone of YOLOv8 is divided into four sections, each containing a single convolution layer followed by a C2f module [32]. It also integrates a PAN and a FPN for feature fusion, along with SPP to increase the receptive field. These architectural improvements allow YOLOv8 to capture multi-scale features, which are crucial for detecting deepfake artifacts in mel-spectrograms. The overall architecture thus comprises the backbone for feature extraction, the neck for fusing those features, and a head that generates bounding boxes and class predictions.

For our deepfake audio detection task, we adapt the YOLOv8 model to use BCE loss as the primary loss function [33]. The BCE loss is defined as:

$$BCE(y, \hat{y}) = -[y \cdot \log(\hat{y}) + (1 - y) \cdot \log(1 - \hat{y})], \quad (1)$$

where y is the ground truth label (0 for real audio, 1 for deepfake) and \hat{y} is the predicted probability of the sample being a deepfake. The BCE loss function is chosen for our deepfake audio detection task due to its ability to handle binary classification problems effectively. It measures the difference between the predicted probability and the actual label, guiding the model towards more accurate predictions.

For optimization, we employ SGD [34], which updates the model parameters by following the gradient of the loss function. The parameter update rule for SGD is mathematically defined as:

$$\theta_{t+1} = \theta_t - \eta \nabla_{\theta} L(\theta_t, x^{(i)}, y^{(i)}). \quad (2)$$

In SGD, the gradients are computed using individual training samples or small batches of samples, which results in faster updates and more frequent parameter adjustments compared to full-batch gradient descent.

The learning rate η plays a crucial role in determining how large each update step is. A smaller learning rate provides more stable but slower convergence, while a larger learning rate speeds up training but risks overshooting the optimal parameter values.

We also introduce momentum to accelerate convergence and avoid oscillations during training. The momentum update rule modifies SGD as:

$$v_{t+1} = \mu v_t + \eta \nabla_{\theta} L(\theta_t, x^{(i)}, y^{(i)}), \quad (3)$$

$$\theta_{t+1} = \theta_t - v_{t+1}.$$

To adapt YOLOv8 for deepfake audio detection, we modify the final layers to output binary classifications (real or fake) instead of multiple object classes. To use audio signals with YOLOv8, we employ a multi-step approach. First, we convert the audio signals into mel-spectrograms. We then organize these mel-spectrograms into appropriate directory structures for YOLO training and create annotation files in YOLO format, specifying each spectrogram's class (real or fake).

The mel-spectrogram transformation is a critical step in our methodology, converting audio data into a visual format that can be analyzed by computer vision techniques. Mel-spectrograms represent the short-term power spectrum of sound based on a nonlinear frequency scale that approximates the human auditory system's response. This transformation allows us to capture temporal and frequency information in a format our adapted YOLOv8 model can effectively process.

The process of creating a mel-spectrogram involves several steps. First, the audio signal is divided into short, overlapping frames. We compute the STFT for each frame, which gives us the magnitude spectrum. This spectrum is then mapped onto the mel scale using a filterbank. The mel scale is a perceptual scale of pitches judged by listeners to be equal in distance from one another. The conversion from frequency f to mel scale m is given by the equation:

$$m = 2595 \cdot \log_{10} \left(1 + \frac{f}{700} \right). \quad (4)$$

This transformation emphasizes lower frequencies, which are more perceptually significant in human hearing, and compresses higher frequencies. The resulting mel-spectrogram provides a compact representation of the audio signal that captures important features for deepfake detection.

Fig. 1 shows an example of a mel-spectrogram generated from an audio sample. The x-axis represents time, the y-axis represents mel frequency bands, and the color

intensity indicates the energy level in each time-frequency bin. This visual representation allows our YOLOv8 model to identify patterns and anomalies that may indicate deepfake audio.

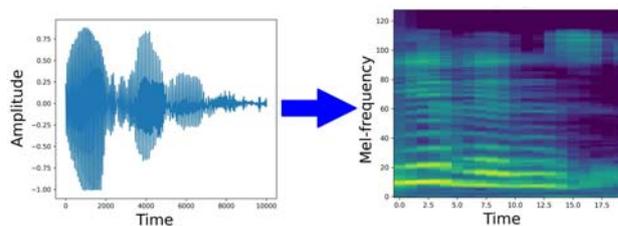


Figure 1 – Transformation of audio signal into mel-spectrogram

In addition to traditional mel-spectrograms, we also form mel-spectrograms as concatenated images from SincConv filters (Fig. 2) [35–36]. This approach allows us to leverage the benefits of learnable bandpass filters in the first layer of our neural network.

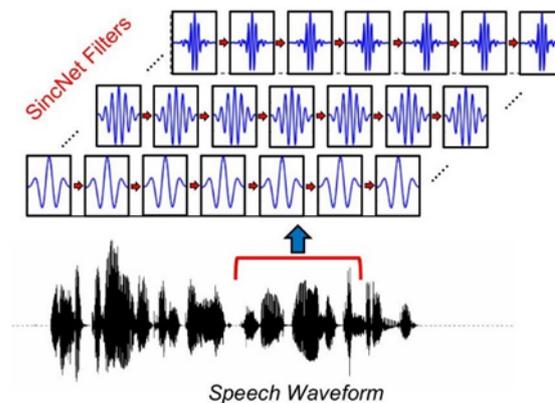


Figure 2 – SincConv filters

SincConv, or Sinc-based Convolutional Neural Networks, is a method introduced in [35] that uses sinc functions to implement band-pass filters in the first layer of a CNN. The SincConv layer learns the low and high cutoff frequencies of band-pass filters, which can be interpreted in the mel-scale, making it particularly suitable for our audio processing task.

The mathematical formulation of a SincConv filter is as follows:

$$h_{\text{sinc}}(t) = \frac{\sin(2\pi ft)}{2\pi f}. \quad (5)$$

The SincConv layer applies these filters to the raw audio waveform, effectively learning to extract relevant frequency information. The output of this layer is then processed to form mel-spectrograms. Figure 3 illustrates concatenated mel-spectrograms formed by using SincConv filters. The image shows how 15 individual mel-spectrograms, each representing the output of a different SincConv filter, are combined into a single image. This

representation allows our YOLOv8 model to analyze multiple frequency bands simultaneously, potentially improving its ability to detect subtle artifacts in deepfake audio.

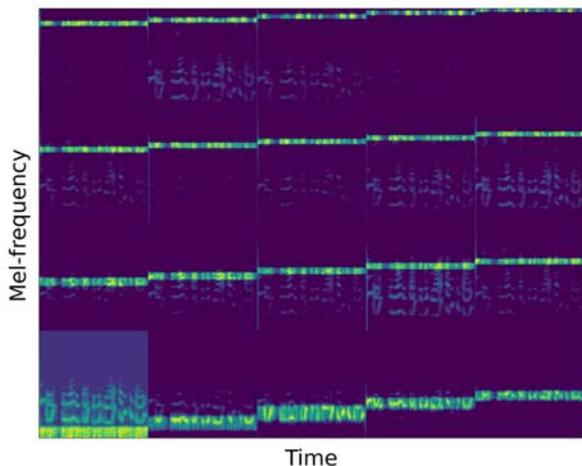


Figure 3 – Concatenated mel-spectrogram

By incorporating both traditional mel-spectrograms and those derived from SincConv filters, we provide our YOLOv8 model with rich, multi-dimensional representations of the audio signals. This approach aims to enhance the model’s capacity to distinguish between genuine and deepfake audio by capturing a wider range of spectral and temporal features.

To evaluate the performance of our YOLOv8-based model, we use the EER metric [37]. EER provides a balanced measure of false positive and false negative errors, making it particularly suitable for assessing the effectiveness of deepfake detection models. The EER is calculated as the point where the false acceptance rate (FAR) equals the false rejection rate (FRR):

$$EER = FAR = FRR \quad (6)$$

where $FAR = \frac{FP}{(FP+TN)}$ and $FRR = \frac{FN}{(FN+TP)}$. FP

represents False Positives, TN represents True Negatives, FN represents False Negatives, and TP represents True Positives. In practice, the EER is often determined by plotting the FAR and FRR curves and finding their intersection point. The lower the EER, the better the performance of the model.

4 EXPERIMENTS

Our methodology aims to address the critical challenge of cross-dataset generalization in deepfake audio detection. In this research we plan to train the YOLOv8 model on mel-spectrograms derived from one dataset and testing it on others, we seek to develop a more robust detection system that can effectively identify deepfakes across various audio sources and manipulation techniques. This approach has the potential to significantly enhance the practical applicability of deepfake audio detection in real-world scenarios.

To further explore the impact of different input representations on detection performance, we plan to train two YOLOv8 medium-size models. The first model will be trained using traditional mel-spectrograms, while the second will use concatenated mel-spectrograms formed from the output of SincConv filters. These SincConv filters cover a frequency range from 0 Hz to 8000 Hz, with a step size of approximately 533.33 Hz per filter. This comparison aims to assess whether the added frequency information provided by the SincConv-based mel-spectrograms enhances the model’s ability to detect deepfakes across datasets.

Our study employs multiple datasets to ensure robust performance across various audio types and deepfake techniques. We primarily train our models on the ASVspoof 2021 LA dataset [6], which serves as a key benchmark in audio spoofing detection, introducing more advanced TTS and VC methods for synthetic speech generation. The LA partition contains 25,380 genuine and 121,461 spoofed utterances in the training set.

After training on the ASVspoof 2021 LA dataset, we assess model generalization by testing on the ASVspoof 2021 DF evaluation set and other datasets. One of these additional datasets is the “In-the-Wild” dataset [38], which contains fakes of politicians and public figures, sourced from publicly accessible platforms. This dataset includes 17.2 hours of fake audio clips and 20.7 hours of real audio clips. By incorporating real-world deepfakes, this dataset exposes the model to more diverse manipulation techniques, providing valuable insights into how well the model performs in uncontrolled environments.

We also test our models on the Fake-or-Real Dataset [39], which includes 111,000 real utterances sourced from open datasets, TED Talks, and YouTube, alongside 87,000 synthetic utterances generated by various TTS techniques. This dataset offers a broad variety of accents, recording conditions, and speech synthesis methods, enabling us to evaluate the model’s performance under different scenarios. The diversity of real and fake utterances in this dataset further strengthens the evaluation by simulating a wide range of conditions that our model might encounter.

Additionally, we utilize the WaveFake dataset [40], which is composed entirely of synthetic speech generated by several TTS and VC architectures, including MelGAN, ParallelWaveGAN, HiFi-GAN, and WaveGlow. This dataset contains 117,985 fake audio files amounting to 196 hours of generated content. Though limited to a single speaker, WaveFake provides a focused evaluation of the model’s ability to detect audio generated by modern speech synthesis techniques.

The training method parameters were set as follows: the number of epochs – 50, the loss function – BCE, and the learning rate – 0.01. SGD was employed as the optimizer, with a momentum of 0.937 and a weight decay of 0.0005. SGD was chosen for its simplicity and effectiveness in avoiding local minima, especially when combined with momentum, which accelerates convergence and helps the model navigate through flat regions of the cost

function. A warm-up learning rate strategy was applied, with a warm-up bias learning rate of 0.1 for the first 3 epochs to ensure smoother convergence.

To enhance generalization and model robustness, various data augmentation techniques were applied, inspired by methods like those in [41]. These included random flips, with a 50% probability of vertical flipping, and horizontal flipping disabled. Small translations and scaling were introduced, with values set to 0.1 for translation and 0.5 for scaling. HSV color augmentations were used, altering hue by 0.015, saturation by 0.7, and value by 0.4, reflecting potential variations in spectrogram images derived from different audio conditions. Additionally, the model applied mixup with a probability of 0, mosaic with a probability of 1.0, and random erasing with a probability of 0.4 to further diversify the training data. These techniques helped the model become more resilient to variations in audio spectrograms, improving its ability to generalize across different datasets.

After training, each model was evaluated on various test datasets, including the ASVspoof 2021 DF evaluation set, the “In-the-Wild” dataset, the Fake-or-Real dataset, and the WaveFake dataset. For each dataset, we computed the EER as the primary performance metric, which provided a balanced measure of false acceptance and false rejection rates.

To gain further insight into the model’s behavior, we visualized feature extraction maps after each model layer. This visualization allowed us to observe how the model processed mel-spectrograms and concatenated representations from SincConv filters, providing deeper understand-

ing into how it distinguished between real and fake audio signals during detection.

5 RESULTS

We present the results of our YOLOv8-based deepfake audio detection model, including visualizations of feature extraction maps and performance metrics.

Fig. 4 and Fig. 5 display the feature extraction maps from the most informative layers of the model when processing traditional mel-spectrograms and concatenated mel-spectrograms respectively. These visualizations highlight how the model captures essential low-level features and identifies key frequency patterns and temporal changes crucial for recognizing deepfake audio artifacts.

The performance of the YOLOv8 model was evaluated using the EER across different test datasets for both traditional and concatenated mel-spectrograms. The EER is a crucial metric for assessing the model’s effectiveness in distinguishing between real and fake audio, providing a balanced measure of false acceptance and false rejection rates. The results are summarized in Table 1.

Table 1 – EER in % of YOLOv8 with different input mel-spectrograms

Model \ Dataset	YOLOv8 with simple mel-spectrograms	YOLOv8 with concatenated mel-spectrograms
ASVspoof DF	28.99	29.67
Fake or Real	39.58	35.3
In-the-wild	51.06	34.55
Wavefake	34.55	43.55

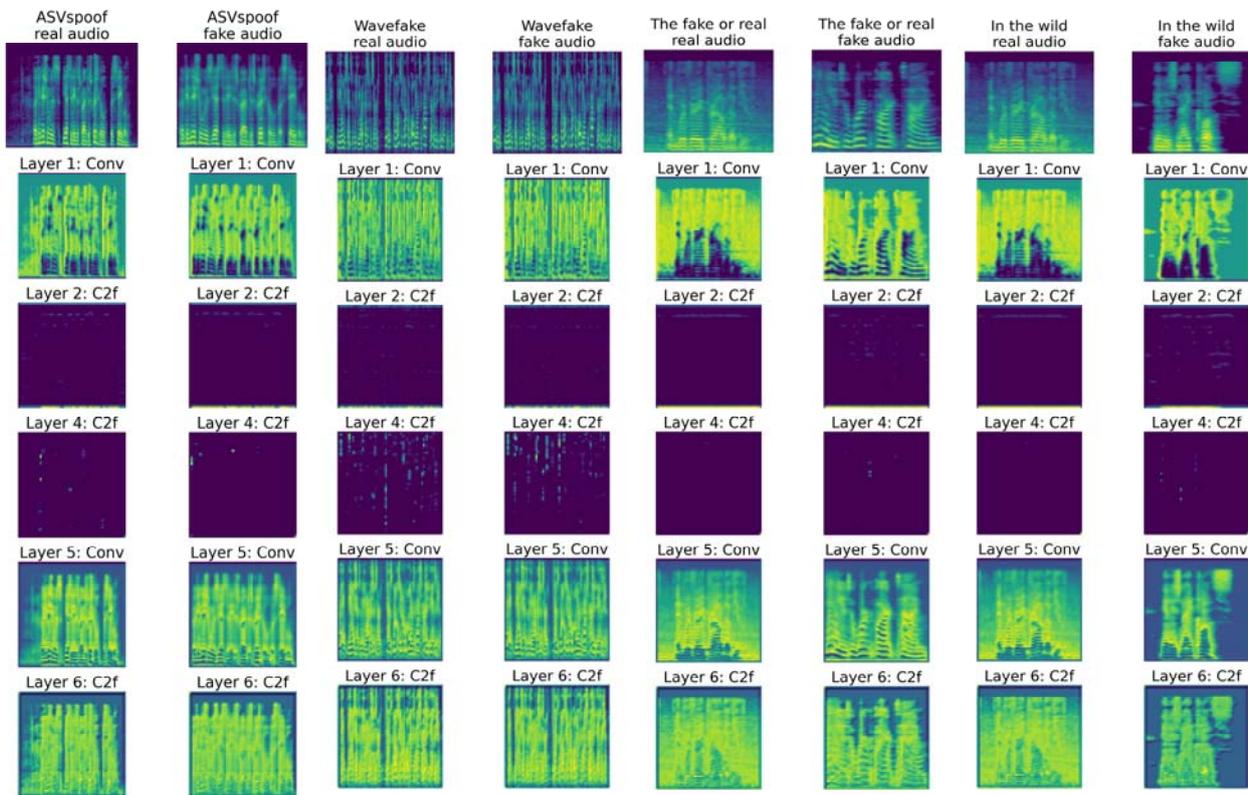


Figure 4 – The feature map of specific layers of YOLOv8 trained on simple mel-spectrograms

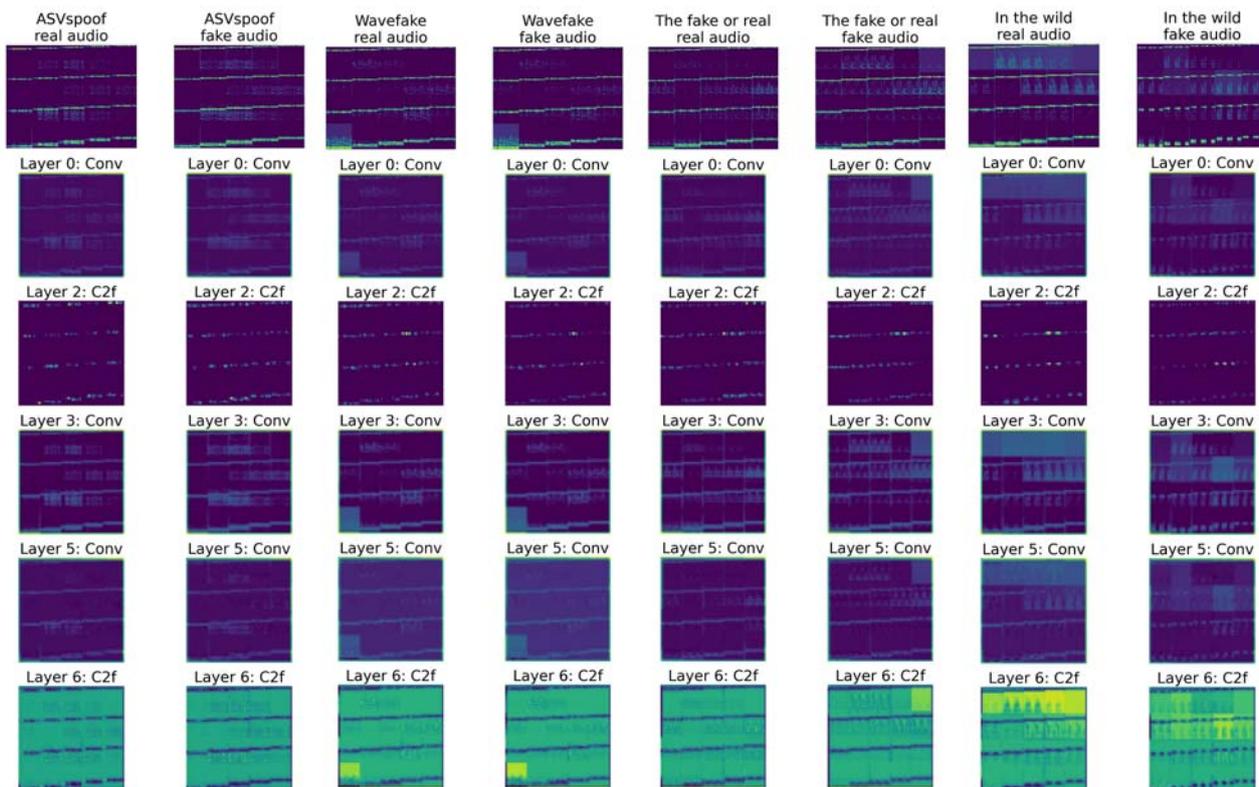


Figure 5 – The feature map of specific layers of YOLOv8 trained on concatenated mel-spectrograms

6 DISCUSSION

The results in Table 1, combined with the visualizations in Figures 4 and 5, offer valuable insights into how different input representations – simple versus concatenated mel-spectrograms – affect YOLOv8’s ability to detect deepfake audio across various datasets. Comparing the results highlights the impact of input type on detection effectiveness, particularly when considering the unique characteristics of each dataset.

For the ASVspooof DF dataset, the EER with simple Mel-spectrograms was slightly lower (28.99%) than with concatenated spectrograms (29.67%). This minor difference suggests that the concatenated input does not add substantial value for ASVspooof DF, where simpler input captures most of the distinguishing features. Figures 4 and 5 support this observation; feature maps generated from both input types appear similar, indicating that YOLOv8 can recognize key deepfake patterns in ASVspooof DF equally well, regardless of input complexity. In this case, concatenated spectrograms do not provide any significant advantage, suggesting that simpler inputs may be sufficient for datasets with clear, identifiable deepfake characteristics.

In the Fake or Real dataset, however, concatenated spectrograms significantly improve performance, reducing the EER from 39.58% with simple spectrograms to 35.3%. This improvement likely stems from the model’s ability to capture additional temporal and spectral information, as concatenation provides a richer context for identifying subtle deepfake cues. Figures 4 and 5 reflect

this difference visually: the feature maps for concatenated spectrograms in Figure 5 show more detailed attention to distinctive regions, highlighting YOLOv8’s enhanced ability to focus on nuanced patterns that simple spectrograms might overlook. This suggests that concatenated spectrograms are beneficial for datasets with higher variability, where added context helps distinguish genuine from fake samples.

The In-the-wild dataset shows the most substantial improvement with concatenated spectrograms, lowering the EER from 51.06% to 34.55%. This dataset is the most challenging due to its uncontrolled recording conditions and varied deepfake manipulations. Figures 4 and 5 illustrate how the model’s focus is more effectively distributed across relevant regions when using concatenated spectrograms, which allows YOLOv8 to capture more complex, multi-dimensional features indicative of deepfake audio. In Figure 5, the feature maps reveal a more coherent and extensive focus across critical regions, demonstrating the model’s improved capability to manage complex acoustic environments. This marked improvement with concatenated inputs underscores the importance of enhanced spectral-temporal representations when dealing with unpredictable, real-world data.

On the Wavefake dataset, in contrast, simple Mel-spectrograms produced a lower EER (34.55%) compared to concatenated spectrograms (43.55%), suggesting that the additional context from concatenated inputs may introduce noise rather than clarity. The homogeneity of the Wavefake dataset likely renders the additional informa-

tion unnecessary, and the model may perform best with a simpler, more focused input. Figures 4 and 5 further illustrate this difference, as feature maps in Figure 4 display a more targeted focus on specific regions for simple spectrograms, while the concatenated input in Figure 5 shows a diffused and less concentrated attention. This dispersion could explain the decrease in performance with concatenated spectrograms, as YOLOv8 may struggle to identify consistent patterns amidst additional, potentially irrelevant information.

Overall, the comparison of Figures 4 and 5 highlights the variability in model behavior across datasets with different input types. Concatenated Mel-spectrograms consistently offer improvements for datasets with greater variability (Fake or Real and In-the-wild), allowing YOLOv8 to capture intricate and temporally contextualized features that might otherwise go unnoticed. For datasets with more homogenous patterns, such as ASVspoof DF and Wavefake, simple Mel-spectrograms prove to be more effective by reducing noise and focusing the model's attention on specific, characteristic features. These results suggest that input representation choice should be tailored to dataset characteristics, with concatenated spectrograms being preferable for complex, varied data, while simpler spectrograms may suffice for more uniform datasets.

CONCLUSIONS

The urgent problem of deepfake audio detection is addressed through the development of a YOLOv8-based model that processes mel-spectrogram representations of audio signals.

The scientific novelty of obtained results is that a YOLOv8-based approach for deepfake audio detection is firstly proposed, which leverages both traditional and concatenated mel-spectrograms formed from SincConv filters. The model analyzes visual representations of audio signals to identify patterns indicative of synthetic speech. This approach demonstrates that computer vision techniques can be successfully adapted for audio authenticity verification, showing varying effectiveness across different types of datasets and mel-spectrogram representations.

The practical significance of obtained results is demonstrated through comprehensive experiments across multiple datasets, including ASVspoof 2021 DF, Fake or Real, In-the-wild, and Wavefake. The results reveal that the effectiveness of different mel-spectrogram representations varies significantly depending on the dataset characteristics. Concatenated mel-spectrograms showed superior performance on diverse real-world data (In-the-wild dataset, EER reduction from 51.06% to 34.55%) and the Fake or Real dataset (EER reduction from 39.58% to 35.3%). However, simple mel-spectrograms proved more effective for homogeneous datasets like Wavefake (34.55% vs 43.55% EER) and ASVspoof DF (28.99% vs 29.67% EER). This demonstrates the importance of selecting appropriate input representations based on the specific characteristics of the target audio data.

Prospects for further research include exploring additional mel-spectrogram formation techniques, investigating the impact of different YOLOv8 architectures, and developing adaptive methods that can automatically select the most appropriate mel-spectrogram representation based on dataset characteristics. Future work could also focus on improving the model's robustness against new types of audio deepfakes and reducing computational requirements while maintaining detection accuracy.

REFERENCES

1. Bondy M. Deepfakes, Digital Humans, and the Future of Entertainment in the Age of AI. *Perkins Coie*. URL: <https://perkinscoie.com/insights/blog/deepfakes-digital-humans-and-future-entertainment-age-ai> (date of access: 30.10.2024).
2. Mcuba M., Singh A., Ikuesan R. A., Venter H. The Effect of Deep Learning Methods on Deepfake Audio Detection for Digital Investigation, *Procedia Computer Science*, 2023, № 219, P. 211–219. DOI: 10.1016/j.procs.2023.01.283
3. Salvi D., Liu H., Mandelli S., Bestagini P., Zhou W., Zhang W., Tubaro S. A Robust Approach to Multimodal Deepfake Detection, *Journal of Imaging*, 2023, Vol. 9, №6, P. 122. DOI: 10.3390/jimaging9060122
4. Wang C., Yi J., Tao J., Sun H., Chen X., Tian Z., Ma H., Fan C., Fu R. Fully Automated End-to-End Fake Audio Detection, *DDAM '22: Proceedings of the 1st International Workshop on Deepfake Detection for Audio Multimedia*, 2022, pp. 27–33. DOI: 10.1145/3552466.3556530
5. Wang X., Yamagishi J. Investigating Active-learning-based Training Data Selection for Speech Spoofing Countermeasure, *2022 IEEE Spoken Language Technology Workshop (SLT)*, 2023, pp. 585–592. DOI: 10.1109/SLT54892.2023.10023350
6. Liu X., Wang X., Sahidullah M., Patino J., Delgado H., Kinnunen T. ASVspoof 2021: Towards Spoofed and Deepfake Speech Detection in the Wild, *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 2023, Vol. 31, pp. 2507–2522. DOI: 10.1109/taslp.2023.3285283.
7. Das R. K., Yang J., Li H. Long-range Acoustic and Deep Features Perspective on ASVspoof 2019, *2019 IEEE Automatic Speech Recognition and Understanding Workshop (ASRU)*, 2019, pp. 1018–1025. DOI: 10.1109/ASRU46091.2019.9003845
8. Wang X., Yamagishi J. Investigating Self-supervised Front Ends for Speech Spoofing Countermeasures, *The Speaker and Language Recognition Workshop*, 2021, pp. 100–106. DOI: 10.48550/arXiv.2111.07725
9. Xu L., Lee H., Chen Z., Chen X., Wang J. Modified Cepstral Feature for Speech Anti-spoofing, *Journal of Donghua University (English Edition)*, 2023, Vol. 40, № 2, pp. 193–201. DOI: 10.19884/j.1672-5220.202205007.
10. Leon P., Stewart B., Yamagishi J. Synthetic Speech Discrimination using Pitch Pattern Statistics Derived from Image Analysis, *Interspeech*, 2012, pp. 370–373. DOI: 10.21437/Interspeech.2012-135
11. Hong Y., Liu X., Tao J., Tian Z., Sun H. DNN Filter Bank Cepstral Coefficients for Spoofing Detection, *IEEE Access*, 2017, Vol. 5, pp. 4779–4787. DOI: 10.1109/ACCESS.2017.2687041
12. Schneider S., Baeviski A., Collobert R., Auli M. Wav2Vec: Unsupervised Pre-training for Speech Recognition, *Interspeech*, 2019, pp. 1–9. DOI: 10.21437/interspeech.2019-1873

13. Babu A., Wang P., Mohamed A., Karthik M. XLS-R: Self-supervised Cross-lingual Speech Representation Learning at Scale, *Interspeech*, 2021, pp. 2278–2282. DOI: 10.21437/Interspeech.2022-143
14. Hsu W. N., Bolte B., Tsai Y., Salakhutdinov K., Mohamed T. HuBERT: Self-Supervised Speech Representation Learning by Masked Prediction of Hidden Units, *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 2021. – Vol. 29. – P. 3451–3460.
15. Zbezhkhovska U., Khapilin O. Exploring Challenges and Future Paths in Deepfake Audio Detection, *Proceedings of the 5th Masters Symposium MS-AMLV-2024*. Lviv, Ukraine, March 29–30, 2024, pp. 1–10.
16. López J. A. V., Roddy M. P., Kinnunen T., Tan Z. Spoofing Detection with DNN and One-class SVM for the ASVspoof 2015 Challenge, *Interspeech*, 2015. DOI:10.21437/Interspeech.2015-468
17. Amin T. B., German J. S., Marziliano P. Detecting Voice Disguise from Speech Variability: Analysis of Three Glottal and Vocal Tract Measures, *Journal of the Acoustical Society of America*, 2013, Vol. 20. DOI: 10.1121/1.4879257
18. Rodríguez-Ortega Y., Ballesteros L. D. M., Renza D. A Machine Learning Model to Detect Fake Voice, *International Conference on Applied Informatics*, 2020, Vol. 1277, pp. 3–13. DOI: 10.1007/978-3-030-61702-8_1
19. Wu X., He R., Sun Z., Tan T. A Light CNN for Deep Face Representation with Noisy Labels, *IEEE Transactions on Information Forensics and Security*, 2018, Vol. 13, № 11, pp. 2884–2896. DOI: 10.1109/TIFS.2018.2833032
20. Alzantot M. F., Wang Z., Srivastava M. B. Deep Residual Neural Networks for Audio Spoofing Detection, *Interspeech*, 2019, pp. 1078–1082. DOI: 10.21437/Interspeech.2019-3174
21. Parasu P., Park K. K., Lee Y. C., Lee Y. Investigating Light-ResNet Architecture for Spoofing Detection under Mismatched Conditions, *Interspeech*, 2020, pp. 1111–1115. DOI: 10.21437/Interspeech.2020-2039
22. Gao S. H. Cheng M., Zhao K., Zhang X., Cheng M. M., Heng P. A. Res2Net: A New Multi-scale Backbone Architecture, *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, 2021, Vol. 43, №2, pp. 652–662. DOI: 10.1109/TPAMI.2019.2938758
23. Tak H., Patino J., Todisco M., Evans N. End-to-end Anti-spoofing with RawNet2, *ICASSP 2021–2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2021. – P. 6369–6373. DOI: 10.1109/ICASSP39728.2021.9414234
24. Lai C. I., Gao S. H., Chen Y. C. ASSERT: Antispoofing with Squeeze-Excitation and Residual Networks, *Interspeech*, 2019, pp. 1013–1017. DOI: 10.21437/Interspeech.2019-1794
25. Tak H., Patino J., Todisco M., Evans N. End-to-end Spectro-temporal Graph Attention Networks for Speaker Verification Anti-spoofing and Speech Deepfake Detection, *Proceedings of the 2021 Edition of the Automatic Speaker Verification and Spoofing Countermeasures Challenge*, 2021, pp. 1–8. DOI: 10.21437/ASVSPPOOF.2021-1.
26. Zeinali H., Mowlae P., Patino J., Evans N. Detecting Spoofing Attacks using VGG and SincNet: Butomilia Submission to ASVspoof 2019 Challenge, *Interspeech*, 2019, pp. 1073–1077. DOI:10.21437/interspeech.2019-2892
27. Liu X., Yang J., Sun H., Wang L. Leveraging Positional-related Local-global Dependency for Synthetic Speech Detection, *Proceedings of the 2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2023, pp. 1–5. DOI: 10.1109/ICASSP49357.2023.10096278
28. Ismail A., Elpeltagy M. S., Zaki M. S., Eldahshan K. A. Deepfake Video Detection: YOLO-Face Convolution Recurrent Approach, *PeerJ Computer Science*, 2021, Vol. 7, pp. 2–19. DOI:10.7717/peerj-cs.730
29. Hubálovský Š., Trojovský P., Bacanin N., Venkatachalam K. V. Evaluation of Deepfake Detection using YOLO with Local Binary Pattern Histogram, *PeerJ Computer Science*, 2022. – Vol. 8. DOI: 10.7717/peerj-cs.1086.
30. Sohan M., Sai Ram T., Rami Reddy C. V. A Review on YOLOv8 and Its Advancements, *Data Intelligence and Cognitive Informatics*, 2024, pp. 529–545. DOI: 10.1007/978-981-99-7962-2_39.
31. Redmon J., Farhadi A. YOLOv3: An Incremental Improvement, *CoRR, arXiv preprint arXiv:1804.02767*, 2018. DOI: 10.48550/arXiv.1804.02767
32. Terven J., Córdova-Esparza D. M., Romero-González J. A. A Comprehensive Review of YOLO Architectures in Computer Vision: From YOLOv1 to YOLOv8 and YOLO-NAS, *Machine Learning and Knowledge Extraction*, 2023, Vol. 5, № 4, pp. 1680–1716. DOI: 10.3390/make5040083
33. Reis D., Kupec J., Hong J., Daoudi A. Real-Time Flying Object Detection with YOLOv8, *arXiv preprint arXiv:2305.09972*, 2024, pp. 1–10.
34. Ruder S. An Overview of Gradient Descent Optimization Algorithms, *arXiv preprint arXiv:1609.04747*, 2017, pp. 1–14.
35. Ravanelli M., Bengio Y. Speaker Recognition from Raw Waveform with SincNet, *2018 IEEE Spoken Language Technology Workshop (SLT)*, 2018, pp. 1021–1028. DOI: 10.1109/SLT.2018.8639585
36. Zbezhkhovska U. On Effectiveness and Generalization Capabilities of Deep Learning Models for Deepfake Audio Detection, *Master Thesis. Ukrainian Catholic University, Faculty of Applied Sciences, Department of Computer Sciences*. Lviv, 2024, 44 p.
37. Kinnunen T. H. Lee K. A., Tak H., Evans N. and Nautsch A. t-EER: Parameter-Free Tandem Evaluation of Countermeasures and Biometric Comparators, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2023, Vol. 46, № 5, pp. 2622–2637. DOI: 10.1109/TPAMI.2023.3313648
38. Müller N. M., Partel J., Kinnunen T. Does Audio Deepfake Detection Generalize?, *Interspeech*, 2022, pp. 2783–2787. DOI: 10.21437/Interspeech.2022-108
39. Reimao R., Tzerpos V. FoR: A Dataset for Synthetic Speech Detection, *2019 IEEE International Conference on Speech Technology and Human-Computer Dialogue (SpeD)*, 2019, pp. 1–10. DOI: 10.1109/SpeD.2019.8878970
40. Frank J., Schönherr L. WaveFake: A Data Set to Facilitate Audio DeepFake Detection, *35th Conference on Neural Information Processing Systems (NeurIPS 2021) Track on Datasets and Benchmarks*, 2021, pp. 1–17. DOI: 10.5281/zenodo.5642694
41. Zhong Z., Zheng L., Kang G., Li S., Yang Y. Random Erasing Data Augmentation, *The Thirty-Fourth AAAI Conference on Artificial Intelligence (AAAI-20)*, 2020, pp. 13001–13008.

Received 18.11.2024.
Accepted 17.02.2025.

ВИЯВЛЕННЯ ГЛИБОКИХ ФЕЙКІВ В АУДІО ЗА ДОПОМОГОЮ YOLOV8 ТА МЕЛ-СПЕКТРОГРАМ

Збежховська У. Р. – д-р філософії, провідний науковий співробітник науково-методичного відділу забезпечення якості освітньої діяльності та вищої освіти, Харківський національний університет Повітряних сил імені Івана Кожедуба, Харків, Україна.

АНОТАЦІЯ

Актуальність. Проблема виявлення глибоких фейків у аудіо стає дедалі більш критичною в умовах швидкого розвитку технологій синтезу голосу та можливості їх використання з злочинною метою. Традиційні методи обробки аудіо стикаються з суттєвими викликами у виявленні складних аудіо фейків, особливо під час тестування на різних типах маніпуляцій з аудіо та наборах даних. Об'єктом дослідження є розробка моделі виявлення глибоких фейків у аудіо, яка використовує мел-спектрограми як вхідні дані для комп'ютерних методів зору, зосереджуючи увагу на покращенні можливостей узагальнення між наборами даних.

Мета роботи – покращення узагальнюючих можливостей моделей виявлення глибоких аудіо фейків шляхом використання мел-спектрограм та комп'ютерних методів зору. Це досягається шляхом адаптації YOLOv8, сучасної моделі комп'ютерного зору, для аналізу аудіо та дослідження ефективності різних представлень мел-спектрограм на різноманітних наборах даних.

Метод. Запропоновано новий підхід, що використовує YOLOv8 для виявлення глибоких аудіо фейків через аналіз двох типів мел-спектрограм: традиційних та конкатенованих, сформованих з фільтрів SincConv. Метод трансформує аудіосигнали в візуальні представлення, які можуть оброблятися алгоритмами комп'ютерного зору, що дозволяє виявляти тонкі шаблони, які свідчать про синтетичну мову. Запропонований підхід включає кілька ключових компонентів: оптимізацію функції втрат бінарної крос ентропії для задачі бінарної класифікації, стохастичний градієнтний спуск з моментом (0,937) для ефективного навчання та комплексні методи аугментації даних. Фільтри SincConv охоплюють частотний діапазон від 0 Гц до 8000 Гц з кроком приблизно 533,33 Гц на фільтр, забезпечуючи детальні можливості частотного аналізу. Ефективність оцінюється за допомогою метрики EER на кількох наборах даних: ASVspoof 2021 LA (25 380 справжніх та 121 461 підроблених висловлювань) для навчання, та ASVspoof 2021 DF, Fake-or-Real (111 000 реальних та 87 000 синтетичних висловлювань), In-the-Wild (17,2 години фейкових, 20,7 години реальних), та WaveFake (117 985 фейкових файлів) для тестування узагальнення між наборами даних.

Результати. Експерименти демонструють різну ефективність моделей в залежності від різних представлень вхідних даних. Конкатеновані мел-спектрограми продемонстрували кращу продуктивність на різноманітних реальних наборах даних (In-the-Wild: 34,55% EER, Fake-or-Real: 35,3% EER), тоді як прості мел-спектрограми працювали краще на більш однорідних наборах даних (ASVspoof DF: 28,99% EER, WaveFake: 34,55% EER). Візуалізації карт ознак показують, що шаблони уваги моделі значно різняться в залежності від типів вхідних даних, наприклад, конкатеновані мел-спектрограми демонструють більш розподілений фокус на відповідних областях для складних наборів даних.

Висновки. Експериментальні результати підтверджують доцільність використання YOLOv8 для виявлення глибоких аудіо фейків та демонструють, що ефективність представлень мел-спектрограм значно залежить від характеристик набору даних. Отримані результати свідчать, що представлення вхідних даних слід обирати на основі специфічних властивостей цільових аудіоданих, причому конкатеновані мел-спектрограми є більш підходящими для різноманітних реальних сценаріїв, а прості мел-спектрограми – для більш контрольованих однорідних наборів даних. Дослідження закладає основу для подальших досліджень у галузі адаптивного вибору представлення даних та оптимізації моделей для виявлення глибоких аудіо фейків.

КЛЮЧОВІ СЛОВА: виявлення глибоких фейків, YOLOv8, мел-спектрограми, узагальнюючі можливості.

ЛІТЕРАТУРА

1. Bondy M. Deepfakes, Digital Humans, and the Future of Entertainment in the Age of AI / M. Bondy // Perkins Coie. URL: <https://perkinscoie.com/insights/blog/deepfakes-digital-humans-and-future-entertainment-age-ai> (date of access: 30.10.2024).
2. The Effect of Deep Learning Methods on Deepfake Audio Detection for Digital Investigation / [M. Mcuba, A. Singh, R. A. Ikuesan, H. Venter] // Procedia Computer Science. – 2023. – № 219. – P. 211–219. DOI: 10.1016/j.procs.2023.01.283
3. A Robust Approach to Multimodal Deepfake Detection / [D. Salvi, H. Liu, S. Mandelli et al.] // Journal of Imaging. – 2023. – Vol. 9, №6. – P. 122. DOI: 10.3390/jimaging9060122
4. Fully Automated End-to-End Fake Audio Detection / [C. Wang, J. Yi, J. Tao et al.] // DDAM '22: Proceedings of the 1st International Workshop on Deepfake Detection for Audio Multimedia. – 2022. – P. 27–33. DOI: 10.1145/3552466.3556530
5. Wang X. Investigating Active-learning-based Training Data Selection for Speech Spoofing Countermeasure / X. Wang, J. Yamagishi // 2022 IEEE Spoken Language Technology Workshop (SLT). – 2023. – P. 585–592. DOI: 10.1109/SLT54892.2023.10023350
6. ASVspoof 2021: Towards Spoofed and Deepfake Speech Detection in the Wild / [X. Liu, X. Wang, M. Sahidullah et al.] // © Zbezhkhovska U. R., 2025
DOI 10.15588/1607-3274-2025-1-14
7. Das R. K. Long-range Acoustic and Deep Features Perspective on ASVspoof 2019 / R. K. Das, J. Yang, H. Li // 2019 IEEE Automatic Speech Recognition and Understanding Workshop (ASRU). – 2019. – P. 1018–1025. DOI: 10.1109/ASRU46091.2019.9003845
8. Wang X. Investigating Self-supervised Front Ends for Speech Spoofing Countermeasures / X. Wang, J. Yamagishi // The Speaker and Language Recognition Workshop. – 2021. – P. 100–106. DOI: 10.48550/arXiv.2111.07725
9. Modified Cepstral Feature for Speech Anti-spoofing / [L. Xu, H. Lee, Z. Chen et al.], X. Chen, J. Wang // Journal of Dong Hua University (English Edition). – 2023. – Vol. 40, № 2. – P. 193–201. DOI: 10.19884/j.1672-5220.202205007.
10. Leon P. Synthetic Speech Discrimination using Pitch Pattern Statistics Derived from Image Analysis / P. Leon, B. Stewart, J. Yamagishi // Interspeech. – 2012. – P. 370–373. DOI: 10.21437/Interspeech2012-135
11. DNN Filter Bank Cepstral Coefficients for Spoofing Detection / [Y. Hong, X. Liu, J. Tao et al.] // IEEE Access. – 2017. – Vol. 5. – P. 4779–4787. DOI: 10.1109/ACCESS.2017.2687041

12. Wav2Vec: Unsupervised Pre-training for Speech Recognition / [S. Schneider, A. Baevski, R. Collobert, M. Auli] // *Interspeech*. – 2019. – P. 1–9. DOI: 10.21437/interspeech.2019-1873
13. XLS-R: Self-supervised Cross-lingual Speech Representation Learning at Scale / [A. Babu, P. Wang, A. Mohamed, M. Karthik] // *Interspeech*. – 2021. – P. 2278–2282. DOI: 10.21437/Interspeech.2022-143
14. HuBERT: Self-Supervised Speech Representation Learning by Masked Prediction of Hidden Units / [W. N. Hsu, B. Bolte, Y. Tsai et al.] // *IEEE/ACM Transactions on Audio, Speech, and Language Processing*. – 2021. – Vol. 29. – P. 3451–3460.
15. Zbezhkhovska U. Exploring Challenges and Future Paths in Deepfake Audio Detection / U. Zbezhkhovska, O. Khapilin // *Proceedings of the 5th Masters Symposium MS-AMLV-2024*. – Lviv, Ukraine, March 29–30, 2024. – P. 1–10.
16. Spoofing Detection with DNN and One-class SVM for the AS-VspooF 2015 Challenge / [J. A. V. López, M. P. Roddy, T. Kinnunen, Z. Tan] // *Interspeech*. – 2015. DOI:10.21437/Interspeech.2015-468
17. Amin T. B. Detecting Voice Disguise from Speech Variability: Analysis of Three Glottal and Vocal Tract Measures / T. B. Amin, J. S. German, P. Marziliano // *Journal of the Acoustical Society of America*. – 2013. – Vol. 20. DOI: 10.1121/1.4879257
18. Rodríguez-Ortega Y. A Machine Learning Model to Detect Fake Voice / Y. Rodríguez-Ortega, L. D. M. Ballesteros, D. Renza // *International Conference on Applied Informatics*. – 2020. Vol. 1277. – P. 3–13. DOI: 10.1007/978-3-030-61702-8_1
19. A Light CNN for Deep Face Representation with Noisy Labels / [X. Wu, R. He, Z. Sun, T. Tan] // *IEEE Transactions on Information Forensics and Security*. – 2018. – Vol. 13, № 11. – P. 2884–2896. DOI: 10.1109/TIFS.2018.2833032
20. Alzantot M. F. Deep Residual Neural Networks for Audio Spoofing Detection / M. F. Alzantot, Z. Wang, M. B. Srivastava // *Interspeech*. – 2019. – P. 1078–1082. DOI: 10.21437/Interspeech.2019-3174
21. Parasu P. Investigating Light-ResNet Architecture for Spoofing Detection under Mismatched Conditions / [P. Parasu, K. K. Park, Y. C. Lee, Y. Lee] // *Interspeech*. – 2020. – P. 1111–1115. DOI: 10.21437/Interspeech.2020-2039
22. Res2Net: A New Multi-scale Backbone Architecture / [S. H. Gao, M. Cheng, K. Zhao et al.] // *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*. – 2021. – Vol. 43, № 2. – P. 652–662. DOI: 10.1109/TPAMI.2019.2938758
23. End-to-end Anti-spoofing with RawNet2 / [H. Tak, J. Patino, M. Todisco, N. Evans] // *ICASSP 2021–2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. – 2021. – P. 6369–6373. DOI: 10.1109/ICASSP39728.2021.9414234
24. Lai C. I. ASSERT: Antispoofing with Squeeze-Excitation and Residual Networks / C. I. Lai, S. H. Gao, Y. C. Chen // *Interspeech*. – 2019. – P. 1013–1017. DOI: 10.21437/Interspeech.2019-1794
25. End-to-end Spectro-temporal Graph Attention Networks for Speaker Verification Anti-spoofing and Speech Deepfake Detection / [H. Tak, J. Patino, M. Todisco, N. Evans] // *Proceedings of the 2021 Edition of the Automatic Speaker Verification and Spoofing Countermeasures Challenge*. – 2021. – P. 1–8. DOI: 10.21437/ASVSPPOOF.2021-1.
26. Detecting Spoofing Attacks using VGG and SincNet: Butomilia Submission to ASVspooF 2019 Challenge / [H. Zeinali, P. Mowlaee, J. Patino, N. Evans] // *Interspeech*. – 2019. – P. 1073–1077. DOI:10.21437/interspeech.2019-2892
27. Leveraging Positional-related Local-global Dependency for Synthetic Speech Detection / [X. Liu, J. Yang, H. Sun, L. Wang] // *Proceedings of the 2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. – 2023. – P. 1–5. DOI: 10.1109/ICASSP49357.2023.10096278
28. Deepfake Video Detection: YOLO-Face Convolution Recurrent Approach / [A. Ismail, M. S. Elpeltagy, M. S. Zaki, K. A. Eldahshan] // *PeerJ Computer Science*. – 2021. – Vol. 7. – P. 2–19. DOI:10.7717/peerj-cs.730
29. Hubálovský Š. Evaluation of Deepfake Detection using YOLO with Local Binary Pattern Histogram / [Š. Hubálovský, P. Trojovský, N. Bacanin, Kv. Venkatachalam] // *PeerJ Computer Science*. – 2022. – Vol. 8. DOI: 10.7717/peerj-cs.1086.
30. Sohan M. A Review on YOLOv8 and Its Advancements / M. Sohan, T. Sai Ram, C. V. Rami Reddy // *Data Intelligence and Cognitive Informatics*. – 2024. – P. 529–545. DOI: 10.1007/978-981-99-7962-2_39.
31. Redmon J. YOLOv3: An Incremental Improvement / J. Redmon, A. Farhadi // *CoRR*, arXiv preprint arXiv:1804.02767. – 2018. DOI: 10.48550/arXiv.1804.02767
32. Terven J. A Comprehensive Review of YOLO Architectures in Computer Vision: From YOLOv1 to YOLOv8 and YOLO-NAS / J. Terven, D. M. Córdoba-Esparza, J. A. Romero-González // *Machine Learning and Knowledge Extraction*. – 2023. – Vol. 5, № 4. – P. 1680–1716. DOI: 10.3390/make5040083
33. Real-Time Flying Object Detection with YOLOv8 / [D. Reis, J. Kupec, J. Hong, A. Daoudi] // arXiv preprint arXiv:2305.09972. – 2024. – P. 1–10.
34. Ruder S. An Overview of Gradient Descent Optimization Algorithms / S. Ruder // arXiv preprint arXiv:1609.04747. – 2017. – P. 1–14.
35. Ravanelli M. Speaker Recognition from Raw Waveform with SincNet / M. Ravanelli, Y. Bengio // *2018 IEEE Spoken Language Technology Workshop (SLT)*. – 2018. – P. 1021–1028. DOI: 10.1109/SLT.2018.8639585
36. Zbezhkhovska U. On Effectiveness and Generalization Capabilities of Deep Learning Models for Deepfake Audio Detection / U. Zbezhkhovska // *Master Thesis. Ukrainian Catholic University, Faculty of Applied Sciences, Department of Computer Sciences*. – Lviv, 2024. – 44 p.
37. t-EER: Parameter-Free Tandem Evaluation of Countermeasures and Biometric Comparators / [T. H. Kinnunen, K. A. Lee, H. Tak et al.] // *IEEE Transactions on Pattern Analysis and Machine Intelligence*. – 2023. – Vol. 46, № 5. – P. 2622–2637. DOI: 10.1109/TPAMI.2023.3313648
38. Müller N. M. Does Audio Deepfake Detection Generalize? / N. M. Müller, J. Partel, T. Kinnunen // *Interspeech*. – 2022. – P. 2783–2787. DOI: 10.21437/Interspeech.2022-108
39. Reimao R. FoR: A Dataset for Synthetic Speech Detection / R. Reimao, V. Tzerpos // *2019 IEEE International Conference on Speech Technology and Human-Computer Dialogue (SpeD)*. – 2019. – P. 1–10. DOI: 10.1109/SpeD.2019.8878970
40. Frank J. WaveFake: A Data Set to Facilitate Audio DeepFake Detection / J. Frank, L. Schönherr // *35th Conference on Neural Information Processing Systems (NeurIPS 2021) Track on Datasets and Benchmarks*. – 2021. – P. 1–17. DOI: 10.5281/zenodo.5642694
41. Random Erasing Data Augmentation / [Z. Zhong, L. Zheng, G. Kang et al.] // *The Thirty-Fourth AAAI Conference on Artificial Intelligence (AAAI-20)*. – 2020. – P. 13001–13008.

ПРОГРЕСИВНІ ІНФОРМАЦІЙНІ ТЕХНОЛОГІЇ

PROGRESSIVE INFORMATION TECHNOLOGIES

UDC 004.9

SEGMENTATION OF LOW-CONTRAST IMAGES IN THE BASIS OF EIGEN SUBSPACES OF TYPE-2 FUZZY MEMBERSHIP FUNCTIONS

Akhmetshina L. G. – Dr. Sc., Professor of the Department of Electronic Computing Machinery, Oles Honchar Dnipro National University, Dnipro, Ukraine.

Yegorov A. A. – PhD, Senior Lecturer of the Department of Computer Sciences and Information Technologies, Oles Honchar Dnipro National University, Dnipro, Ukraine.

Fomin A. A. – Post-graduate student of the Department of Electronic Computing Machinery, Oles Honchar Dnipro National University, Dnipro, Ukraine.

ABSTRACT

Context. The study addresses the current task of automating a sensitive image segmentation algorithm based on the Type-2 fuzzy clustering method. The research object is low-contrast greyscale images which are outcomes of standard research methods across various fields of human activity.

Objective. The aim of the work is to create a new set of informative features based on the input data, perform sensitive fuzzy segmentation using a clustering method that employs Type-2 fuzziness, and implement automatic defuzzification in eigen subspace of membership functions.

Method. A method for segmenting low-contrast images is proposed. It consists of the following steps: expanding the feature space of the input data, applying singular value decomposition (SVD) to the extended dataset with subsequent automatic selection of the most significant components, which serve as input for fuzzy clustering using Type-2 fuzzy sets. Clustering is performed using the T2FCM method, which allows the automatic selection of the number of fuzzy clusters based on an initially larger guaranteed number, followed by the merging of close clusters (proximity was defined in the study using a weighted Euclidean distance). After fuzzy clustering, the proposed method integrates its results (fuzzy membership functions) with the input data for clustering, preprocessed using fuzzy transformations. The resulting matrix undergoes another fuzzy transformation, followed by SVD and the automatic selection of the most significant components. A grayscale image is formed based on the weighted sum of these selected components, to which the adaptive histogram equalization method is applied, resulting in the final segmentation output. The proposed segmentation method involves a small number of control parameters: the initial number of fuzzy clusters, the error of the T2FCM method, the maximum number of iterations, and the coefficient of applied fuzzy transformations. Adjusting these parameters to the processed images does not require significant effort.

Results. The developed algorithm has been implemented as software, and experiments have been conducted on real images of different physical nature.

Conclusions. The experiments confirmed the efficiency of the proposed algorithm and recommend its practical application for visual analysis of low-contrast grayscale images. Future research prospects may include analyzing the informative potential of the algorithm when using other types of transformations of fuzzy membership functions and modifying the proposed algorithm for segmenting images of various types.

KEYWORDS: Image Segmentation, Fuzzy Clustering, Type-2 Fuzzy Clustering, orthogonal transformation, singular value decomposition, singular subspaces.

ABBREVIATIONS

2DPCA is a Two-Dimensional Principal Component Analysis;
FCM is a fuzzy clustering algorithm;
T2 refers to Type-2 fuzziness;
T2FCM is a fuzzy clustering algorithm based on Type-2 fuzziness;
MRI is Magnetic Resonance Imaging;
PCA is a Principal Component Analysis method.

NOMENCLATURE

N is a number of pixels of an image;
 M is a width of an image;
 L is a height of an image;
 C is a vector of the coefficients of algorithm;
 dC is a vector of differences of neighbouring elements of vector C of algorithm;
 dC_{\min} is a minimum element of vector dC ;
 dC_{\max} is a maximum element of vector dC ;
 c is a number of fuzzy clusters;

dC_a is a coefficient of algorithm;
 K is a coefficient of algorithm;
 I is an input image;
 I^{out} is a segmented image;
 I_{max}^l is a maximal value of processing window;
 I_{min}^l is a minimal value of processing window;
 I_{max}^g is a maximal value of image;
 I_{min}^g is a minimal value of image;
 I_0 is a center pixel of processing window;
 I^i is image at various stages of algorithm, $i \in \{svd, tr, u, s\}$;
 u_{ji} is a degree of membership of the object i to cluster j ;
 m is a weight coefficient that characterizes a measure of fuzziness;
 v_j is a center of cluster j ;
 d_{ij} is an Euclidean distance between a center of cluster v_j and an instance of original data x_i ;
 U is a membership function;
 w_j is a singular value from singular value decomposition, $j \in \{1, 1-\delta\}$;
 x is a pixel coordinate;
 y is a pixel coordinate;
 z is a pixel coordinate.

INTRODUCTION

Image segmentation refers to a high level of processing and it is a mandatory stage in there most image analysis technologies. Currently, there is no universal algorithm for its implementation and the result largely depends on the quality of the initial data [1]. Low-contrast images are often used in practice, and the insufficient quality of which is due to both the features of the equipment (heterogeneity of fields, nonlinearity of sensor's characteristics, noise) and the process of their formation (anatomical features of the analysis objects, dynamic distortions, etc.). In addition to randomness, which can be controlled in accordance with the theory of probability, the objective property of images is the presence of uncertainty and ambiguity, which must be taken into account during processing.

The object of study is the process of grayscale low-contrast images' segmentation that are the result of standard research methods.

The segmentation process often involves the synthesis of the new information parameters formed on the basis of initial data which make it possible to increase the values of the initial brightness characteristics' variations and to separate the information from different sources. This is an ambiguous task related to the field of artificial intelligence. The reliability of the result depends significantly on the type and characteristics of the initial data and, as a rule, there is no a priori information about the system of their formation and noise component.

To increase the validity and sensitivity of segmentation it is necessary to take into account the ambiguity and uncertainty that are always present in digital images.

The subject of study is the technology of grayscale low-contrast images' segmentation based on the use of the T2FCM method (fuzzy clustering type_2) with automatic determination of the number of clusters with further conversion to the singular subspaces of the obtained membership functions using singular value decomposition and synthesis of the resulting image based on its eigen images of orthogonal components.

The known fuzzy clustering algorithms [2–13] are usually characterized by uncertainty in determining the number of clusters and initializing initial values, ambiguity of the defuzzification process and also, they do not take into account the spatial information. Besides, they are sensitive to the noise component.

The purpose of the work is to increase the sensitivity and reliability of grayscale low-contrast images' segmentation and, to a certain extent, to neutralize the sensitivity to the influence of noise factors (due to the synthesis of a segmented image based on its eigen images of fuzzy membership functions).

1 PROBLEM STATEMENT

It is supposed that the low-contrast grayscale image I is given as a set of brightness values I_i for $i=1, 2, \dots, N$ pixels with coordinates (x, y) , $x = 1, 2, \dots, M$, $y = 1, 2, \dots, L$.

The synthesis problem of a segmented image can be represented as the problem of visualizing the result of its clustering on the basis of a fuzzy membership functions' matrix $U: \langle u_{ki}, I_i \rangle$, where u_{ki} is a fuzzy membership function of an instance of initial data I_i with coordinates (x, y) to the k -th cluster, $k=1, 2, \dots, c$, and $\forall k (\forall i (\sum u_{ki} = 1))$.

The IU transformation: $f(\langle u_{ki}^m \rangle, \langle d_{ki} \rangle) \rightarrow \min$, where d_{ki} is the Euclidean distance between the centre of the cluster v_j and the object I_i ; m is the weight coefficient that characterizes the measure of fuzziness. The clustering parameters c, m are chosen experimentally.

Creation of a segmented image $I^{out} : f(\forall k (\forall i (u_{ki}))$.

2 REVIEW OF THE LITERATURE

Image segmentation is a complex analysis procedure. It belongs to the tasks of unsupervised learning. Modern algorithms can be used in conditions of almost complete lack of information about the data distribution laws. They use the measure of object parameters' proximity in a multidimensional space and rely solely on heuristic considerations about the nature and characteristics of the investigated set.

Inaccuracy and uncertainty are present in all digital images and can lead to errors while forming data for analysis based on brightness characteristics [2]. The concept of "ambiguity of grey" reflects the fact that the accuracy of brightness values is primarily limited by the dig-

itization process (quantization in amplitude and spatial discretization), as well as, by the precision of the hardware systems. For example, ambiguity in medical images is caused by the forming fields' heterogeneity or movement of a patient. It's the weather conditions in satellite images. It's the inaccuracy and unevenness of the measurement grid in geological field images. It's the uneven background, significant noise, aberration artifacts, etc. in microscopic images. Geometric fuzziness manifests when determining object boundaries because of deformations from motion capture, insufficient resolution, low contrast and noise influence. It should be noted separately the need to consider the peculiarities of a human eye as an instrument of visual analysis which adheres to Weber's laws. In particular, it cannot detect changes in the grey level below the threshold of visual perception [3, 4].

The procedure of image segmentation to enhance analysis reliability must correspond to the contradictory requirements – noise removal while preserving fine details, as well as the delineation of object boundaries without excessive detailing and emerging artifacts. One of the ways to solve the problem of improving images' quality and the reliability of their analysis is based on the formation of the new informative features on the assumption of the initial data, which change the relationship (the space structure) of the analyzed features [5]. The use of fuzzy logic due to nonlinearity allows to increase the influence of variations in the brightness properties and eliminate ambiguity of the initial data [6]. The processing medical images using a fuzzy approach allows to increase resolution, highlight the features of the individual areas' structure and improve the accuracy of segmentation [7]. It is necessary to carry out preliminary conversion (improving) for noisy low contrast images before the targeted processing is performed. Most often, the local characteristics of each pixel are converted.

The key problem for the implementation of sensitive segmentation is the formation of a relevant set of inputs that provide a solution to the problem. The synthesis of information parameters is considered as the transforming process of the images' initial brightness characteristics into a new virtual space where there is a redistribution of brightness characteristics, which allowing to increase the analyzed brightness range and (or) the metric distance between the object of interest and the "background". Choosing different types of source image transformations and segmentation methods leads to different results.

Fuzzy clustering methods provide a set-theoretic partitioning into subsets and associate each object with a fuzzy set with a membership function that varies in the interval $[0, 1]$ [8]. The first developed fuzzy clustering method was the FCM (Fuzzy C-Means) algorithm. In 1980 J. Bezdek proved its convergence and in 1981 he generalized the algorithm in the case of arbitrary fuzzy sets [9]. Today there are a significant number of FCM method modifications aimed at solving the specific problems with maximum consideration of images' specifics [10–13].

The FCM algorithm problems can be formulated as follows:

1. There is no theoretical justification for the importance of the choice of the clusters' number and the fuzziness degree;
2. The use of pseudorandom initial values of centroids leads to different results;
3. Noise sensitivity;
4. The spatial information which is essential for image processing is not used [14, 15];
5. The process of defuzzification is ambiguous.

The membership functions obtained within the framework of FCM method can be interpreted as an ensemble of multiparameter data with the possibility of using multivariate analysis algorithms for its processing, for example, orthogonalization methods. The ideas of projection methods into eigen subspaces as one of the tools for mathematical processing of experimental data, were presented in the works [16, 17].

In particular, methods called principal component analysis (PCA) involve the use of statistical principles to reduce the number of inputs in order to extract the most significant factors from the input data. This approach in the problems of processing of images which are two-dimensional structures was applied in practice only in the 2000s [18]. Two-dimensional principal component analysis (2DPCA) and other orthogonalization techniques are currently used to solve such important problems as compression of visual information, feature extraction in object recognition and search for video images, reduction of calculations in image processing, etc. [19, 20]. The image covariance matrix is constructed directly with using the original image matrices and its eigenvectors are computed to highlight the features of the image that determine the internal structure of the experimental data.

This article deals with the information possibilities of using the method of multivariate information analyzing based on fuzzy clustering and singular decomposition in the analysis of low-contrast greyscale images.

3 MATERIALS AND METHODS

The formalization of the term "fuzzy set" involves generalizing the concept of membership reflecting the idea that elements of a set can have a common property to varying degrees. Unlike the probability which is connected with the uncertainty regarding an object's membership in a crisp set, fuzzy logic provides a foundation for developing a more flexible approach to data analysis and allows to make a decision among a set of alternative options. The general scheme of fuzzy image processing includes the following stages: input parameter formation, fuzzification, processing/interpretation of membership function values and defuzzification on the basis of which the final result is formed [21, 22].

The mutual mapping of grey levels and fuzzy membership function surfaces can be interpreted as a specific type of nonlinear coding-decoding, in terms of fuzzy logic is data fuzzification-defuzzification. The complexity of the fuzzy approach is the uncertainty in the

formation of the final result in a multidimensional space. The methods for its implementation are determined mainly by the purpose of processing and differ in the peculiarities of taking into account the topology and properties of the analyzed image, for example, the dynamic range of brightness, the used color model, the number of channels, the size of the ensemble, etc.

Image I of size $N = M \times L$ can be represented as an array of fuzzy sets regarding the analyzed property, in particular, brightness, with the value of the membership function $u_{x,y}$ varying in the interval $[0,1]$, for each pixel:

$$I = \bigcup_{x=1}^M \bigcup_{y=1}^L u_{x,y} \quad (1)$$

The fuzzification process can be carried out in various ways. For example, fuzzification based on histogram analysis refers to global methods that take into account the grey level of each pixel on the basis of which the membership function to one or more classes, such as very dark, slightly bright, medium, etc., is determined according to the given requirements. Its use requires prior knowledge about the analyzed image; for example, the minimum and maximum of grey levels' frequencies. However, the accuracy of these points' detecting on the histogram does not need to be very high because the concept of fuzziness is used.

Neighbourhood-based fuzzification takes into account a defined neighbourhood of pixels and typically requires more computational time compared to the histogram-based approach.

This approach requires additional data analysis because noise and outliers (anomalous values) can lead to false values of the membership functions for a subset of pixels. For example, for a window of size $r \times r$ the membership function of the central pixel can be described as follows:

$$u = 1 - \left[1 + \frac{1}{\max(I_i)} \sum_{i=1}^{r-1} \|I_0 - I_i\| \right]^{-1} \quad (2)$$

or

$$u = 1 - \left[1 - \frac{I_{\max}^l - I_{\min}^l}{I_{\max}^g - I_{\min}^g} \right]^{-1}, \quad (3)$$

where, I_{\max}^l , I_{\min}^l , I_{\max}^g , I_{\min}^g , are the local and global extremes of the window and the image as a whole, respectively; I_0 is the central pixel.

To solve high-level problems, it is necessary to extract properties of the analyzed image (such as the object length, the region homogeneity, the entropy, the average value, etc.), which are then subjected to fuzzification.

Fuzzy clustering techniques that iteratively determine the matrix U – the degree of pixels' belonging to fuzzy clusters can also be interpreted as a fuzzification stage.

The simplest way to perform the defuzzification procedure of the clustering result is to select the value corresponding to the maximum of the membership function. When using this approach, a new parameter whose visualization provides the image segmentation is synthesized as follows:

$$I_{x,y}^{out} = \max(u_{k,x,y}), \forall k \in [1, c], \quad (4)$$

where $u_{k,x,y}$ is the membership function of a pixel in the output image with coordinates x, y to the k -th fuzzy cluster. The resulting image will always be greyscale.

Another method of visualization is to form an output image I^{out} based on the centres of the clusters (v):

$$I_{x,y}^{out} = \sqrt{\sum_{j=1}^q v_{I_{x,y,j}^{in}}^2}. \quad (5)$$

However, the values of the membership function can have comparable or even equal values for different clusters, which leads to ambiguity and therefore, unreliability of defuzzification. Fig.1 c shows the type of two membership functions' graphs for clustering into 6 clusters for the model image in Fig. 1 a whose histogram is presented in Fig. 1 b where the two membership functions have a substantial number of identical values.

On the other hand, each class, by definition, also contains the information suitable for the analysis that can be lost when performing expression-based defuzzification (4, 5). If the membership function of each class is interpreted as an image where each one carries specific information, then during segmentation, it is necessary to form a "composite" image based on the image merging of all membership functions.

Fuzzy logic of type_2 (T2) makes it possible to consider the problems with a higher degree of uncertainty, in particular, in the methods of image representation and in the algorithms for their processing [23].

The theoretical basis of our approach to the defuzzification process is the fact that by specifying a specific number of c classes for fuzzy clustering, a three-dimensional matrix U containing a set of c values of membership functions for each field pixel is gotten. The dimension of the third z coordinate is equal to the specified number of classes. The latter means that within the framework of FCM method it is possible to obtain automatically an ensemble of multidimensional data and apply methods for multidimensional information processing. Each membership function can be interpreted as an image, and the U array can be interpreted as a multidimensional image with c channels.

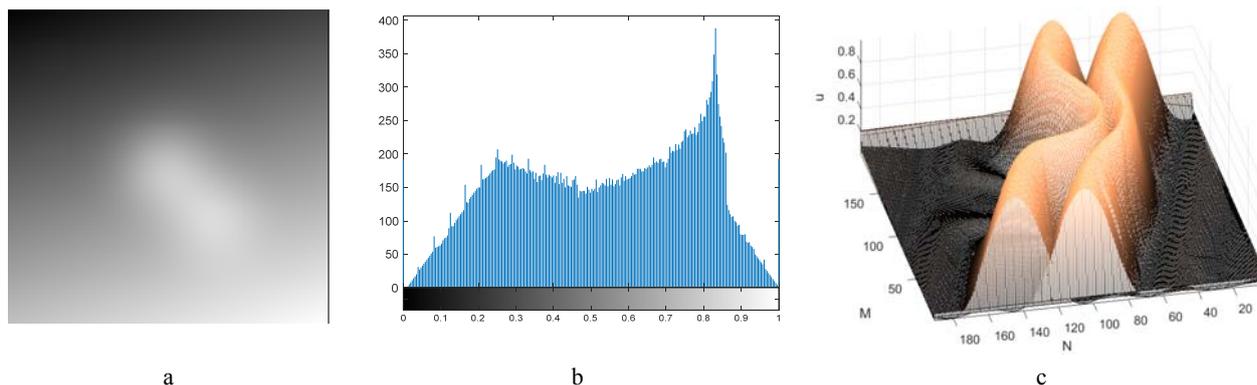


Figure 1 – View of the Fuzzy Membership Functions: a – model image; b – histogram; c – membership functions of two classes

In particular, the application of singular value decomposition (SVD) for multidimensional data enables a transition to a new (eigen) orthogonal subspace, where each new component results from a linear composition of all original parameters. This allows for a comprehensive analysis. The informational contribution of each new component can be determined by analyzing the spectrum of normalized singular values $w_i, i \in [1, n]$, and $w_1 \geq w_2 \geq \dots \geq w_n \geq 0$.

Applying the singular transformation to the matrix of fuzzy cluster membership functions forms an orthonormal “eigen images” basis of these membership functions. Each of the “eigen images” contains H_i % of all the information contained in the original ensemble of the matrix U . This circumstance allows defuzzification to be performed based on the information contained in all membership functions. By additionally imposing, for example, a constraint:

$$\sum_{i=1}^p \hat{w} \geq H, \quad (6)$$

where H is a specified boundary value, such as 95% or 99%, $p \leq n$, it becomes possible to influence the sensitivity of segmentation. The physical meaning of the expression (6) is a filtration of non-essential information components (noise).

Based on the above, an image segmentation algorithm is proposed that, firstly, synthesizes new informative parameters based on the original brightness characteristics, secondly, applies a type-2 fuzzy clustering algorithm with automatic cluster number determination, and thirdly, performs automatic defuzzification using the “eigen images” of membership functions.

The algorithm includes the following steps:

1. Transforming the original greyscale image brightness values using a window transformation (window size 3x3), forming a 9-dimensional ensemble including the brightness of neighbouring pixels to take into account for spatial characteristics.

2. Performing an orthogonal transformation (SVD) on the expanded original data, followed by automatic selection of the most significant components [24] based on calculating the coefficient vector C by the formula:

$$C_i = \frac{\left| \sum_{j=1}^{1+c} (V_s)_{i,j} \right| + \left| \sum_{j=1}^{1+c} (V_s)_{j,i} \right|}{2}, i \in [1, 1+c] \quad (7)$$

where V_s contains the right singular vectors for the SVD. This vector C is sorted in descending order, and a difference vector dC is created, containing the differences for each neighbouring pair in the sorted vector C . The value dC_a is then calculated as:

$$dC_a = \frac{\sum_{j=1}^c \frac{dC_j}{c} + \frac{dC_{\min} + dC_{\max}}{2}}{2}, \quad (8)$$

where dC_{\min} and dC_{\max} are the minimum and maximum elements of vector dC , respectively. This threshold value dC_a is used to determine the number of the most significant elements of the left singular vector matrix. The selected indices correspond to the original indices in vector C prior to sorting I^{svd} .

3. Scaling each component of matrix I^{svd} to the interval $[0,1]$, forming a multidimensional input matrix for fuzzy data clustering.

4. Performing fuzzy clustering on the scaled matrix I^{svd} using the T2FCM [25] method with dynamic compression of the fuzzy membership function [9]. This involves initially setting a larger number of fuzzy clusters, which dynamically decreases during training by merging close clusters. The weighted Euclidean distance is used to determine closeness, calculated (for the distance between the centers of the k -th and l -th clusters) as follows:

$$d_{k,l}^0 = \sqrt{\sum_{j=1}^q (S_{u_k} \cdot v_{k,j}^t - S_{u_l} \cdot v_{l,j}^t)^2}, \quad (9)$$

where values S_{u_k} and S_{u_l} are calculated as:

$$S_{u_j} = \frac{\overline{u_j}}{u_{\min}}, (\forall j \in \{1, \dots, c\}), \quad (10)$$

where $\overline{u_j}$ is the average membership value for the j -th fuzzy cluster, and $\overline{u_{\min}}$ is the minimum value of the vector \overline{u} of average membership values for each cluster.

During each training iteration t , the MFT2 matrix a^t is determined based on the difference between the “upper” u_h^t and “lower” u_l^t membership functions as follows:

$$(u_{h1}^t)_{i,k} = \left((u^t)_{i,k} \right)^{1 - \left((u^t)_{i,k} \right)^+ + \left((u^t)_{i,k} \right)^{-} \left(\overline{u^t} \right)_k^{-0.5}}, \quad (11)$$

$$(u_{l1}^t)_{i,k} = \left((u^t)_{i,k} \right)^{1 + \left((u^t)_{i,k} \right)^+ - \left((u^t)_{i,k} \right)^{-} \left(\overline{u^t} \right)_k^{+0.5}}, \quad (12)$$

$$(u_{h2}^t)_{i,k} = \left((u^t)_{i,k} \right)^{K + 0.75 + \left(\overline{u^t} \right)_k / 2}, \quad (13)$$

$$(u_{l2}^t)_{i,k} = \left((u^t)_{i,k} \right)^{K + 1.25 + \left(\overline{u^t} \right)_k / 2}, \quad (14)$$

$$\left(\overline{u^t} \right)_k = \left((u^t)_{i,k} \right)^{1 - \max \left(\left((u^t)_{i,k} \right)^+, 1 - \left((u^t)_{i,k} \right)^- \right)}, \quad (15)$$

$$(u_h^t)_{i,k} = \left((u_{h1}^t)_{i,k} + (u_{h2}^t)_{i,k} \right) / 2, \quad (16)$$

$$(u_l^t)_{i,k} = \left((u_{l1}^t)_{i,k} + (u_{l2}^t)_{i,k} \right) / 2, \quad (17)$$

where $\overline{u^t}_k$ is the average membership for the k -th cluster, and $i \in [1, N]$, K is the coefficient, values of which are recommended to be chosen within the interval $[0, 0.2]$. This coefficient has a significantly impact the transformation results. The final membership function matrix a^t is then interpreted as a multidimensional image.

5. Transforming the scaled matrix I^{svd} by the formula:

$$I_{i,j}^{tr} = \left(I_h^{svd} \right)_{i,j} - \left(I_l^{svd} \right)_{i,j}, \quad (18)$$

where I_h^{svd} and I_l^{svd} are calculated as per formulas (11) and (12), respectively, forming matrix I^{tr} .

6. Constructing matrix I^u as the union of the columns of matrices I^{tr} and a^t .

7. Applying the transformation described in step 4 to matrix I^u to compute matrix a^t (formulas (11)–(17)), resulting in matrix I^{T2} .

8. An orthogonal transformation (SVD) is applied to the matrix I^{T2} . Following this, the most significant components are selected from the left singular vectors, as described in step 2. A grayscale image I^s is formed using a weighted sum of selected significant components based on the values of the vector C , corresponding to the selected significant components of the left singular vectors, scaled so that their sum equals 1.

9. Applying the adaptive histogram equalization method (using a uniform transformation function) to the image I^s , forming the final grayscale image I^{out} .

4 EXPERIMENTS

Traditionally, the number of clusters for fuzzy segmentation is determined empirically, and defuzzification is based on the maximum of the membership function. Our algorithm employs the T2FCM method with automatic determination of the final number of clusters. The following control parameters were used: fuzziness coefficient $m = 2$, exit condition $\varepsilon \leq 10^{-5}$, maximum iteration number set at 100, and an initial number of fuzzy clusters $c = 9$ (the experiments were conducted on medical images, and therefore, this amount is more than enough). The starting initialization of the cluster centers for the T2FCM method was carried out by filling them with values from the original data vectors. Coefficient $K = 0.075$ (for formulas (13) and (14)) was used in all experiments.

It was found that for some images, the best results could be obtained without applying step 7 in the proposed algorithm (using matrix I^u in step 8 instead of I^{T2}).

The presented experiments were conducted on images of various physical nature, termed “low-contrast”. Examples of images and histograms are shown in Figure 2: MRI (a, b) and X-ray (c, d), respectively. Unlike typical low-contrast images, characterized by narrow histograms in low brightness regions, these images have features such as multimodal histograms, a full or nearly full intensity range with significant dark and light regions, smooth brightness transitions in areas of interest, and blurred boundaries, lack of a priori information about the presence and location of anomalies that can be compared with the noise level making visual analysis challenging.

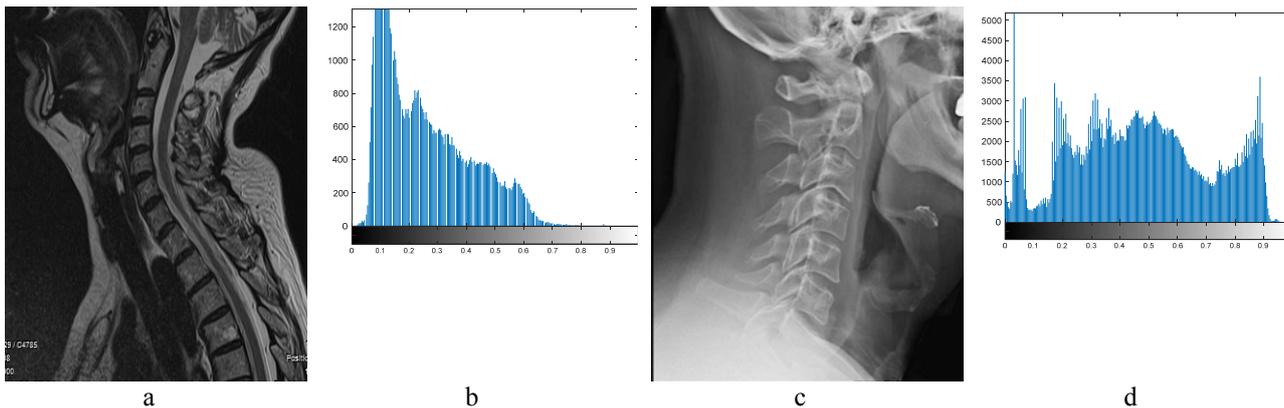


Figure 2 – Low-contrast images and their histograms: a, b – MRI; c, d – X-ray

5 RESULTS

Figure 3 illustrates various segmentation methods applied to the MRI image shown in Figure 2 a: Figure 3 a depicts the FCM method with $c=6$ (since medical images typically contain 5–7 different tissues) based solely on the intensity characteristics using expression (4); Figure 3 b shows segmentation with an extended feature space as described in Section 1; Figure 3 c presents the proposed algorithm without Step 7; and Figure 3 d shows the result of the full proposed algorithm.

Figures 4 a and 4 b present the membership functions of two different classes of the membership function U , while Figures 4 c and 4 d show the first and second eigen images of the membership functions.

Figure 5 a shows the segmentation result of the X-ray image shown in Figure 2 c, and Figure 5 c presents the result for the optoelectronic metallographic image shown in Figure 5 b.

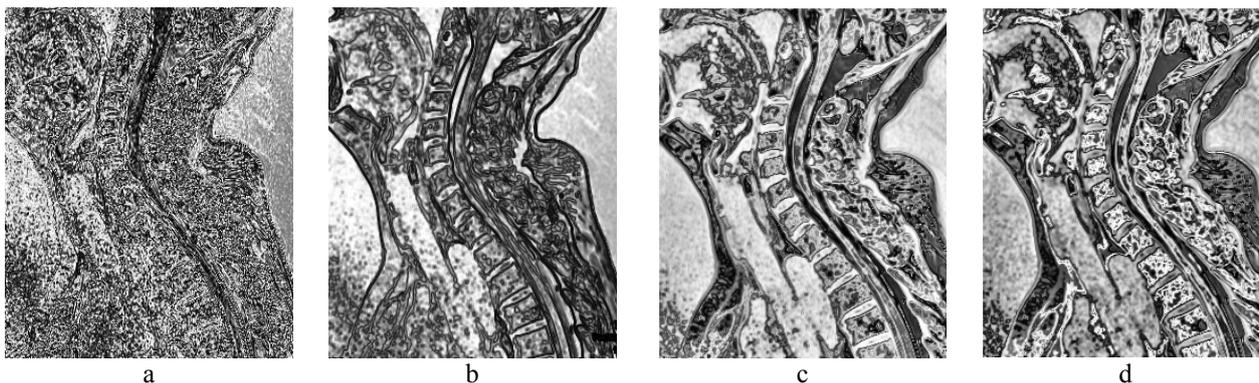


Figure 3 – Visualization of fuzzy clustering results for the MRI image (6 classes): a – original image; b – maximum membership function based on the original data; c – with an extended feature space (step 1); d – proposed method

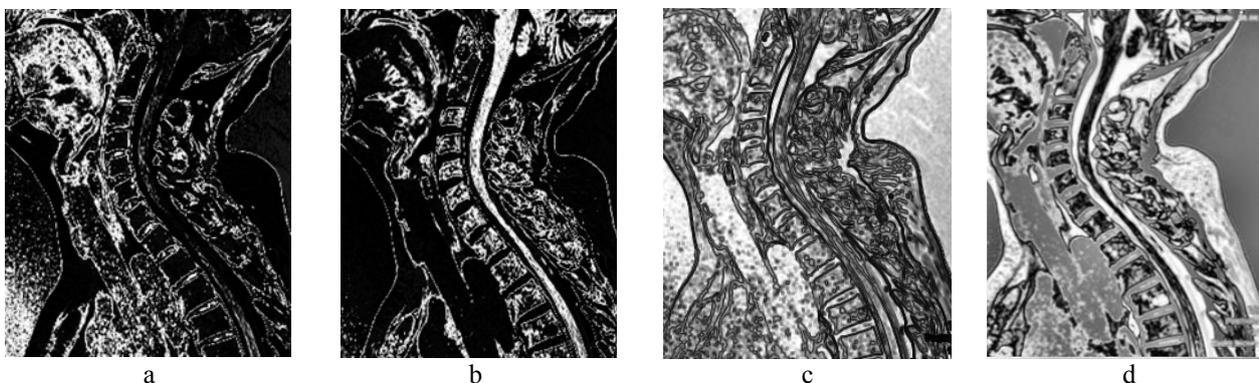


Figure 4 – Visualization of fuzzy clustering results for the MRI image into 6 classes: a, b – two arbitrary fuzzy classes; c, d – first and second eigen images of membership functions

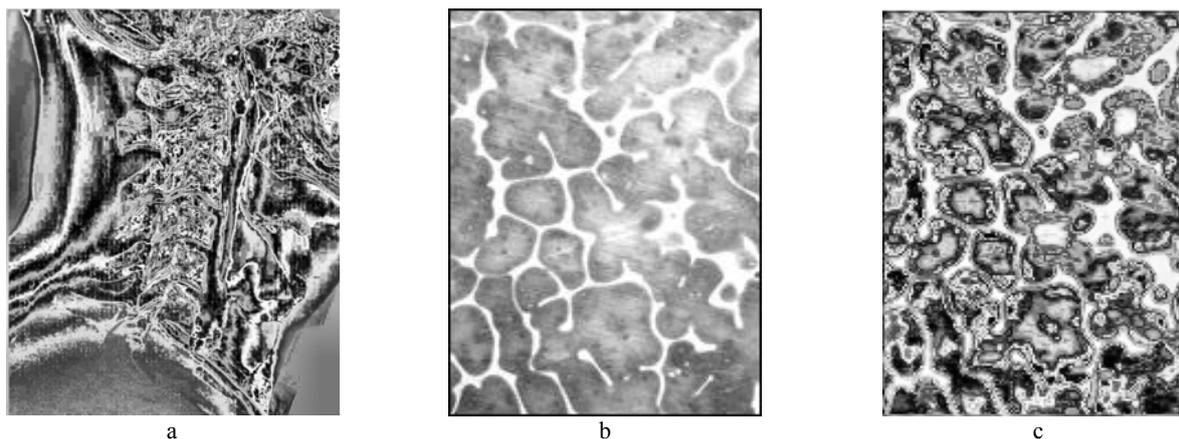


Figure 5 – Image segmentation: a – segmentation of the X-ray image from Figure 2 c; c – segmentation of the optoelectronic metallographic image from Figure 5 b

6 DISCUSSION

Fuzzy clustering is a sensitive tool, particularly responsive to the quality and composition of the original dataset. Visual analysis in Figure 3a shows that defuzzification by maximum membership function based solely on brightness forms a “noisy” image. Expanding the original brightness space per steps 1 and 2 significantly improves the result (Figure 3 b).

For example, for Figure 2 a, Table 1 presents the percentage of information H contained in each of the eigen images from the singular value decomposition of the membership functions. The first principal component carries the most information among all the obtained membership functions, while the second one contains the highest amount of information among the remaining components, and so on.

Table 1 – Eigenvalues of singular value decomposition (%)

	w_1	w_2	w_3	w_4	w_5	w_6
H_0	45.858	22.102	13.881	9.553	4.751	3.853
H_1	56.533	21.511	11.140	6.348	3.516	0.952

The analysis of eigenvalues shows increased informativeness for the first three principal components, from 81.841% to 89.184%, for cases where only the brightness of pixels (H_0) is used as the initial data and its conversion in accordance with step 1 (H_1).

In addition to the above, the information contained in the membership function of each class is largely lost during defuzzification based on expressions (4), (5) (Fig. 4 a, b).

The process of orthogonalizing membership functions enables conversion from peer cluster space to non-peer eigen image space. As seen in Figure 3 c, the use of the eigen space of membership functions from the singular value decomposition during defuzzification eliminates excessive detail and artifacts, enhances the clarity of object delineation, and simplifies the procedure for visually analyzing the results.

The automatic defuzzification method for the fuzzy clustering process, implemented by the proposed

algorithm, demonstrates a significant improvement in the sensitivity and informativeness of the synthesized image. It ensures clear delineation of the objects of interest and reveals the structure of various tissues compared to the original image.

CONCLUSIONS

A relevant problem has been solved: a sensitive algorithm for automatic segmentation of low-contrast grayscale images was developed based on type-2 fuzzy clustering.

The scientific novelty lies in the first-time proposal of a method for automatically forming segmented images in the eigen space of type-2 fuzzy clustering membership functions, enhancing accuracy and sensitivity.

The algorithm incorporates mechanisms for accounting for spatial components (expansion of the feature space based on window transformation), a clustering method using type-2 fuzziness, and result formation in the orthogonal space of membership functions. The singular value decomposition method was applied to the input data for fuzzy clustering, which were preliminarily subjected to fuzzy transformations and to the membership functions of the T2FCM method, with automatic determination of the final number of clusters, followed by the selection of the most significant components. The final result is formed based on the weighted sum of these selected components.

The practical significance of the results obtained lies in the fact that the proposed algorithm ensures improved accuracy and sensitivity in the segmentation of low-contrast grayscale images of various physical origins.

The prospects for further research involve studying the proposed algorithm’s application to segmenting various types of images, analyzing its informational potential when using alternative types of functional dependencies for forming type-2 fuzzy membership functions and other transformation’s types of fuzzy membership functions for result generation.

ACKNOWLEDGEMENTS

The work is supported by the state budget scientific research project of Dnipro National University “Increasing the efficiency of information processing in computer systems and networks” (state registration number 0122U001400) and “Intelligent data processing in computer information systems” (state registration number 0122U001397).

REFERENCES

1. Haussecker H., Tizhoosh H. R., Jahne B., H. Haussecker, P. Geissler. Fuzzy Image Processing: An Overview, *Handbook of Computer Vision and Applications, Vol. 2, Signal Processing and Pattern Recognition*, Boston, Academic Press, 1999. Section 22, P. 722. DOI:10.1016/B978-012379777-3/50017-0
2. Pratt W. K. Digital Image Processing. New York; Chichester; Weinheim; Brisbane; Singapore; Toronto, John Wiley & Sons, 2001, 723 p. DOI: 10.1002/0471221325
3. Russ J. C., Neal F. B. The Image Processing Handbook 7th Edition. Boca Raton, CRC Press, 2016, 1053 p. DOI: 10.1201/b18983
4. Chi Z., Yan H., Pham T. Editors: L. A. Zadeh et al. Fuzzy Algorithms: With Applications to Image Processing and Pattern Recognition, Vol. 10, *Advances in Fuzzy Systems – Applications and Theory*. Singapore; New Jersey; London; Hong Kong: World Scientific, 1996, 240 p. DOI: 10.1142/3132
5. Pham D. L., Xu C., Prince J. L. Current methods in medical image segmentation, *Annual Review of Biomedical Engineering*, 2000, Vol. 2, pp. 315–337. DOI: 10.1146/annurev.bioeng.2.1.315
6. Sokashe M. S. Computer assisted method for cervical vertebrae segmentation from x-ray images, *International Journal of Advanced Research in Computer and Communication Engineering*, 2013, Vol. 2, № 11, pp. 4387–4388.
7. Rutkovskiy L. Metody i tekhnologii iskusstvennogo intellekta. Moscow, Telekom, 2010, 600 p.
8. Bezdek J. C. Selected Applications in Classifier Design, *Pattern Recognition with Fuzzy Objective Function Algorithms*. New York; London, Plenum Press, 1981, Chap. 6, pp. 203–239. DOI: 10.1007/978-1-4757-0450-1
9. Yegorov A., Akhmetshina L. Optimizaciya yarkosti izobrazhenij na osnove nejro-fazzi tekhnologij. Monografiya. Dnipro, Lambert, 2015, 139 p.
10. A modified fuzzy clustering algorithm based on dynamic relatedness model for image segmentation / [X. Gao, Y. Zhang, H. Wang et al.] // *Visual Computing*. – 2023. – Vol. 39, P. 1583–1596. DOI: 10.1007/s00371-022-02430-4
11. Jaffar M. A., Naveed N., Ahmed B., Hussain A., Mirza A. M. Fuzzy c-means clustering with spatial information for color image segmentation, *International Conference on Electrical Engineering (ICEE): Third International Conference, Lahore, 09–11 April 2009: proceedings*. Lahore, IEEE, 2009, pp. 1–6. DOI: 10.1109/ICEE.2009.5173186
12. Zhang X., Sun Y., Liu H., Hou Z., Zhao F., Zhang C. Improved clustering algorithms for image segmentation based on non-local information and back projection, *Information Sciences*, 2021, Vol. 550, № 6, pp. 129–144. DOI: 10.1016/j.ins.2020.10.039
13. Li B. N., Chui C. K., Chang S., Ong S. H. Integrating spatial fuzzy clustering with level set methods for automated medical image segmentation, *Computers in Biology and Medicine*, 2011, Vol. 41, № 1, pp. 1–10. DOI: 10.1016/j.compbiomed.2010.10.007
14. Chuang K.-S., Tzeng H.-L., Chen S., Wu J., Chen T.-J. Fuzzy c-means clustering with spatial information for image segmentation, *Computerized Medical Imaging and Graphics*, 2006, Vol. 30, № 1, pp. 9–15. DOI: 10.1016/j.compmedimag.2005.10.001.
15. Pearson K. On lines and planes of closest fit to systems of points in space, *The London, Edinburgh and Dublin Philosophical Magazine and Journal of Sciences*, 1901, Series 6, Vol. 2, № 11, pp. 559–572. DOI: 10.1080/14786440109462720
16. Hotelling H. Analysis of a complex of statistical variables into principal components, *Journal of Educational Psychology*, 1933, Vol. 24, № 6, pp. 417–441. DOI: 10.1037/h0071325
17. Yang J., Zhang D., Frangi A. F., Yang J.-Y. Two-dimensional PCA: A new approach to appearance-based face representation and recognition, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2004, Vol. 26, № 1, pp. 131–137. DOI: 10.1109/TPAMI.2004.1261097
18. Nandi D., Ashour A., Chakraborty S., Dey N. et al. Principal Component Analysis in Medical Image Processing: A Study, *International Journal of Image Mining*, 2015, Vol. 1, № 1, pp. 65–86. DOI: 10.1504/IJIM.2015.070024
19. Aslam S., Rabie T. F. Principal Component Analysis in Image Classification: A review, *2023 Advances in Science and Engineering Technology International Conferences (ASET), Dubai, 20–23 February 2023: proceedings*. Dubai, IEEE, 2023, pp. 1–7. DOI: 10.1109/ASET56582.2023.10180847
20. Aneja D., Rawat T. K. Fuzzy Clustering Algorithms for Effective Medical Image Segmentation, *Intelligent Systems and Applications*, 2013, Vol. 11, № 6, pp. 55–61. DOI: 10.5815/ijisa.2013.11.06
21. Haussecker H., Tizhoosh H. R., Jahne B., Haussecker H., Geissler P. Fuzzy Image Processing: An Overview, *Handbook of Computer Vision and Applications, Vol. 2, Signal Processing and Pattern Recognition*. Boston, Academic Press, 1999, Section 22, pp. 683–727. DOI: 10.1016/B978-012379777-3/50017-0
22. Tizhoosh H. R., Kerre E. E., Nachttegael M. (eds.) Fuzzy Image Enhancement: An Overview, *Fuzzy Techniques in Image Processing*, Vol. 52, *Studies in Fuzziness and Soft Computing*. Heidelberg, Physica, 2000, Chap. 5, pp. 137–171. DOI: 10.1007/978-3-7908-1847-5_5
23. Wu H., Wu Y., Luo J. An Interval Type-2 Fuzzy Rough Set Model for Attribute Reduction, *IEEE Transactions on Fuzzy Systems*, 2009, Vol. 17, № 2, pp. 301–315. DOI: 10.1109/TFUZZ.2009.2013458.
24. Akhmetshina L. G., Yegorov A. A. Segmentaciya slabokontastnyh izobrazhenij v ortogonal'nom bazise funkcij prinalozhnosti nechetkoj klasterizacii, *Visnyk Hersons'kogo nacional'nogo tehnichnogo universytetu*, 2019, № 2(69), Part 2, pp. 160–165.
25. Akhmetshina L., Yegorov A. Low-Contrast Image Segmentation by Using of the Type-2 Fuzzy Clustering Based on the Membership Function Statistical Characteristics, *Intellectual Systems of Decision Making and Problems of Computational Intelligence (ISDMCI): XV International Scientific Conference, Kherson, 21–25 May, 2019: proceedings*. Kherson: Springer, 2019, pp. 689–700.

Received 21.11.2024.
Accepted 15.01.2025.

СЕГМЕНТАЦІЯ СЛАБКОКОНТРАСТНИХ ЗОБРАЖЕНЬ У БАЗИСІ ВЛАСНИХ ПІДПРОСТОРІВ НЕЧІТКИХ ФУНКЦІЙ НАЛЕЖНОСТІ ТИПУ-2

Ахметшина Л. Г. – д-р техн. наук, професор кафедри електронних обчислювальних машин Дніпровського національного університету ім. О. Гончара, Дніпро, Україна.

Єгоров А. О. – канд. техн. наук, ст. викл. кафедри комп'ютерних наук та інформаційних технологій Дніпровського національного університету ім. О. Гончара, Дніпро, Україна.

Фомін А. А. – аспірант кафедри електронних обчислювальних машин Дніпровського національного університету ім. О. Гончара, Дніпро, Україна.

АНОТАЦІЯ

Актуальність. Розглянуто актуальне завдання автоматизації чутливого алгоритму сегментації зображень на основі методу нечіткої кластеризації типу-2. Об'єктом дослідження є слабкоконтрастні напівтонові зображення, які є результатом стандартних методів дослідження в різних галузях діяльності людини.

Мета роботи – створення нового набору інформативних ознак на основі вихідних даних, виконання чутливої нечіткої сегментації на основі метода кластеризації з використанням нечіткості 2-го порядку, реалізація автоматичної дефазифікації у власному підпросторі функцій приналежності.

Метод. Запропоновано метод сегментації слабкоконтрастних зображень, який складається з наступних кроків: розширення простору ознак вхідних даних, застосування сингулярного розкладу до розширеного набору даних з наступним автоматичним відбором найбільш значущих компонентів, що є вхідними даними для нечіткої кластеризації з використанням нечітких множин типу-2. Ця кластеризація відбувається за допомогою методу T2FCM, який дозволяє автоматично підбирати кількість нечітких кластерів на основі початкового задання гарантовано більшої кількості з наступним злиттям близьких кластерів (в роботі близькість визначалась на основі зваженої Евклідової відстані). Після виконання нечіткої кластеризації в запропонованому методі здійснюється об'єднання її результатів (нечіткої функції належності) з вхідними для нечіткої кластеризації даними, які попередньо оброблюються нечітким перетворенням. Результуюча матриця знов підлягає нечіткому перетворенню, після чого до отриманих результатів застосовується сингулярний розклад з наступним автоматичним відбором найбільш значущих компонентів. На основі зваженої суми цих відібраних компонентів формується напівтонове зображення, до якого застосовується метод адаптивної еквалізації гістограми, в результаті чого і отримується кінцевий результат сегментації. Запропонований метод сегментації має невелику кількість керуючих параметрів: початкову кількість нечітких кластерів, помилку методу T2FCM та максимальну кількість ітерацій, а також коефіцієнт використаних нечітких перетворень, причому їх налаштування на зображення, що оброблюються, не вимагають значних зусиль.

Результати. Розроблений алгоритм реалізовано програмно, проведено експерименти на реальних зображеннях різної фізичної природи.

Висновки. Проведені експерименти підтвердили працездатність запропонованого алгоритму та дозволяють рекомендувати його для використання на практиці при вирішенні задач візуального аналізу слабкоконтрастних напівтонових зображень. Перспективи подальших досліджень можуть полягати в аналізі інформативних можливостей алгоритму при використанні інших типів перетворень нечітких функцій належності та в модифікації запропонованого алгоритму для сегментації зображень різного типу.

КЛЮЧОВІ СЛОВА: сегментація зображень, нечітка кластеризація, нечітка кластеризація типу_2, ортогональні перетворення, сингулярне розкладання, власні підпростори.

ЛІТЕРАТУРА

1. Haussecker H. Fuzzy Image Processing: An Overview / H. Haussecker, H. R. Tizhoosh // Handbook of Computer Vision and Applications. – Vol. 2: Signal Processing and Pattern Recognition / B. Jahne, H. Haussecker, P. Geissler. – Boston: Academic Press, 1999. – Section 22. – P. 722. DOI:10.1016/B978-012379777-3/50017-0
2. Pratt W. K. Digital Image Processing / W. K. Pratt. – New York; Chichester; Weinheim; Brisbane; Singapore; Toronto: John Wiley & Sons, 2001. – 723 p. DOI: 10.1002/0471221325
3. Russ J. C. The Image Processing Handbook 7th Edition / J. C. Russ, F. B. Neal. – Boca Raton : CRC Press, 2016. – 1053 p. DOI: 10.1201/b18983
4. Chi Z. Fuzzy Algorithms: With Applications to Image Processing and Pattern Recognition / Z. Chi, H. Yan, T. Pham – Vol. 10: Advances in Fuzzy Systems – Applications and Theory / [Editors: L. A. Zadeh et al]. – Singapore; New Jersey; London; Hong Kong: World Scientific, 1996. – 240 p. DOI: 10.1142/3132
5. Pham D. L. Current methods in medical image segmentation / D. L. Pham, C. Xu, J. L. Prince // Annual Review of Bio-
medical Engineering. – 2000. – Vol. 2. – P. 315–337. DOI: 10.1146/annurev.bioeng.2.1.315
6. Sokashe M. S. Computer assisted method for cervical vertebrae segmentation from x-ray images / S. V. Sokashe // International Journal of Advanced Research in Computer and Communication Engineering. – 2013. – Vol. 2, № 11. – P. 4387–4388.
7. Rutkovskiy L. Metody i tekhnologii iskusstvennogo intellekta / L. Rutkovskiy. – Moscow : Telekom, 2010. – 600 p.
8. Bezdek J. C. Selected Applications in Classifier Design / J. C. Bezdek // Pattern Recognition with Fuzzy Objective Function Algorithms / J. C. Bezdek – New York; London: Plenum Press, 1981. – Chap. 6. – P. 203–239. DOI: 10.1007/978-1-4757-0450-1
9. Yegorov A. Optimizaciya yarkosti izobrazhenij na osnove nejro-fazzi tehnologij. Monografiya / A. Yegorov, L. Akhmetshina. – Dnipro: Lambert, 2015. – 139 p.
10. A modified fuzzy clustering algorithm based on dynamic relatedness model for image segmentation / [X. Gao, Y. Zhang, H. Wang et al.] // Visual Computing. – 2023. – Vol. 39, P. 1583–1596. DOI: 10.1007/s00371-022-02430-4

11. Fuzzy c-means clustering with spatial information for color image segmentation / [M. A. Jaffar, N. Naveed, B. Ahmed et al.] // International Conference on Electrical Engineering (ICEE): Third International Conference, Lahore, 09–11 April 2009; proceedings. – Lahore: IEEE, 2009. – P. 1–6. DOI: 10.1109/ICEE.2009.5173186
12. Improved clustering algorithms for image segmentation based on non-local information and back projection / [X. Zhang, Y. Sun, H. Liu et al.] // Information Sciences. – 2021. – Vol. 550, № 6. – P. 129–144. DOI: 10.1016/j.ins.2020.10.039
13. Integrating spatial fuzzy clustering with level set methods for automated medical image segmentation / [B. N. Li, C. K. Chui, S. Chang, S. H. Ong] // Computers in Biology and Medicine. – 2011. – Vol. 41, № 1. – P. 1–10. DOI: 10.1016/j.compbiomed.2010.10.007
14. Fuzzy c-means clustering with spatial information for image segmentation / [K. -S. Chuang, H.-L. Tzeng, S. Chen, J. Wu, T. -J. Chen] // Computerized Medical Imaging and Graphics. – 2006. – Vol. 30, № 1. – P. 9–15. – DOI: 10.1016/j.compmedimag.2005.10.001.
15. Pearson K. On lines and planes of closest fit to systems of points in space / K. Pearson // The London, Edinburgh and Dublin Philosophical Magazine and Journal of Sciences. – 1901. – Series 6, Vol. 2, № 11. – P. 559–572. DOI: 10.1080/14786440109462720
16. Hotelling H. Analysis of a complex of statistical variables into principal components. / H. Hotelling // Journal of Educational Psychology. – 1933. – Vol. 24, № 6. – P. 417–441. DOI: 10.1037/h0071325
17. Two-dimensional PCA: A new approach to appearance-based face representation and recognition / [J. Yang, D. Zhang, A. F. Frangi, J. -Y. Yang] // IEEE Transactions on Pattern Analysis and Machine Intelligence. – 2004. – Vol. 26, № 1. – P. 131–137. DOI: 10.1109/TPAMI.2004.1261097
18. Principal Component Analysis in Medical Image Processing: A Study / [D. Nandi, A. Ashour, S. Chakraborty et al.] // International Journal of Image Mining. – 2015. – Vol. 1, № 1. – P. 65–86. DOI: 10.1504/IJIM.2015.070024
19. Aslam S. Principal Component Analysis in Image Classification: A review / S. Aslam, T. F. Rabie // 2023 Advances in Science and Engineering Technology International Conferences (ASET), Dubai, 20–23 February 2023: proceedings. – Dubai: IEEE, 2023. – P. 1–7. DOI: 10.1109/ASET56582.2023.10180847
20. Aneja D. Fuzzy Clustering Algorithms for Effective Medical Image Segmentation / D. Aneja, T. K. Rawat // Intelligent Systems and Applications. – 2013. – Vol. 11, № 6. – P. 55–61. DOI: 10.5815/ijisa.2013.11.06
21. Haussecker H. Fuzzy Image Processing: An Overview / H. Haussecker, H. R. Tizhoosh // Handbook of Computer Vision and Applications. – Vol. 2: Signal Processing and Pattern Recognition / B. Jahne, H. Haussecker, P. Geissler. – Boston: Academic Press, 1999. – Section 22. – P. 683–727. DOI: 10.1016/B978-012379777-3/50017-0
22. Tizhoosh H. R. Fuzzy Image Enhancement: An Overview / H. R. Tizhoosh // Fuzzy Techniques in Image Processing. – Vol. 52: Studies in Fuzziness and Soft Computing / E. E. Kerre, M. Nachtgael (eds.). – Heidelberg: Physica, 2000. – Chap. 5 – P. 137–171. DOI: 10.1007/978-3-7908-1847-5_5
23. Wu H. An Interval Type-2 Fuzzy Rough Set Model for Attribute Reduction / H. Wu, Y. Wu, J. Luo // IEEE Transactions on Fuzzy Systems. – 2009. – Vol. 17, № 2. – P. 301–315. – DOI: 10.1109/TFUZZ.2009.2013458.
24. Akhmetshina L. G. Segmentaciya slabokontrastnyh izobrazhenij v ortogonal'nom bazise funkcij prinadlezhnosti nechetkoj klasterizacii / L. G. Akhmetshina, A. A. Yegorov // Visnyk Hersons'kogo nacional'nogo tehničnogo universytetu. – 2019. – № 2(69), Part 2. – P. 160–165.
25. Akhmetshina L. Low-Contrast Image Segmentation by Using of the Type-2 Fuzzy Clustering Based on the Membership Function Statistical Characteristics / L. Akhmetshina, A. Yegorov // Intellectual Systems of Decision Making and Problems of Computational Intelligence (ISDMCI): XV International Scientific Conference, Kherson, 21–25 May, 2019: proceedings. – Kherson: Springer, 2019. – P. 689–700.

АНАЛІЗ ПІДХОДІВ ДОСТУПУ ДО ДАНИХ У МУЛЬТИ-КЛАУД СЕРЕДОВИЩІ

Касерес А. – аспірант Національного технічного університету України «Київський політехнічний інститут імені Ігоря Сікорського», кафедра інформаційних технологій в телекомунікаціях, Київ, Україна.

Глоба Л. С. – професор, д-р техн. наук, кафедра інформаційних технологій в телекомунікаціях інституту телекомунікаційних систем національного технічного університету України «Київський політехнічний інститут імені Ігоря Сікорського», Київ, Україна.

АНОТАЦІЯ

Актуальність. Мультихмарна система характеризується послідовним або одночасним використанням послуг від різних хмарних постачальників для виконання програм. Така система є бажаною інфраструктурою для переважної більшості ІТ-бізнесу сьогодні. Наразі існують різноманітні підходи для об'єднання хмарних платформ кількох постачальників. У даній статті досліджувались практичні підходи для досягнення мультихмарної сумісності, зосереджуючись на абстрактному доступі до даних між різними постачальниками cloud-сховищ та multi-cloud розподілі обчислювальних ресурсів. Представлено ключові технології та методології безперебійного керування даними, такі як використання мультихмарних шлюзів зберігання даних (на прикладі S3Proху), впровадження платформ керування даними (Apache NiFi) та використання універсальних хмарних бібліотек (Apache Libcloud). У роботі висвітлено переваги і недоліки обраних підходів і проведено експерименти по визначенню вартості і продуктивності для них. Результатом проведених досліджень є визначення вартості і продуктивності для різних підходів доступу до даних у мультихмарних середовищах.

Мета. Дослідити різні підходи мультихмарного доступу до даних і визначити найбільш оптимальний за характеристиками продуктивності і вартості.

Метод. Запропоновано оптимізацію мультихмарних інфраструктур на основі експериментальних даних. Експериментальне моделювання включає в себе емпіричні вимірювання швидкодії і порівняння витрат на зберігання. Визначення продуктивності базується на вимірюванні часу читання даних і затримки. Для оцінки вартості використовується модель ціноутворення AWS S3. Описано підходи оптимізації з урахуванням розмірів файлів і обсягом зберігання даних: об'єднання сильних сторін різних мультихмарних підходів і динамічне перемикання між рішеннями. Запропоновано алгоритм вибору мультихмарних підходів, де враховані критерії вартості і швидкодії, а також їх пріоритетності.

Результати. У ході експерименту отримано значення вартості, необхідної для зберігання і завантаження даних різних обсягів (100 ГБ, 1 ТБ, 10 ТБ), а також продуктивності для передавання файлів різного розміру (100 КБ, 1 МБ, 10 МБ) для технологій мультихмарних шлюзів, платформ управління даними, та хмарно-незалежних бібліотек. S3Proху має найшвидший доступ до файлів для великих обсягів даних. Apache Libcloud показує кращу вартість на невеликих об'ємах. Проте ці обидва підходи значно переважають Apache NiFi. Дане дослідження може сприяти розвитку методів ефективного керування ресурсами у мультихмарних середовищах.

Висновки. Отримані результати дають можливість визначити пріоритетність вибору означених парадигм, щоб допомогти організаціям розробити та розгорнути ефективні мультихмарні стратегії, які дозволять їм скористатися перевагами унікальних функцій кожного хмарного провайдера, зберігаючи при цьому уніфіковане, гнучке та ефективне середовище зберігання та обчислення.

КЛЮЧОВІ СЛОВА: хмарні обчислення, мультихмарний доступ, універсальні хмарні бібліотеки.

АБРЕВІАТУРИ

DMP – Data management platform.

НОМЕНКЛАТУРА

A – множина мультихмарних підходів;

A_{opt} – оптимальний мультихмарний підхід;

C_i – вартість використання хмарного сервісу i -го провайдера;

C_{min} – мінімальна вартість використання хмарного сервісу;

CC – критерій вартості;

CT – критерій часу доступу до файлу;

FS – розмір файлу;

K – кількість цінових факторів;

M – кількість передавань файлу;

N – кількість досліджуваних мультихмарних стратегій;

P_i – продуктивність хмарного сервісу i -го провайдера;

P_{max} – максимальна продуктивність хмарного сервісу;

T – середній час доступу до файлу;

V – об'єм зберігання.

ВСТУП

Проблему об'єднання хмарних платформ кількох вендорів (постачальників послуг) у вигляді єдиного простору зберігання та обробки даних можна зустріти у різних дослідженнях [1–3]. Глобальне впровадження мультихмарних стратегій зростає, і організації всіх розмірів та галузей все більше використовують переваги декількох хмарних провайдерів. Опитування Gartner [4] показує, що 81% користувачів публічних хмар працюють з двома чи більше cloud-провайдерами. Flexera 2024 State of the Cloud Report [5] підтверджує це, показуючи, що 89% усіх організацій використовують мультихмарні технології (рис. 1).



Рисунок 1 – Стан використання хмарних технологій

Оскільки все більше компаній приймають цю тенденцію, проблеми, пов'язані з керуванням доступом до даних у різних постачальників хмарних сховищ, стають дедалі очевиднішими. Це потребує повного розуміння різноманітних підходів, які використовують для абстрактного доступу до даних у мультихмарних середовищах. Зазвичай величезні обсяги даних зберігаються в одній хмарі, а їх обробка відбувається в іншій, а результати обробки можуть бути збережені ще на одній cloud-платформі [3].

Управління доступом до даних у різних постачальників хмарних сховищ і розподіл обчислювальних ресурсів між різними хмарами може становити значні труднощі. Через відсутність стандартних інтерфейсів для досягнення мультихмарної сумісності нерідко інтеграцію потрібно щоразу розробляти вручну [6]. Щоб використати потенціал мультихмарних сховищ і обчислень, організації повинні визначити найкращий підхід до абстрактного доступу до даних у різних хмарних провайдерів, зберігаючи при цьому уніфіковане та ефективне середовище.

Об'єкт дослідження – мультихмарні стратегії та їхнє глобальне впровадження.

Предмет дослідження – методи та проблеми управління доступом до даних у мультихмарних середовищах, а також порівняння різних підходів до абстрактного доступу до даних з точки зору вартості та продуктивності.

© Касерес А., Глоба Л. С., 2025
DOI 10.15588/1607-3274-2025-1-16

Метою проведених досліджень є постановка та проведення експерименту, який порівнює різні підходи мультихмарного доступу до даних на основі вартості і продуктивності. Оцінюючи та ранжуючи ці підходи, організації можуть приймати обґрунтовані рішення, які оптимізують їх мультихмарне розгортання та максимізують переваги, отримані від використання кількох хмарних постачальників.

На сьогоднішній день широко застосовують такі технології доступу до даних у мультихмарних середовищах, використовуючи верхні рівні абстракції:

- Multi-cloud шлюзи (Multi-cloud storage gateways);
- платформи керування даними (Data management platforms або DMPs);
- брокери хмарних сервісів (Cloud Service Brokers);
- універсальні хмарні бібліотеки (Cloud-agnostic libraries)
- контейнеризація (Containerization);
- оркестрація контейнерів (Container Orchestration).

1 ПОСТАНОВКА ЗАДАЧІ

Для визначення ефективності технології доступу до даних у мультихмарних середовищах, використовуючи верхні рівні абстракції моделі OSI, опишемо вхідні і вихідні параметри.

Дано: у якості вхідних параметрів виступають підходи абстрактного доступу до даних у мультихмарних середовищах $A = \{A_1, A_2, \dots, A_N\}$.

Для оцінки та порівняння різних підходів будуть використані наступні критерії:

- вартість C_i використання хмарного сервісу i -го провайдера, де $1 \leq i \leq N$.
- продуктивність (час відгуку) P_i хмарного сервісу i -го провайдера, де $1 \leq i \leq N$.

Критерії та обмеження оцінки якості результатів:

- для вартості важлива мінімізація значення $C \rightarrow \min$, при цьому $C = \{C_1, C_2, \dots, C_N\}$;
- для продуктивності важлива максимізація значення $P \rightarrow \max$, $P = \{P_1, P_2, \dots, P_N\}$.

Знайти: дане дослідження має на меті отримати мінімальну вартість і максимальну продуктивність.

Продуктивність визначимо як відношення обсягу даних до часу обробки (1):

$$P = \frac{FS}{T} \quad (1)$$

Вартість визначається як сума складових вартості (2):

$$C = \sum_i C_i \quad (2)$$

2 ОГЛЯД ЛІТЕРАТУРИ

Існує багато способів використання ресурсів різних хмар. У [7–9] наведено огляд мультихмарних технологій та прототипів, серед яких використання мультихмарних платформ, шлюзів, прототипи для комплексної обробки подій в хмарному середовищі, розробки для обробки комерційних хмарних симуляцій, фреймворки, відкриті бібліотеки. Найбільш ефективним методом для забезпечення взаємодії є надання гнучких стандартів для роботи у хмарних сервісах.

Зусилля зі стандартизації мультихмарного доступу стосуються ініціатив галузевих консорціумів, організацій зі стандартизації та відкритих спільнот для створення спільного набору специфікацій і рекомендацій, які сприяють сумісності доступу до даних у кількох хмарах.

Деякі зусилля зі стандартизації в мультихмарному просторі доступу до даних включають:

- Інтерфейс керування хмарними даними (CDMI) [10], розроблений Асоціацією індустрії мереж зберігання даних (SNIA), є стандартом, який визначає функціональний інтерфейс для доступу та керування даними, що зберігаються в хмарі. CDMI дозволяє користувачам створювати, отримувати, оновлювати та видаляти елементи даних із хмари, а також виявляти базову інфраструктуру хмарного зберігання й керувати нею.

Позитивним стороною даної технології можна назвати відсутність прив'язки до одного хмарного постачальника, що дає можливість переміщувати дані між хмарами, також являється відкритим стандартом і підтримується багатьма провайдерами.

Разом з тим, стандарт може негативно вплинути на продуктивність в окремих випадках, а також вимагати додаткових зусиль інтеграції з існуючими системами управління даними.

- Відкритий інтерфейс хмарних обчислень (OCCI) [11] – це набір специфікацій, які визначають API загального призначення для керування різноманітними ресурсами хмарних обчислень, включаючи служби зберігання, обчислення та мережеві служби. OCCI прагне сприяти взаємодії між різними хмарними провайдерами та сприяти розробці портативних хмарних програм.

Цей інтерфейс спрощує інтеграцію та управління хмарними ресурсами, може адаптуватися до різних потреб користувачів і технологій.

Водночас, за рахунок модульної структури може призвести до складності у виборі конфігурації під конкретні потреби.

Таким чином, описані стандарти мають деякі спільні переваги, такі як спрощення інтеграції та управління хмарними ресурсами, дозволяє використовувати різні хмарні платформи без прив'язки до одного вендора. Основні недоліки визначаються як потенційні проблеми з продуктивністю і складність впровадження.

Хоча стандартизацію часто вважають найбільш перспективним способом забезпечення взаємодії з кількома хмарами, впровадження стандартів відбува-

ється повільно, оскільки постачальники хмарних послуг зацікавлені в просуванні своїх власних API і технологій.

Таким чином, відсутність загально визначених стандартів вимагає дослідження альтернативних рішень щодо інтеоперабельності [9].

Для подальшого дослідження будуть розглянуті наступні парадигми мультихмарного доступу до даних:

- multi-cloud шлюзи;
- платформи керування даними;
- хмарно-незалежні бібліотеки.

1. Multi-cloud шлюзи – це апаратне або програмне забезпечення, яке надає плавний доступ до даних, що зберігаються в кількох хмарних провайдерах, діючи як посередник між програмами користувача та службами хмарного зберігання [12].

Мультихмарні шлюзи сховищ працюють, перетворюючи API-інтерфейси різних хмарних провайдерів в єдиний API, полегшуючи переміщення даних і управління ними. Вони дозволяють користувачам отримувати доступ і керувати даними в різних хмарних провайдерах без необхідності впровадження та підтримки кількох інтеграцій API. Спрощена архітектура доступу до даних через шлюз зберігання проілюстрована на рис. 2.

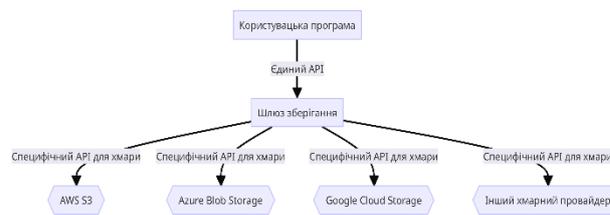


Рисунок 1 – Спрощена схема шлюзу мультихмарного сховища

Доступ до хмарного сховища через шлюз зазвичай має такі переваги:

- Спрощене керування даними за допомогою єдиного уніфікованого API.
- Унікає блокування постачальника та забезпечує стійкість даних шляхом розподілу даних між кількома хмарними провайдерами.
- Потенційно покращена продуктивність доступу до даних за допомогою кешування та методів оптимізації.

При цьому до недоліків можна віднести:

- Додаткові витрати за користування шлюзом.
- Зниження продуктивності або збільшення затримки через додатковий рівень абстракції.
- Залежність від постачальника шлюзу щодо оновлень, виправлення помилок і підтримки.

Розробками у цій сфері є: Cloudbian Hyper-Store, Nasuni Cloud File Services, Panzura Freedom, Morro Data CloudNAS, S3Proху, які оптимізують використання сховища та підвищують безпеку, одночасно забезпечуючи мультихмарний доступ до даних.

2. Платформи керування даними – це програмні рішення, які, на відміну від шлюзів зберігання, призначені для полегшення не лише керування даними та доступу до них у мультихмарних середовищах, але й міграції, захисту та керування даними між різними постачальниками хмарних сховищ. Ці платформи абстрагують базові служби хмарного зберігання та забезпечують централізовану площину

управління даними, що дозволяє організаціям запроваджувати політики, автоматизувати процеси та отримувати аналіз своїх даних у кількох хмарах [13].

Архітектура та потік доступу DMP дуже схожі на шлюзи зберігання (рис. 2). Замість цього пропонується зосередитися на ключових відмінностях для кінцевого користувача.

Діаграма послідовності дій показана на рис. 3.

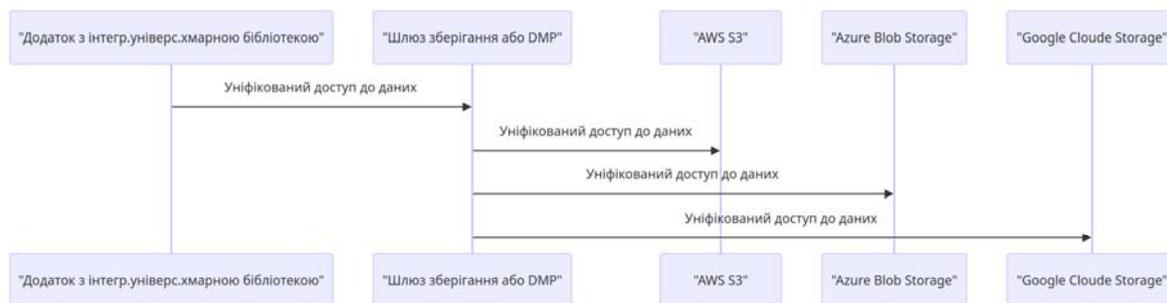


Рисунок 3 – Діаграма послідовності дій щодо інтеграції хмарно-незалежної бібліотеки в кінцевий додаток

Переваги платформ управління даними:

- Оптимізують керування даними, обробку та аналітику в багатьох хмарних провайдерах, пропонуючи більш розширені функції, ніж шлюзи зберігання.
- Покращують управління даними та відповідність вимогам, забезпечуючи централізований контроль і видимість даних із різних джерел, у тому числі мультихмарних служб зберігання.
- Підтримують розширену обробку даних і можливість аналітики, які зазвичай недоступні в шлюзах зберігання.

Недоліки платформ управління даними:

- Можуть бути складнішими у впровадженні та обслуговуванні завдяки додатковим функціям і функціям, які надає платформа.
- Можуть призвести до підвищення вартості використання платформи.
- Деякі платформи можуть не підтримувати всіх хмарних постачальників або певні служби зберігання, що може призвести до потенційних проблем із сумісністю або обмежень.

Серед популярних прикладів DMPs можна назвати Talend Data Fabric, Informatica Intelligent Cloud Services, Dell Boomi AtomSphere, SnapLogic Intelligent Integration Platform, а також рішення з відкритим кодом Apache Nifi.

3. Універсальні хмарні бібліотеки – це програмні інструменти, які забезпечують єдиний уніфікований інтерфейс для взаємодії з декількома службами хмарного сховища на рівні програмування додатків, усуваючи необхідність працювати з кількома спеціальними API для cloud-середовища. Ці бібліотеки часто розробляються та підтримуються спільнотою учасників, що робить їх ініціативами з відкритим вихідним кодом і дозволяє будь-яким розробникам додавати відсутні функції та налаштовувати інструмент відповідно до потреб власної програми. Однак, оскільки ці

бібліотеки потрібно впроваджувати в кінцеву програму на набагато нижчому рівні, вони можуть потребувати більше зусиль у розробці порівняно з попередніми двома рішеннями.

Приклади універсальних хмарних бібліотек і фреймворків включають Apache Libcloud і JClouds. Вони пропонують розробникам можливість взаємодії з широким колом постачальників хмарних сховищ через єдиний API. Хоча обидві бібліотеки схожі за своєю здатністю надавати інтерфейси, що не залежать від хмари, вони відрізняються підходом до інтеграції хмарного сховища [14].

Використання універсальних хмарних бібліотек додає значних зусиль для реалізації кінцевого додатку. Абстрагуючи деталі доступу до даних, специфічні для хмари, це вводить потребу в реалізації власного спеціального API, який зазвичай більш обмежений і менш документований, ніж API хмарних провайдерів. Проілюструємо процес інтеграції хмарно-незалежної бібліотеки за допомогою діаграм 3 і 4.

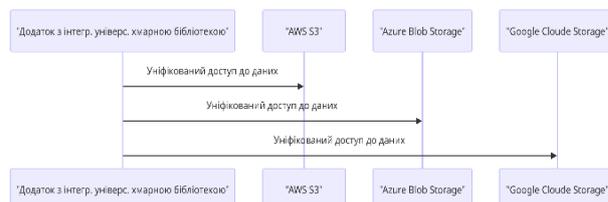


Рисунок 4 – Діаграма послідовності дій щодо інтеграції хмарно-незалежної бібліотеки в кінцевий додаток

Переваги універсальних хмарних бібліотек включають:

- Рентабельність, оскільки більшість рішень з відкритим кодом можна використовувати безкоштовно.
- Розвинуті засоби налаштування, що дозволяють необмежену адаптацію або розширення.

– Швидше виправлення помилок завдяки розробці, керованій спільнотою.

В якості недоліків можна відзначити:

– Зазвичай вимагає більше технічних знань і зусиль для впровадження та підтримки.

– Підтримка та документація можуть бути обмежені, залежно від популярності проєкту та участі спільноти.

– Потенційні проблеми сумісності або обмеження певних хмарних постачальників або служб зберігання [15].

Для аналізу кожного з описаних вище підходів, необхідно провести експериментальні дослідження, у ході яких буде отримано значення показників вартості і продуктивності для подальшої оцінки ефективності впровадження і доцільності застосування кожного зі способів в конкретних умовах.

Для проведення експерименту обрано рішення, які є репрезентативними для кожної парадигми. Визначені технології характеризуються своєю доступністю, тобто мають ліцензію для некомерційного використання, а також стабільну підтримку:

– S3Proху [16] – репрезентує мультихмарні шлюзи. Це серверний компонент, який служить як посередник між додатком та об'єктом сховища.

– Apache NiFi [17] – представник платформ управління даними, розроблений Apache Software Foundation. Це система для візуального програмування потоків даних між різними системами, яка дозволяє легко налаштовувати, керувати та проводити моніторинг потоків даних, включаючи інтеграцію з різними джерелами даних, їх обробку та передачу в різні системи або сховища.

– Apache Libcloud [18] – представник хмарних бібліотек та ініціатив з відкритим кодом. Це бібліотека Python, яка пропонує уніфікований API для взаємодії з широким колом хмарних провайдерів. Забезпечує високий рівень абстракції, який спрощує роботу з різними службами хмарного сховища, що робить його придатним для розробників, які віддають перевагу одній мові програмування та потребують широкого спектру підтримки постачальників.

3 МАТЕРІАЛИ ТА МЕТОДИ

У експерименті буде виконано порівняння параметрів вартості C і продуктивності P [19].

Параметр вартості в даному експерименті спрямований на оцінку фінансових наслідків, пов'язаних із впровадженням і експлуатацією кожного рішення – S3Proху, Apache NiFi та Apache Libcloud. Усі розрахунки вартості в цьому експерименті базуються на моделях ціноутворення AWS S3 [20] станом на 1 червня 2024 року.

Розглянемо прямі витрати, пов'язані з функціонуванням цих рішень, дозволяючи користувачам приймати обґрунтоване рішення з урахуванням їхніх конкретних моделей використання та бюджетних обмежень.

Аналіз продуктивності має на меті оцінити можливість доступу до даних, які пропонує кожне рішення, і зосереджений на вимірюванні часу читання даних і затримки.

Для проведення цього експерименту зосереджуємось на таких складових вартості:

– Витрати на зберігання. Розраховуються на основі обсягу даних, що зберігаються в хмарі. Ми розглядаємо різні обсяги даних (100 ГБ, 1 ТБ і 10 ТБ).

– Витрати на екземпляр (віртуальна машина). Ці витрати пов'язані із запуском екземплярів віртуальних машин на хмарних платформах, зокрема на AWS EC2 для даного експерименту. Розглядаємо розміри екземплярів віртуальних машин, які можуть обробляти обсяги даних, що використовуються в тестах.

– Витрати на запит. Такі витрати оцінюють фінансовий вплив кількості запитів PUT, GET і LIST, зроблених під час операцій доступу до даних. Ці витрати зазвичай залежать від кількості запитів, зроблених на місяць.

Усі складові вартості були взяті з офіційного сайту провайдера AWS [20].

Таким чином, загальна вартість для i -го підходу буде розрахована за (3):

$$C_i = \sum_{j=1}^k C_j. \quad (3)$$

Оскільки всі рішення, з якими проводяться експерименти, є відкритими, ліцензійні витрати не стягуються.

Щоб оцінити продуктивність, файли різних розмірів використано для перевірки часу доступу:

- 100 КБ;
- 1 МБ;
- 10 МБ.

Кожен файл отримано 20 разів для кожного сценарію, щоб отримати найбільш достовірне середнє значення. Для узгодженості використано стабільне кабелне з'єднання зі швидкістю 100 Мбіт/с.

Середній час доступу до файлу для i -го підходу буде розраховано за (4):

$$T_i = \frac{\sum_{j=1}^M T_j}{M}. \quad (4)$$

Наступні кроки описують процес вимірювання продуктивності:

– Провести порівняльний аналіз для кожного рішення, щоб виміряти час, необхідний для читання даних від різних постачальників хмарних сховищ.

– Оцінити затримку для кожного рішення, визначивши час, витрачений на обробку та повернення запитів.

– Проаналізувати пропускну здатність кожного рішення, вимірюючи обсяг даних, до яких можна отримати доступ і передати їх за одиницю часу.

Продуктивність у нашому випадку – це швидкість обробки даних, тобто, чим менше значення середнього часу доступу до файлу, тим краще.

Таким чином, після проведення вимірювань для кожного підходу будуть визначені вартість і швидкість доступу до файлів.

4 ЕКСПЕРИМЕНТИ

Для проведення експериментів програми було встановлено у ізолюванні докер-контейнери із наступним обмеженням ресурсів:

- 2 Virtual CPU;
- 2 GB RAM;
- 2 GB SWAP;
- 16 GB Virtual disk.

1. Експеримент S3Proху

Для проведення експерименту з S3 Proху було використано офіційний Docker-образ [21]. Для вимірювання продуктивності було написано прикладну програму на Python, доступну у відкритому доступі [22].

Експеримент, проведений із застосуванням технології S3Proху, дає уявлення про його можливості як мультимарного рішення для доступу до даних (табл. 1).

Таблиця 1 – Швидкість доступу до файлів S3Proху

Розмір файлу	Середній час доступу, с
100 КБ	0,0732
1 МБ	0,1055
10 МБ	0,3652

Продуктивність S3Proху оцінювалася шляхом вимірювання середнього часу доступу до файлів різного розміру. При стабільному кабельному з'єднанні на 100 Мбіт/с кожен файл було завантажено 20 разів та визначено середнє значення.

Виявлено, що середній час доступу до файлу розміром 100 КБ становить 0,0732 секунди. Середній час доступу дещо збільшився до 0,1055 секунди для файлу розміром 1 МБ і суттєво збільшився до 0,3652 секунди для файлу розміром 10 МБ.

У нашому дослідженні рішення S3Proху фактор вартості оцінювався на основі компонентів, поданих раніше: витрати на зберігання, витрати на екземпляри та витрати на запити. У таблиці 2 подано детальний розподіл цих витрат для даних розміром 100 ГБ, 1 ТБ і 10 ТБ.

Таблиця 2 – Вартість S3Proху

Об'єм даних	Вартість зберігання, USD	Витрати на екземпляр віртуальної машини, USD	Вартість запити, USD
100 ГБ	2,45	7,2	7,27
1 ТБ	24,5	14,4	89,25
10 ТБ	245	72	889,18

2. Експеримент Apache NiFi

Для даного підходу використано Docker-образ [23]. Код програми, який реалізує вимірювання продуктивності, знаходиться у відкритому доступі [24].

Результати тестування Apache NiFi наведено в таблиці 3.

Таблиця 3 – Швидкість доступу до файлів Apache NiFi

Розмір файлу	Середній час доступу, с
100 КБ	7,2302
1 МБ	7,2502
10 МБ	7,2402

Рішення Apache NiFi також було проаналізовано на основі середнього часу доступу до файлів різного розміру. Середній час доступу був значно довшим порівняно з S3Proху: файл розміром 100 КБ займав у середньому 7,2302 секунди. Час доступу для більших файлів, 1 МБ і 10 МБ, був подібним і становив 7,2502 секунди і 7,2402 секунди відповідно.

Для рішення Apache NiFi оцінка вартості відбувається за тією ж схемою, що й аналіз S3Proху. Вартість реалізації Apache NiFi представлена у таблиці 4.

Таблиця 4 – Вартість Apache NiFi

Об'єм даних	Вартість зберігання, USD	Витрати на екземпляр, USD	Вартість запити, USD
100 ГБ	2,45	12,6	7,28
1 ТБ	24,5	25,2	89,27
10 ТБ	245	126	867,88

3. Експеримент Apache Libcloud

Так як Apache Libcloud – бібліотека, що інтегрується прямо у код кінцевого застосунку, даний підхід не потребує проміжного кроку із запитом в окремий контейнер. Відповідно, код бібліотеки було взято з її офіційного репозиторію [25] і інтегровано згідно з наявними інструкціями. Результуючий код експерименту можна побачити по посиланню [26].

Експеримент із Apache Libcloud, бібліотекою Python, створеною для взаємодії з багатьма популярними постачальниками хмарних послуг за допомогою єдиного API, показав наступні результати (табл. 5):

Таблиця 5 – Швидкість доступу до файлів Apache Libcloud

Розмір файлу	Середній час доступу, с
100 КБ	0,0919
1 МБ	0,1059
10 МБ	0,2927

У випадку рішення Apache Libcloud спостерігалося, що середній час доступу до файлів різного розміру був трохи довшим, ніж для S3Proху, але значно коротшим, ніж для Apache NiFi. У таблиці 6 наведено розподіл витрат на впровадження Apache Libcloud.

Таблиця 6 – Вартість Apache NiFi

Об'єм даних	Вартість зберігання, USD	Витрати на екземпляр, USD	Вартість запиту, USD
100 ГБ	2,45	0	9,06
1 ТБ	24,5	0	92,84
10 ТБ	245	0	928,48

Розглянувши кожне рішення окремо, важливо провести порівняльний аналіз між трьома підходами: S3Proху, Apache NiFi та Apache Libcloud. Цей порівняльний огляд дозволяє більш повно зрозуміти сильні та слабкі сторони кожного підходу по відношенню один до одного. Загальна вартість по кожному підходу представлена у таблиці 7.

Таблиця 7 – Загальна вартість

Підхід	Об'єм	Загальна вартість, USD
S3Proху	100 ГБ	16,92
	1 ТБ	128,15
	10 ТБ	1206,18
Apache NiFi	100 ГБ	22,33
	1 ТБ	138,97
	10 ТБ	1238,88
Apache Libcloud	100 ГБ	11,51
	1 ТБ	117,34
	10 ТБ	1173,48

Загальна швидкість доступу до файлів по кожному підходу представлена у таблиці 8.

Таблиця 8 – Швидкість доступу до файлів

Підхід	Об'єм	Середній час доступу, с
S3Proху	100 КБ	0,0732
	1 МБ	0,1055
	10 МБ	0,3652
Apache NiFi	100 КБ	7,2302
	1 МБ	7,2502
	10 МБ	7,2402
Apache Libcloud	100 КБ	0,0919
	1 МБ	0,1059
	10 МБ	0,2927

5 РЕЗУЛЬТАТИ

В ході експерименту були отримані дані по вартості і швидкості передавання (тобто час, за який передається файл певного розміру) для кожного з підходів мультимарного доступу до даних. У випадку критерію продуктивності, перевагу варто надати тому підходу, який має найменший середній час доступу, для вартості – найнижчу ціну. Аналізуючи таблиці з відповідними даними можна зробити висновок, що Apache NiFi має найвищу ціну і найбільший час доступу, а, значить, серед обраних підходів, має найнижчий пріоритет вибору, якщо мова йде про вартість і швидкодію. Найкращі ж результати показало використання Apache Libcloud у вартісному показнику. Що стосується продуктивності, то і Apache Libcloud, і S3Proху показали досить близькі результати, проте на розмірі файлів до 1 МБ трохи виграє S3Proху.

Графіки загальної вартості і швидкості доступу до файлів показані відповідно на рис. 5 і 6.

Тепер можна розглянути підходи для оптимізації.

Об'єднання сильних сторін:

– S3Proху має найшвидший доступ до файлів розміром до 1 МБ;

– Apache Libcloud показує найкращу вартість на різних обсягах.

– Apache NiFi підходить для сценаріїв, де швидкодія не є критичною, проте необхідна складна оркестрація даних.

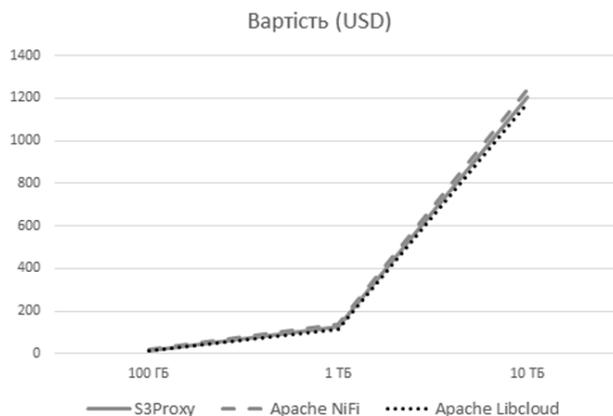


Рисунок 5 – Графік залежності вартості від обсягу зберігання

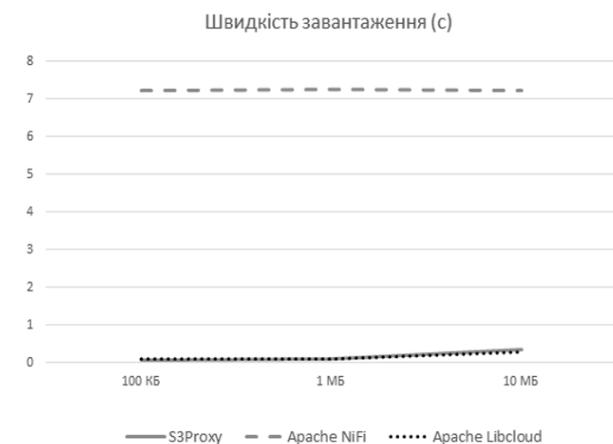


Рисунок 6 – Графік залежності часу передачі даних від розміру файлу

Динамічне перемикання між рішеннями:

– S3Proху для невеликих і середніх файлів (до 10 МБ), де важлива висока швидкість передавання;

– Apache Libcloud, коли важливим є зниження вартості;

– Apache NiFi, якщо необхідна більш складна логіка передачі.

Для оптимізації використання мультимарних підходів при наявності параметрів швидкодії і вартості пропонується наступний алгоритм:

1. Визначення вхідних параметрів. У нашому випадку це: розмір файлу, об'єм зберігання, критерій вартості, критерій часу доступу до файлу.

Для CC і CT , який параметр є більш важливим, призначаємо 1, інакше – 0.

2. Порівнюємо CC і CT .

3. Обираємо оптимальний підхід при $CT=1$, використовуючи функцію, що визначає середній час доступу для підходу A і заданого розміру файлу FS :

$$A_{opt} = \arg \min_A T(A, FS). \quad (5)$$

4. Обираємо оптимальний підхід при $CC=1$, використовуючи функцію, що визначає вартість для підходу A і заданого об'єму зберігання V :

$$A_{opt} = \arg \min_A C(A, V). \quad (6)$$

Для підходів Apache Libcloud і S3Proху правило прийняття рішень буде мати наступний вигляд:

ЯКЩО $FS \leq 1$ МБ ТА $CT=1$,

ТО обираємо S3Proху,

ІНАКШЕ обираємо Apache Libcloud.

Для опрацювання більш складних сценаріїв можна обрати Apache NiFi.

Алгоритм в загальному вигляді представлено на рис. 7.

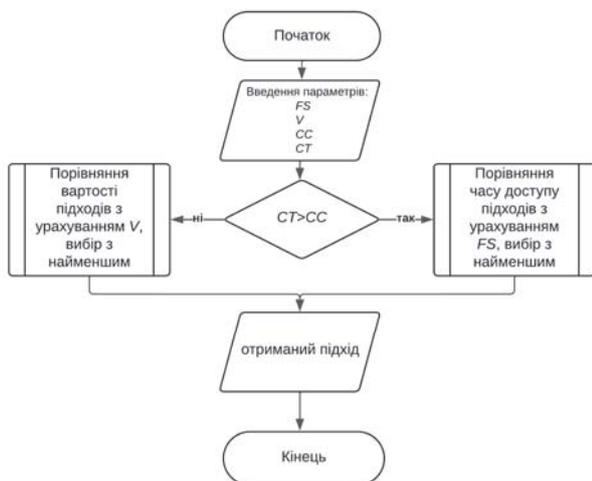


Рисунок 7 – Загальний алгоритм вибору мультимарного підходу

6 ОБГОВОРЕННЯ

Проведені дослідження показали, що підхід хмарних універсальних бібліотек (Apache Libcloud) і мультимарних шлюзів (S3Proху), маючи незначний розрив у експериментальних даних, можуть бути рекомендовані до впровадження. Для вибору підходу під конкретну задачу варто звернути на особливості кожного з них.

Хмарні універсальні бібліотеки:

– Дозволяють розробникам писати код, який може взаємодіяти з різними хмарними сервісами без залежності від конкретного провайдера.

– Спрощують процес інтеграції і роботи з хмарними сервісами.

– Можуть не підтримувати всі функції, які доступні у конкретних хмарних сервісах.

Мультимарні шлюзи:

– Дозволяють інтегрувати різні хмарні сервіси в єдине сховище, що полегшує управління даними в різних середовищах.

– Використовуються для зниження складнощів міграції даних між різними хмарними платформами.

– Додаткові шари абстракції, які вводить шлюз, можуть призвести до зниження продуктивності.

– Можуть вимагати додаткових витрат на обслуговування та підтримку.

Але для всебічного аналізу означених підходів може бути недостатнім використання лише цих критеріїв для оцінки ефективності впровадження. Тому подальше дослідження пропонується виконувати у напрямку оцінки складності впровадження кожного з рішень і побудови багатокритеріальної моделі визначення ефективності.

ВИСНОВКИ

В ході роботи було досліджено такі підходи мультимарного доступу до даних:

– multi-cloud шлюзи;

– платформи керування даними;

– універсальні хмарні бібліотеки.

Для кожного з підходів було обрано технологію, на основі якої і проводився експеримент.

Наукова новизна. Встановлено, що універсальні хмарні бібліотеки, представлені Apache Libcloud, демонструють найкращі показники з точки зору вартості та продуктивності. Трохи поступається експериментальними даними S3Proху, який є представником мультимарних шлюзів. Найгірші показники має Apache NiFi (платформи керування даними).

Практична значимість дослідження. Результати дослідження можуть бути використані організаціями для вибору найбільш відповідного підходу до мультимарного доступу до даних відповідно до їх конкретних потреб і умов.

Подальші дослідження пропонується спрямувати на визначення критерію складності впровадження кожного з підходів. І далі, на основі усіх зазначених критеріїв, пропонується створити модель, яка б змогла окреслити методологію вибору найбільш ефективних інструментів.

ЛІТЕРАТУРА

1. Petcu D. Multi-Cloud: Expectations and Current Approaches / D. Petcu // MultiCloud 2013 – Proceedings of the International Workshop on Multi-Cloud Applications and Federated Clouds – 2013. – P. 1–6. DOI: 10.1145/2462326.2462328.
2. Saceres A. State-of-the-Art Architectures for Interoperability of Heterogeneous Clouds / A. Saceres, L. S. Globa // Telecommunications and Computer Engineering (TCSET): 2022 IEEE 16th International

- Conference on Advanced Trends in Radioelectronics, Lviv-Slavske, Ukraine, 22–26 February 2022, : proceedings. – Los Alamitos: IEEE, 2022. – P. 704–709. DOI: 10.1109/TCSET55632.2022.9766965.
3. Ferrera A. J. Multi-Cloud Platform-as-a-Service Model, Functionalities and Approaches / A. J. Ferrera, D. G. Perez, R. S. Gonzalez // *Procedia Computer Science*. – 2016. – Vol. 97. – P. 63–72.
 4. Goasduff L. Why Organizations Choose a Multicloud Strategy [Electronic resource] / L. Goasduff. – Access mode: <https://www.gartner.com/smarterwithgartner/why-organizations-choose-a-multicloud-strategy>.
 5. Flexera 2024 State of the Cloud Report [Electronic resource]. – Access mode: <https://info.flexera.com/CM-REPORT-State-of-the-Cloud>.
 6. Козіна О. А. Архітектура проміжного програмного забезпечення для узгодження даних в мультихмарних системах / О. А. Козіна, В. І. Панченко, О. М. Рисованій // Вісник Національного технічного університету «ХПІ». – 2021. – № 2 (6). – С. 101–112. DOI: 10.20998/2411-0558.2021.02.07.
 7. Multi-Cloud: A Comprehensive Review / [H. A. Imran, U. Latif, A. A. Ikram et al.] // 2020 IEEE 23rd International Multitopic Conference (INMIC), Bahawalpur, Pakistan, 05–07 November 2020: proceedings. – Los Alamitos: IEEE, 2021. – P. 1–5. DOI: 10.1109/INMIC50486.2020.9318176.
 8. Interoperability in the Heterogeneous Cloud Environment: A Survey of Recent User-Centric Approaches / [I. Mansour, R. Sahandi et al.] // ICC '16: Proceedings of the International Conference on Internet of things and Cloud Computing, Cambridge, United Kingdom, 22–23 March 2016: proceedings. – New York, NY, USA: Association for Computing Machinery. – P. 1–7. DOI: 10.1145/2896387.2896447.
 9. Framework Semantic and Standard Approaches in Multi-clouds to Achieve Interoperability: A Survey / [Z. A. Adhoni, N. Dayanand Lal et al.] // *Journal of Integrated Science and Technology*. – 2022. – Vol. 10, No. 2. – P. 67–72.
 10. Sheldon R. Cloud Data Management Interface (CDMI) [Electronic resource] / R. Sheldon. – Access mode: <https://www.techtarget.com/searchstorage/definition/Cloud-Data-Management-Interface>.
 11. What is Open Cloud Computing Interface (OCCI) [Electronic resource]. – Access mode: <https://www.analyticssteps.com/blogs/what-open-cloud-computing-interface-occi>.
 12. Bennett S. Cloud Storage Gateway Vendors: 7 Top Solutions [Electronic resource] / S. Bennett. – Access mode: <https://www.resilio.com/blog/cloud-storage-gateway-vendors>.
 13. Data Management Platforms: An Empirical Taxonomy / [J. A. Hüllmann, A. Sivakumar, S. Krebber et al.] // 34th Bled eConference, Bled, Slovenia, 27–30 June 2021: proceedings. Available at: <https://aisel.aisnet.org/bled2021/9>.
 14. Integrated Management of IaaS Resources / [F. Meireles, B. Malheiro et al.] // *European Conference on Parallel Processing. Euro-Par 2014: Parallel Processing Workshops*. – 2014. – Cham : Springer, 2014. – P. 73–84. DOI: 10.1007/978-3-319-14313-2_7.
 15. Malinina J. Cloud Agnostic – an Approach You Might Want to Avoid [Electronic resource] / J. Malinina. – Access mode: <https://medium.com/coderbyte/cloud-agnosticism-and-its-hidden-cost-4d3ed6d963f>.
 16. Parker E. S3 File Gateway for Efficient, Multi-Cloud File Access [Electronic resource] / E. Parker. – Access mode: <https://www.resilio.com/blog/s3-file-gateway>.
 17. Yousry A. Data Management: A Guide to Apache NiFi [Electronic resource] / A. Yousry. – Access mode: <https://medium.com/@ansam.yousry/data-management-a-guide-to-apache-nifi-21a29ecc4591>.
 18. Apache Libcloud [Electronic resource]. – Access mode: <https://libcloud.apache.org/>.
 19. Interoperability and Portability for Cloud Computing: A guide [Electronic resource]. – Access mode: <https://www.omg.org/cloud/deliverables/CSCC-Interoperability-and-Portability-for-Cloud-Computing-A-Guide.pdf>.
 20. AWS Pricing [Electronic resource]. – Access mode: https://aws.amazon.com/ru/pricing/?aws-products-pricing.sort-by=item.additionalFields.productNameLowercase&aws-products-pricing.sort-order=asc&awsf.Free%20Tier%20Type=*all&awsf.tech-category=*all.
 21. Docker Image for S3-Proxy. Reverse proxy for AWS S3 with basic authentication [Electronic resource]. – Access mode: <https://hub.docker.com/r/pottava/s3-proxy>.
 22. Caceres A. Link to GitHub repository with experiment s3Proxy [Electronic resource] / A.Caceres. – Access mode: https://github.com/MA3STR0/multicloud_experiments/tree/master/multicloud_storage/s3proxy.
 23. Docker Image for Apache NiFi. Apache NiFi unofficial binary build [Electronic resource]. – Access mode: <https://hub.docker.com/r/apache/nifi>.
 24. Caceres A. Link to GitHub repository with experiment Apache NiFi [Electronic resource] / A.Caceres. – Access mode: https://github.com/MA3STR0/multicloud_experiments/tree/master/multicloud_storage/apache_nifi.
 25. Repository Apache Libcloud [Electronic resource]. – Access mode: <https://github.com/apache/libcloud>.
 26. Caceres A. Link to GitHub repository with experiment Libcloud [Electronic resource] / A.Caceres. – Access mode: https://github.com/MA3STR0/multicloud_experiments/tree/master/multicloud_storage/libcloud.

Received 30.08.2024.
Accepted 14.01.2025.

ANALYSIS OF DATA ACCESS APPROACHES IN A MULTI-CLOUD ENVIRONMENT

Caceres A. – Postgraduate student, Department of Information Technologies in Telecommunications, National Technical University of Ukraine “Igor Sikorsky Kyiv Polytechnic Institute”, Kyiv, Ukraine.

Globa L. – Professor, Dr. Sc., Department of Information Technologies in Telecommunications, Institute of Telecommunication Systems, National Technical University of Ukraine “Igor Sikorsky Kyiv Polytechnic Institute”, Kyiv, Ukraine.

ABSTRACT

Context. A multi-cloud system is characterized by the sequential or simultaneous use of services from different cloud providers to run applications. Such a system is a preferred infrastructure for the vast majority of IT businesses today. Currently, there are various approaches to combining cloud platforms from multiple vendors. This article explores practical approaches to achieve multi-cloud interoperability, focusing on abstract data access between different cloud storage providers and multi-cloud computing resource allocation. Key technologies and methodologies for uninterrupted data management are presented, such as the use of multi-cloud storage gateways (using S3Proxy as an example), the implementation of data management platforms (Apache NiFi), and the use of cloud-agnostic libraries (Apache Libcloud). The paper highlights the advantages and disadvantages of the selected approaches and conducts experiments to determine the cost and performance of these strategies. The result of the research is to determine the cost and performance of different approaches to data access in multi-cloud environments.

Objective. To investigate different approaches to multi-cloud data access and determine the most optimal in terms of cost and performance.

Method. We propose the optimization of multi-cloud infrastructures based on experimental data. Experimental modeling includes empirical measurements of performance and comparison of storage costs. The determination of performance is based on the measurement of data reading time and latency. The AWS S3 pricing model is used to estimate the cost. Optimization approaches are described, considering file sizes and data storage, combining the strengths of different multi-cloud approaches and dynamic switching between solutions. An algorithm for selecting multi-cloud approaches is proposed, which takes into account the criteria of cost and performance, as well as their priority.

Results. The experiment yielded values for the cost of storing and downloading data of different sizes (100 GB, 1 TB, 10 TB), and the performance of transferring files of different sizes (100 KB, 1 MB, 10 MB) for multi-cloud gateway technologies, data management platforms, and cloud-agnostic libraries. S3Proxy was found to have the fastest file access for large data volumes, while Apache Libcloud showed better value for smaller volumes. Both approaches significantly outperformed Apache NiFi. This study can contribute to the development of methods for efficient resource management in multi-cloud environments.

Conclusions. The obtained results can assist in prioritizing the selection of these paradigms, aiding organizations in developing and deploying effective multi-cloud strategies. This approach enables them to leverage the distinctive features of each cloud provider while maintaining a unified, flexible, and efficient storage and computing environment.

KEYWORDS: cloud computing, multi-cloud access, cloud-agnostic libraries.

REFERENCES

1. Petcu D. Multi-Cloud: Expectations and Current Approaches, *MultiCloud 2013 – Proceedings of the International Workshop on Multi-Cloud Applications and Federated Clouds*, 2013, pp. 1–6. DOI: 10.1145/2462326.2462328.
2. Caceres A., Globa L. S. State-of-the-Art Architectures for Interoperability of Heterogeneous Clouds, *Telecommunications and Computer Engineering (TCSET): 2022 IEEE 16th International Conference on Advanced Trends in Radioelectronics, Lviv-Slavske, Ukraine, 22–26 February 2022, : proceedings*. Los Alamitos, IEEE, 2022, pp. 704–709. DOI: 10.1109/TCSET55632.2022.9766965.
3. Ferrera A. J., Perez D. G., Gonzalez R. S. Multi-Cloud Platform-as-a-Service Model, Functionalities and Approaches, *Procedia Computer Science*, 2016, Vol. 97, pp. 63–72.
4. Goasduff L. Why Organizations Choose a Multicloud Strategy [Electronic resource]. Access mode: <https://www.gartner.com/smarterwithgartner/why-organizations-choose-a-multicloud-strategy>.
5. Flexera 2024 State of the Cloud Report [Electronic resource]. Access mode: <https://info.flexera.com/CM-REPORT-State-of-the-Cloud>.
6. Kozina O. A., Panchenko V. I., Risovaniy O. M. Arkhitektura promizhnoho prohramnoho zabezpechennya dlya uzgodzhennya danykh v mul'tykhmarinykh systemakh, *Visnik Nacionalnogo tehnicnogo universitetu “HPI”*, 2021, No. 2 (6), pp. 101–112. DOI: 10.20998/2411-0558.2021.02.07.
7. Imran H. A., Latif U., Ikram A. A. et al. Multi-Cloud: A Comprehensive Review, *2020 IEEE 23rd International Multitopic Conference (INMIC), Bahawalpur, Pakistan, 05–07 November 2020: proceedings*. Los Alamitos, IEEE, 2021, pp. 1–5. DOI: 10.1109/INMIC50486.2020.9318176.
8. Mansour I., Sahandi R. et al. Interoperability in the Heterogeneous Cloud Environment: A Survey of Recent User-Centric Approaches, *ICC '16: Proceedings of the International Conference on Internet of things and Cloud Computing, Cambridge, United Kingdom, 22–23 March 2016: proceedings*. New York, NY, USA: Association for Computing Machinery, pp. 1–7. DOI: 10.1145/2896387.2896447.

9. Adhoni Z. A., Dayanand Lal N. et al. Framework Semantic and Standard Approaches in Multi-clouds to Achieve Interoperability: A Survey, *Journal of Integrated Science and Technology*, 2022, Vol. 10, No. 2, pp. 67–72.
10. Sheldon R. Cloud Data Management Interface (CDMI) [Electronic resource]. Access mode: <https://www.techtarget.com/searchstorage/definition/Cloud-Data-Management-Interface>.
11. What is Open Cloud Computing Interface (OCCI) [Electronic resource]. Access mode: <https://www.analyticssteps.com/blogs/what-open-cloud-computing-interface-occi>.
12. Bennett S. Cloud Storage Gateway Vendors: 7 Top Solutions [Electronic resource]. Access mode: <https://www.resilio.com/blog/cloud-storage-gateway-vendors>.
13. Hüllmann J. A., Sivakumar A., Krebber S. et al. Data Management Platforms: An Empirical Taxonomy, *34th Bled eConference, Bled, Slovenia, 27–30 June 2021: proceedings*. Available at: <https://aisel.aisnet.org/bled2021/9>.
14. Meireles F., Malheiro B. et al. Integrated Management of IaaS Resources, *European Conference on Parallel Processing. Euro-Par 2014: Parallel Processing Workshops, 2014*. Cham, Springer, 2014, pp. 73–84. DOI: 10.1007/978-3-319-14313-2_7.
15. Malinina J. Cloud Agnostic – an Approach You Might Want to Avoid [Electronic resource]. Access mode: <https://medium.com/coderbyte/cloud-agnosticism-and-its-hidden-cost-4d3ed6d963f>.
16. Parker E. S3 File Gateway for Efficient, Multi-Cloud File Access [Electronic resource]. Access mode: <https://www.resilio.com/blog/s3-file-gateway>.
17. Yousry A. Data Management: A Guide to Apache NiFi [Electronic resource]. Access mode: <https://medium.com/@ansam.yousry/data-management-a-guide-to-apache-nifi-21a29ecc4591>.
18. Apache Libcloud [Electronic resource]. Access mode: <https://libcloud.apache.org/>.
19. Interoperability and Portability for Cloud Computing: A guide [Electronic resource]. Access mode: <https://www.omg.org/cloud/deliverables/CSCC-Interoperability-and-Portability-for-Cloud-Computing-A-Guide.pdf>.
20. AWS Pricing [Electronic resource]. Access mode: https://aws.amazon.com/ru/pricing/?aws-products-pricing.sort-by=item.additionalFields.productNameLowercase&aws-products-pricing.sort-order=asc&awsf.Free%20Tier%20Type=*all&awsf.tech-category=*all.
21. Docker Image for S3-Proxy. Reverse proxy for AWS S3 with basic authentication [Electronic resource]. Access mode: <https://hub.docker.com/r/pottava/s3-proxy>.
22. Caceres A. Link to GitHub repository with experiment s3Proxy [Electronic resource]. Access mode: https://github.com/MA3STR0/multicloud_experiments/tree/master/multicloud_storage/s3proxy.
23. Docker Image for Apache NiFi. Apache NiFi unofficial binary build [Electronic resource]. Access mode: <https://hub.docker.com/r/apache/nifi>.
24. Caceres A. Link to GitHub repository with experiment Apache NiFi [Electronic resource]. Access mode: https://github.com/MA3STR0/multicloud_experiments/tree/master/multicloud_storage/apache_nifi.
25. Repository Apache Libcloud [Electronic resource]. Access mode: <https://github.com/apache/libcloud>.
26. Caceres A. Link to GitHub repository with experiment Libcloud [Electronic resource]. Access mode: https://github.com/MA3STR0/multicloud_experiments/tree/master/multicloud_storage/libcloud.

МЕТОД USE CASE В УПРАВЛІННІ ІТ-ПРОЄКТОМ НА ОСНОВІ МЕТОДОЛОГІЇ AGILE

Свінцицька О. М. – канд. економ. наук, доцент, доцент кафедри комп'ютерних наук, Державний університет Житомирська політехніка, м. Житомир, Україна.

Пулеко І. В. – канд. техн. наук, доцент, доцент кафедри комп'ютерної інженерії та кібербезпеки, Державний університет Житомирська політехніка, м. Житомир, Україна.

Граф М. С. – доктор філософії, завідувачка кафедри комп'ютерних наук, Державний університет Житомирська політехніка, м. Житомир, Україна.

Петросян Р. В. – старший викладач кафедри комп'ютерних наук, Державний університет Житомирська політехніка, м. Житомир, Україна.

АНОТАЦІЯ

Актуальність. Розглянуто роль та процес формування вимог користувача на основі методу Use Case в оцінці трудомісткості Agile проекту на етапі попередньої оцінки менеджментом компанії. Вже з середини 70-х років відомо, що помилки у вимогах є найчисленнішими, найдорожчими та трудомісткими для виправлення в проєктах. У зв'язку з цим, підвищується ступінь важливості управління вимогами в ІТ-проєктах з використанням сучасних технологій та методів їх формування та оцінки.

Мета. Формування та оцінка вимог користувача в управлінні ІТ-проєктами на основі методу Use Case та їх вплив на один з показників ефективності проєкту на етапі планування, зокрема трудомісткість.

Метод. В статті запропоновано новий авторський підхід до формування та оцінки вимог користувача в Agile-проєктах з урахуванням впливу ризиків та оцінки складності системи на основі методу Use Case, і як результат дослідження та пропозиції по досягненню поставленої мети, запропоновано математичну модель оцінки трудомісткості проєкту. Математичний шаблон моделі дозволяє враховувати додаткові змінні, які можуть впливати на проєкт, зокрема кількість рівнів користувачів, наявний функціонал та ризики технічного та організаційного характеру, є гнучкою і може адаптуватися до різних потреб конкретного проєкту, що відповідає принципам методології Agile. Кількість компонентів у формулі можуть бути змінені, щоб врахувати важливість різних змінних, або розширена, щоб враховувати додаткові змінні, які можуть впливати на проєкт.

Результати. Розроблена та апробована на прикладі мобільного додатку математична модель оцінки трудомісткості проєкту на основі методу use case, що містить набір вихідних даних для розробки продукту та обмежень щодо зміни вимог користувачів та організаційно-технічних ризиків. Запропонована математична модель дозволяє швидко, точно та ефективно визначати сценарії трудомісткості витрат проєкту різного типу та рівня складності та може слугувати ефективним інструментом прийняття управлінських рішень.

Висновки. Загальні висновки, отримані після аналізу методів формування та оцінки вимог користувача в практиці Agile управління, наступні. На етапі планування робіт базову модель оцінки проєкту, яка базується на експертній оцінці кожної функціональної вимоги, замінено на більш сучасну та складну, яка базується на методі use case та враховує зміни вимог користувачів та інші ризики розробки продукту. Для складання нової моделі використовуються графічні, аналітичні та математичні інструменти, зокрема, діаграма use case, коригуючі коефіцієнти, що враховують складність актора та use case, коефіцієнти, що враховують організаційний та технічний ризики, і як результат, отримуємо математичний формат розрахунку трудомісткості проєкту. Такий підхід дозволяє швидко адаптуватися до різних типів проєктів, а за умови правильного визначення вихідних даних, модель дозволяє отримати досить точні оцінки на ранньому етапі планування проєкту. Практично отримані результати дослідження демонструють потенціал запропонованої математичної моделі, яка може мати логічне продовження через верифікацію моделі на більшій вибірці та оцінку її стійкості до різних типів проєктів і ризиків.

КЛЮЧОВІ СЛОВА: вимоги користувачів, ІТ-продукт, Agile-проєкт, User Story, Use Case, мобільний додаток.

АБРЕВІАТУРИ

API – Application Programming Interface;
ACF – Actor Complexity Factor;
BA – Business analysis;
CR – Communication and Risk;
Dev – Development;
GUI – Graphical User Interface;
ІТ – інформаційні технології;
MVP – minimum viable product;
OF – Organizational Factor;
OCF – Organizational Complexity Factor;
PM – Project management;
TF – Technical Factor;

TCF – Technical Complexity Factor;
QA – Quality Assurance;
UI – User Interface;
UML – Unified Modeling Language;
UCCF – Use Case Complexity Factor.

НОМЕНКЛАТУРА

E – множина всіх етапів проєкту;
 F – множина фіч;
 G – показник ризику;
 i – етапи роботи над проєктом;
 J – множина use case;
 K_i – коефіцієнт складності етапу;

K_{ACF} – середня складність актора;
 K_{UCCF} – середня складність use case;
 K_{TCF} – показник технічної складності проекту;
 K_{OCF} – показник організаційної складності проекту;
 PF – продуктивність розробки програмного забезпечення на j -му use case;
 c_i – кількість акторів для i -го use case;
 c_j – кількість use case для j -го актора;
 T – загальна трудомісткість проекту;
 $T_i(f)$ – час, необхідний для розробки фічі f на етапі i роботи над проектом;
 X – множина акторів.

ВСТУП

Проекти програмного забезпечення мають дуже високу ймовірність провалу, і основною причиною цього є поганий процес розробки та оцінки вимог на етапі планування. В Agile методології формування вимог є основним і найважливішим етапом життєвого циклу розробки програмного забезпечення [3]. А помилка в оцінці вимог може збільшити витрати та час виконання та призвести до збою проекту програмного забезпечення. Подолання цих недоліків потребує застосування нового підходу до вибору методів збору, формування та оцінки вимог користувача. Незважаючи на широкий спектр існуючих методів, таких як інтерв'ю, анкетування, прототипування, використання прецедентів та ін., сучасні IT-проекти вимагають більш гнучких та ефективних підходів. Саме тому набули популярності такі методи, як сценарії користування (use case) та історії користувачів (user story), які дозволяють більш гнучко описувати функціональність системи та легко адаптувати її до змінних вимог. В статті проаналізовано їх переваги та недоліки, здійснено вибір методу use case для побудови математичної моделі прогнозування оцінки трудомісткості проекту на основі вимог користувача, яку практично апробовано на проекті мобільного додатку.

Мета дослідження. Моделювання трудомісткості IT-проекту на основі процесу формування та оцінки вимог користувача методом Use Case.

Об'єкт дослідження є процес побудови математичної моделі оцінки трудомісткості Agile проекту, що враховує вимоги користувача IT-продукту, ризик розробки та інструменти методу Use Case.

Предмет дослідження. Поєднання традиційних методів формування та оцінки вимог користувача з інноваційними, які побудовані на основі використання природної мови, концептуальних моделей, які допомагають описати математично та візуалізувати процеси в Agile-проектах, а також знаходять своє відображення в показниках трудомісткості.

1 ПОСТАНОВКА ПРОБЛЕМИ

Розробка та оцінка вимог є ключовою та центральною для кожного успішного проекту. © Свінцицька О. М., Пулеко І. В., Граф М. С., Петросян Р. В., 2025
DOI 10.15588/1607-3274-2025-1-17

кілька причин, чому IT-проекти зазнаються невдач, серед них є причини, які пов'язані з вимогами, зокрема, методами їх збору, оцінки та документування. Існуючі методи часто мають ряд обмежень, таких як суб'єктивність, залежність від експертних оцінок та недостатня гнучкість для адаптації до змінних вимог користувача. Тому виникає потреба у розробці нових підходів до оцінки проектів, які б дозволили забезпечити більш точне визначення трудомісткості проекту на ранніх етапах, при цьому враховували фактори ризику та зміну вимог користувача. Рішення даної проблеми може полягати у створенні моделі, яка б дозволила адаптувати до таких умов ранню оцінку тривалості та вартості проекту, а також мінімізувати вплив людського фактору на точність оцінки, використовуючи більш об'єктивні методи розрахунку.

Одним із перспективних підходів до вирішення цієї проблеми є застосування методу use case як основи для побудови математичної моделі оцінки трудомісткості проекту. Це дозволить врахувати складність кожного use case, його взаємозв'язки з акторами, а також врахувати відповідні ризики, пов'язані зі зміною вимог користувача. Така модель дозволить створити більш точну, гнучку та об'єктивну систему оцінки проектів на ранньому етапі реалізації.

Для прийняття рішення про застосування даного підходу доцільно провести дослідження на основі порівняння двох варіантів оцінки проекту. Перший – традиційний, шляхом прямого підрахунку витрат годин на розробку проекту та застосування до них коригуючих коефіцієнтів, що сформовані експертним шляхом на основі попереднього досвіду. Задача тут полягає у розрахунку загальної тривалості проекту на основі показника трудомісткості T , враховуючи набір фіч F та їхні характеристики, що сформовані на основі вимог користувача, які враховують етапи розробки проекту, i (UI, Dev, QA, PM, BA, CR) та відповідний коефіцієнт їх складності, K_i . Для цього пропонується формалізувати цей варіант оцінки через шаблон математичної моделі. Тоді, математична модель визначення загальної тривалості проекту T , матиме такий вигляд (формула (1)):

$$T = \sum (f \in F) \sum (i \in E) T_i(f) \cdot K_i, \quad (1)$$

де $F = \{f_1, f_2, \dots, f_m\}$.

Поряд цим, такий формат розрахунку є достатньо спрощеним, має ряд обмежень, таких як: суб'єктивність та залежність від експертних оцінок, висока трудомісткість самих оціночних робіт, відсутність гнучкості в залежності від змін впливу зовнішнього контексту. Застосування методу use case дозволить створити нову математичну модель, яка враховуватиме такі фактори, як: складність use case, ризику проекту та продуктивність розробки.

Отже, постановка задачі полягає у створенні математичної моделі для прогнозування оцінки тривалості проекту T^* на ранніх етапах розвитку, з урахуванням складності акторів, K_{ACF} та складності use case, K_{UCCF} , продуктивності розробника, PF та рівня ризику проекту, G . При цьому, задано:

$$X = \{x_1, x_2, \dots, x_n\};$$

$$J = \{j_1, j_2, \dots, j_m\}.$$

В якості критерію оцінки визначено загальну трудомісткість проекту при дотриманні заданих обмежень:

- технічних (наприклад, вибір технології, платформ);
- організаційних (наприклад, стандарти якості, досвід команди);
- зміни вимог користувачів (наприклад, зміна функціональності).

Використання методу use case в якості основи для математичної моделі дозволить досягти більш точної оцінки та прогнозів тривалості проекту порівняно з традиційними методами.

2 ОГЛЯД ЛІТЕРАТУРИ

Дослідники зробили великий внесок у сферу формування та оцінки вимог в Agile-проектах, але ступінь, до якого їхні пропозиції були прийняті на практиці, залишається недостатньо розкритим. Сучасні дослідження показують, що 71% програмних проєктів зазнають невдачі через погане управління та неправильну оцінку вимог на етапі планування. Доповідь Standish CHAOS [1], яка досліджувала 9236 IT-проектів, виявила, що трьома основними причинами невдач проєктів були: відсутність введення даних користувачами, неповні вимоги або зміни вимог. Крім того, дослідження свідчать, що [2]: деякі вимог не вистачає в кінці ітерації, а деякі вимоги були неоднозначними та непослідовними, що з часом потребувало додаткової роботи над вимогами; деякі вимоги втрачаються або недостатньо структуровані чи мають незрозуміле походження, що призводить до втрати їх розуміння.

Методологія Agile була призначена головним чином для того, щоб допомогти розробникам створити проєкт, який може швидко адаптуватися до перетворення запитів і підходить для проєктів з високим ступенем невизначеності. Agile – продукт будується інкрементально, тобто замість початкового жорсткого планування створення і випуску всього продукту, одразу невеликі цінні інкременти продукту плануються та випускаються поступово [13].

Розглянемо публікації та основні нормативні документи, що визначають основні процеси, пов'язані з формуванням та визначенням вимог користувачів. Стандарт IEEE Std 610.12-1990 містить глосарій термінів, пов'язаних із програмною інженерією, відповідно до якого вимога – це «умова або здатність, необхідна користувачеві для вирішення проблеми або досягнення об'єктивних цілей» [4]. Поряд з цим, вітчизняні автори публікацій [5, 6] доповнюють це

визначення можливостями, якими повинна володіти система, щоб виконати контракт або задовільнити стандартам, специфікаціям або іншим формальним документам [5]. Інший підхід більш узагальнений. Це набір потреб потенційних користувачів щодо властивостей, якостей та функцій програмного продукту, який потрібно розробити або модифікувати [6].

Стандарт ISO/IEC 25062 [7] прямо не згадує вимоги, але неявно забезпечує основу для визначення вимог щодо зручності їх використання. Стандарт ISO 9241-210:2019 спрямований на визначення процесів для визначення вимог до взаємодії користувача з інформаційно-комунікаційними технологіями. Він надає загальні рекомендації щодо того, як слід збирати, аналізувати та документувати вимоги користувачів під час розробки систем, що включають в себе інтерфейси користувача [8]. Стандарт ISO/IEC/IEEE 29148 містить різні аспекти збору, документування, аналізу, перевірки та управління вимогами, включаючи вимоги користувачів, надає загальні вказівки для їх управління та обробки [9].

Авторами в наукових виданнях публікуються різні підходи опису процесів і методів, які можна використовувати для збору інформації про користувачів та їхні завдання [10, 11, 12]. Деякі автори чітко розмежовують потреби користувача та вимоги користувача [11]. Вимоги користувачів зазвичай включаються як частина вимог зацікавлених сторін документ специфікації, як описано в ISO/IEC/IEEE 29148. В авторській статті [2] наводяться практичні результати досліджень щодо структурування вимог за різними ознаками в процесі розробки проєктів. В Agile-проектах зачату формування вимог здійснюється двома основними способами – з використанням природної мови, тобто мови, якою користувачі або стейкхолдери проєкту спілкуються (наприклад user story та use case) та з використанням концептуальних моделей (model based – Use Case Diagrams, Class Diagrams, Activity Diagrams, State Diagrams). Найбільш широке поширення отримав перший підхід, як найбільш гнучкий, а другий носить допоміжний характер. Поряд з цим використання графічних елементів є широко поширеним в практиці формування вимог. Основною метою використання графічних елементів є надання додаткової інформації у вигляді анотацій або допомогти візуалізувати вимоги. Вони є більш наочною формою бачення вимог. Вибір методу залежить від типу проєкту, специфіки бізнесу та вподобань команди. Найчастіше використовують комбінацію різних методів для найкращого розуміння та документування вимог стейкхолдерів [2].

З розвитком продуктів та інформаційних технологій вимоги користувачів стають дедалі складнішими, і в результаті генеруються більш сучасні методи отримання вимог, за допомогою яких вимоги стають вимірними, тестованими, пов'язаними

з бізнес-потребами, і описаними з рівнем деталізації достатнім для конструювання системи.

Отже, одна з проблем в управлінні Agile-проектами – це відсутність загальноприйнятих термінів та підходів, якими користуються фахівці для опису вимог користувача та їх впливу на оцінку трудомісткості проєктів. У різних джерелах проблему пропонується вирішувати, виходячи з різних умов, що склалися в менеджменті проєкту та мало уваги приділяється ролі управління вимогами в цих процесах. З огляду на розбіжність робимо висновок про важливість вибору та обґрунтування підходів та методів до формування та аналізу вимог при розробці IT-проектів на основі Agile.

3 МАТЕРІАЛИ І МЕТОДИ

В цьому розділі статті автором наводиться опис запропонованої моделі оцінки трудомісткості Agile проєкту на основі методу use case. Варто відзначити, що формування вимог – це постійний процес, який починається на етапі пресеєлу (попередньої оцінки) і триває протягом усього життєвого циклу проєкту. На кожному етапі вимоги уточнюються, деталізуються і можуть змінюватися, використовуються різні інструменти та методи. Тому, автором проводиться дослідження та наводяться пропозиції щодо вирішення проблеми формування та оцінки вимог в управлінні IT-проектом методом use case на етапі попередньої оцінки проєкту за візією замовника. На цьому етапі головною метою є отримати первинне уявлення про проєкт, визначити його масштаб, складність та потенційну вартість за різними складовими з можливістю гнучкої адаптації проєкту до змінних умов замовника. Зазвичай це високорівневі, неформалізовані вимоги, які замовник висловлює у загальних рисах. Вони слугують для створення першої оцінки проєкту та визначення подальших дій. Запропонована автором модель дозволяє вирішити поставлене завдання. В нашому дослідженні така модель буде мати математичну формалізовану складову, що забезпечить досягнення поставленої задачі, тобто створити модель оцінки тривалості Agile-проекту в годинах на етапі попередньої оцінки з метою підвищити її точність та забезпечити більш гнучке управління проєктом у відповідності до змінних вимог користувача та впливу організаційно-технічних ризиків.

Розглянувши існуючі методи формування та опису вимог будемо поєднувати графічні, аналітичні та математичні інструменти для побудови моделі оцінки трудомісткості проєкту. На основі діаграми use case, яка використана в даному дослідженні для візуалізації бізнес-вимог проєкту мобільного додатку відеоредактора, опишемо етапи побудови моделі.

На першому етапі автором пропонується застосувати коригуючі коефіцієнти до заданої продуктивності розробки для i -го use case, що враховують складність кожного актора та відповідного варіанта використання (use case), що

відображені на діаграмі (рис.3) з метою визначення рівня складності системи. Для цього розподілимо акторів та use case на прості, середні і складні за критеріями, представленими в табл.2 та табл.3.

Таблиця 2 – Критерії розподілу акторів на прості, середні і складні

Класифікація актора	Характеристика актора
Simple	Зовнішня система, яка взаємодіє із системою, яка використовує чітко визначений API
Average	Зовнішня система, яка взаємодіє із системою, що використовує стандарт протоколів зв'язку (напр. TCP/IP, HTTP, база даних)
Complex	Людина-актор, що використовує GUI інтерфейс програми

З вище представлених характеристик акторів, нами зроблено такі висновки щодо ідентифікації акторів на прості чи складні:

- якщо актор взаємодіє з декількома варіантами використання без додаткових умов, то він, швидше за все, є простим;
- якщо взаємодія між актором і системою є складною, в нашому прикладі це додаткові дії, які повинен виконати актор, щоб забезпечити реалізації варіанту використання, то актор, швидше за все, є складним;
- важливість актора для системи. якщо актор є важливим для системи, то він, швидше за все, є складним.

Наступним етапом побудови моделі є розрахунок ACF , який визначається як сума добутків кількості акторів за категорією складності та його ваги. Отримане середнє значення з цього числа на одного актора і буде K_{ACF} .

Далі аналізуємо систему за критерієм кількості простих, середніх та складних варіантів використання (табл. 3).

Таблиця 3 – Класифікація use case на прості, середні та складні

Класифікація use case	Межі use case
Simple	межі 1–3
Average	межі 4–7
Complex	8 і більше

На основі отриманих даних розраховуємо коефіцієнт середньої складності use case за даними $UCCF$ та відповідний K_{UCCF} , які визначаються аналогічно ACF та K_{ACF} відповідно.

Після аналізу складності системи в моделі автором пропонується використати показник впливу організаційно-технічних ризиків, G . Отже, наступним етапом побудови моделі є опис його розрахунку.

На цьому етапі формується набір субфакторів та їх відносна важливість в розробці проєкту. Окремо формується набір технічних та організаційних субфакторів. Команда експертів присвоює значення та вагу кожному з субфакторів, виходячи зі складності проєкту, зокрема: значення 0 – якщо субфактор не має

значення, 3 – якщо субфактор має середній вплив та 5 – якщо субфактор має дуже великий вплив на проект. На основі сформованого набору із 13 технічних субфакторів та їх ваги визначається K_{TCF} . Це дозволяє оцінити складність розробки та впровадження програмного забезпечення. Чим вище значення K_{TCF} , тим складнішим вважається проект. Формула для розрахунку:

$$k_{TCF} = 0,6 + \frac{TF}{100}, \quad (2)$$

де TF – це індекс, який враховує конкретні технічні аспекти проекту. Його значення, обчислюється на основі інших параметрів, таких як кількість функцій, складність алгоритмів, використання нових технологій тощо.

0,6 – це константа, яка додає базовий рівень складності до проекту. Значення константи 0,6 відображає мінімальний рівень складності, який присутній у будь-якому проекті, а також дозволяє більш точно порівнювати різні проекти за їхньою складністю.

Оцінка організаційних ризиків має аналогічний підхід до розрахунку, як і для TCF . Показник організаційної складності проекту K_{OCF} враховує різноманітні фактори, які можуть вплинути на загальну ефективність проекту, наприклад, досвід команди, технологічна складність проекту, розмір проекту тощо. Для цього враховано 8 субфакторів, кожному з яких визначено вагу у загальному їх значенні для розробки проекту. Формула для розрахунку матиме вигляд:

$$k_{OCF} = 1,4 + (-0,03 \cdot OF), \quad (3)$$

де 1,4 – це константа, яка додає базовий рівень складності до проекту. Навіть якщо всі інші фактори, що впливають на проект, є ідеальними, завжди існує певний рівень складності, пов'язаний з організацією роботи, комунікацією, управлінням проектом тощо.

-0,03 – це коефіцієнт, який показує, як сильно фактор OF впливає на загальну складність. Мінус перед числом означає, що зі збільшенням OF значення K_{OCF} зменшується. Іншими словами, чим вище значення OF , тим менш складним вважається середовище проекту.

Визначивши фактори впливу на загальний рівень ризику проекту та згрупувавши їх дві категорії отримаємо загальний показник впливу організаційно-технічних ризиків G , що розраховується за формулою:

$$G = K_{TCF} \cdot K_{OCF}, \quad (4)$$

Розрахувавши коефіцієнти, що пропонується використовувати для коригування продуктивності розробки системи для i -го use case, переходимо до наступного етапу розробки моделі – визначення трудомісткості проекту, T^* , що завершує процес побудови загальної моделі. В нашому випадку це формалізований шаблон математичної моделі, яка дозволяє розрахувати загальну тривалість проекту у годинах, T^* (формула 5):

$$T^* = \sum (c_i \cdot PF_i \cdot k_{ACF}) + (c_j \cdot PF_j \cdot k_{UCCF}) + (c_k \cdot PF_k \cdot G), \quad (5)$$

де c_i – кількість акторів для i -го use case за даними рисунку 3; PF_i – продуктивність розробки для i -го use case, що задано на рівні 40 годин на розробку, як базовий рівень для даної команди на основі попереднього досвіду роботи.

Запропоновані коефіцієнти та математичний шаблон розрахунку трудомісткості проекту на основі методу use case, може використовуватись до різних проектів з метою побудови різних сценаріїв розвитку проекту при змінних вимогах користувача до функціоналу та існуючих ризиків.

Отже, наведені автором етапи демонструють детальний підхід до оцінки трудомісткості проекту на основі методу Use Case на етапі попередньої оцінки проекту для керівників та бізнес-аналітиків ІТ-проекту. Модель враховує зміну вимог користувачів, що виражена через коефіцієнти складності актора та use case, а також як технічні аспекти проекту (складність функціоналу, технології), так і організаційні фактори (досвід команди, стабільність вимог), які закладені у відповідному коефіцієнті.

Запропонована математична модель дозволяє об'єднати всі отримані дані та розрахувати загальну трудомісткість проекту. При зміні вхідних параметрів, оцінюваний показник трудомісткості автоматично змінює своє значення. Такий підхід дозволяє швидко адаптуватися до різних типів проектів, а за умови правильного визначення вхідних даних, модель дозволяє отримати досить точні оцінки на ранньому етапі планування проекту.

4 ЕКСПЕРИМЕНТИ

Для вирішення поставленої задачі, в даній частині роботи, застосуємо набір вихідних даних для побудови запропонованої математичної моделі та опишемо послідовність етапів робіт для отримання результатів від її застосування в умовах Agile проекту створення мобільного додатку-відеоредактора, що базується на концепції методу use case.

В даному дослідженні для прикладу обрано ІТ-продукт, що забезпечує покращення конкурентних позицій замовника, який бажає створити мобільний додаток-відеоредактор для молодих талантів. Аутсорсингова компанія отримала Vision проекту від замовника в якості бізнес-вимог до продукту. Vision проекту – це велика ідея, яку замовник реалізує через продукт. У ньому містяться основні тези про продукт:

яким він буде, яку мету переслідує бізнес, хто користувач. У Vision від замовника продукту Filmy є інформація про ринок, користувачів, монетизацію та ідейні цілі Filmy. Також у документі перераховані фічі, які необхідно реалізувати. На основі цього документу буде тривати підготовка до написання вимог різних рівнів. Відповідно до Vision проекту проджект з бізнес-аналітиком готують опис користувацьких вимог для подальшої роботи в Confluence. Перша версія продукту дозволить користувачам створювати відео, відео дуети та ділитися ними онлайн. Далі планується розширювати функціональність за рахунок великої бібліотеки відео та аудіо ефектів. Однак, оскільки даний проект має за мету подальшу комерціалізацію і у замовника відсутні чіткі цілі подальшого розвитку функціоналу, що робить задачу достатньо ризикованою для реалізації та унеможливає чітке планування витрат на проект, в тому числі трудових і фінансових. В цих умовах робота над проектом лягає в площину концепції управління проектами Agile у поєднанні із сучасними інструментами моделювання вимог користування та їх оцінки на основі методу use case.

Для отримання вихідних даних, їх зберігання та використання використовувався пакет програмного забезпечення від компанії Atlassian, зокрема, це Jira і Confluence. На рисунку 1 відображено основні продукти Atlassian, які застосовуються в процесі розробки проекту програмного забезпечення.



Рисунок 1 – Інструменти Atlassian в процесі розробки IT-проекту

Jira – інструмент управління проектами, який дозволяє стежити за завданнями, багами та планувати релізи програмного забезпечення. В Atlassian Marketplace є багато додатків для інтеграцій, які вирішують супутні задачі по управлінню проектом. В нашому дослідженні використовувалися плагіни для командної роботи щодо оцінки тривалості задач по проекту за традиційним підходом: Agile Poker for Jira – Planning & Estimation (рис. 2) та Cost Monitoring for Jira (рис. 3).

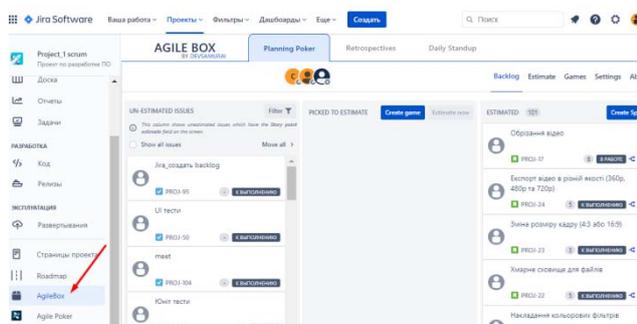


Рисунок 2 – Плагін Agile Poker for Jira – Planning & Estimation



Рисунок 3 – Плагін Cost Monitoring for Jira

Для збереження даних та спільної роботи над проектом використовувався інструмент Confluence. Тут накопичені командою знання поєднані із можливостями для спільної роботи.

Для реалізації моделі та обробки даних використовувався пакет аналітичних функцій Excel. Для моделювання та візуалізації вимог стейкхолдерів з точки зору користувачів та їх взаємодії з системою використовувався метод Use Case Diagram у графічному інструменті draw.io.

Use Case Diagram для мобільного додатку відеоредактора Filmy відображена на рис. 3.

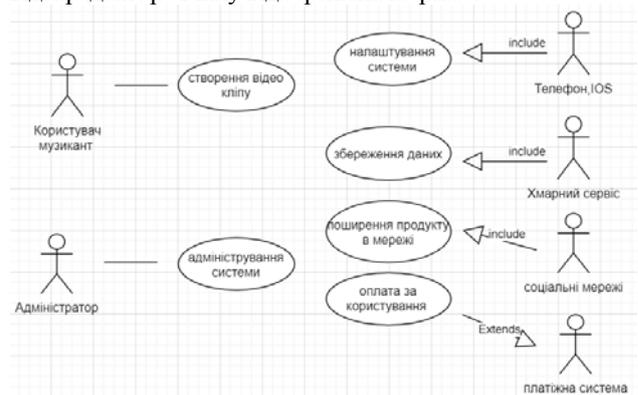


Рисунок 3 – Use Case Diagram для мобільного додатку відеоредактора Filmy

В нашому дослідженні візуалізація допомагає оцінити складність системи, з точки зору взаємодії користувачів із системою. Класифікація акторів додатку на рівні складності наведена в таблиці 4.

Таблиця 4 – Класифікація акторів на прості, середні та складні

Складність актора	Актори	К-ть акторів	Вага	Складність актора
Simple	X3, X4	2	1	2
Average	X2, X5	2	2	4
Complex	X1, X6	2	3	6

На основі отриманих даних розраховуємо K_{ACF} , який становить в нашому випадку 2,0.

В нашій моделі також передбачено аналіз системи за рівнем складності. В таблиці 4 наведено характеристику системи за критерієм кількості простих, середніх та складних варіантів використання (табл. 5).

Таблиця 5 – Розподіл use case на прості, середні та складні

Класифікація use case	Межі use case	Актори	К-ть use case	Вага	Складність use case
Simple	межі 1–3	X2, X3, X4, X6	4	0,15	0,6
Average	межі 4–7	X5	4	0,35	1,4
Complex	8 і більше	X1	8	0,5	4

Отже, в нашому прикладі, складний актор – музикант, який взаємодіє з більш ніж 7-ма варіантами використання. Телефон та адміністратора відносимо до середніх акторів, які взаємодіють з 4-ма варіантами використання, що можна побачити на рисунку 2 (Use Case Diagram). Маркетолог, банк, хмарний сервіс і соціальні мережі визначено як прості актори, оскільки вони взаємодіють з одним варіантом використання. Варто відмітити, що представлена класифікація демонструє зв'язки системи на вищому рівні з її зовнішніми атрибутами. Кожен елемент системи має свій набір складових, який деталізує та описує вимоги для потреб наступного рівня. Наприклад, банк як простий актор бере участь у двох варіантах використання: «Оплатити користування додатком» та «Прийняти платіж за користування додатком». В свою чергу варіант використання «Оплатити користування додатком» передбачає такі дії банку:

- прийняття запиту на оплату від додатку;
- перевірка запиту на оплату на наявність помилок;
- обробка запиту на оплату;
- надання додатку інформації про успіх або невдачу оплати.

Варіант використання «Прийняти платіж за користування додатком» для банку, як актора, передбачає такі дії банку:

- прийняття платежу від користувача;
- перевірка платежу на наявність помилок;
- зарахування платежу на рахунок додатку.

Представлена класифікація акторів та use case на прості, середні та складні є досить умовною, оскільки кількість варіантів використання та акторів може змінюватися в залежності від конкретних вимог до додатку, з якими взаємодіє користувач. Наявність простих, середніх та складних зв'язків в мобільному додатку Filmu дає нам підстави зробити висновок, що аналізована система відноситься до середнього рівня складності, оскільки:

- у додатку є шість акторів, з яких чотири є складними. Це означає, що додаток повинен підтримувати взаємодію з широким спектром зовнішніх систем;
- у додатку є 16 варіантів використання, з яких 8 є складними. Це означає, що додаток повинен підтримувати широкий спектр функцій;
- додаток повинен підтримувати такі технології, як API, протоколи зв'язку та GUI. Це вимагає від додатку наявності значної функціональності та складності.

На основі отриманих даних в таблиці 4 розраховуємо K_{UCCF} , що складе 2,08.

Для розрахунку показника G з нашого математичного шаблону моделі, побудуємо відповідні таблиці (таблиця 6 та 7), в яких наведено дані щодо переліку факторів ризику, експертної оцінки впливу організаційно-технічних ризиків на їх ваги у проекті.

Таблиця 6 – Показники технічної складності проекту

Фактор	Опис	Вага	Оцінка	TF
T1	Розподілена система	2.0	5	10
T2	Вимоги до швидкодії та часу відгуку	1.0	5	5
T3	Зручність використання кінцевим користувачем	1.0	3	3
T4	Внутрішня обчислювальна складність	1.0	2	2
T5	Повторне використання коду	1.0	3	3
T6	Легкість інсталяції	0.5	1	0.5
T7	Легкість використання	0.5	5	2.5
T8	Портативність на інші платформи	2.0	2	4
T9	Обслуговування системи	1.0	2	2
T10	Паралельна обробка	1.0	3	3
T11	Функції безпеки	1.0	5	5
T12	Доступ для третіх сторін	1.0	1	1
T13	Навчання кінцевих користувачів	1.0	1	1
				Загальний TF: 42

На основі даних таблиці розраховуємо TF , що складає для нашого проекту 42, а далі виконуючи умову формули 2, отримуємо K_{TCF} на рівні 1.02.

Для визначення відповідного коефіцієнта на основі факторів організаційної складності використаємо дані таблиці 8.

Таблиця 8 – Показники організаційної складності проекту

Фактор	Опис	Вага	Оцінка	OF
E1	Знайомство з використовуваним процесом розробки	1.5	3	4.5
E2	Досвід у розробці подібних додатків	0.5	3	1.5
E3	Досвід команди в об'єктно-орієнтованій розробці	1.0	2	2.0
E4	Компетентність провідного аналітика	0.5	5	2.5
E5	Мотивація команди	1.0	2	2.0
E6	Стабільність вимог	2.0	1	2.0
E7	Залучення співробітників на неповну зайнятість	-1.0	0	0.0
E8	Складність мови програмування	-1.0	4	-4.0
				Загальний OF: 10.5

В нашому випадку OF складає 10,5, а далі виконуючи умову формули 3, отримуємо K_{OCF} на рівні 1,085.

Підставляючи всі розрахункові величини у нашу формулу 5 отримуємо значення T^* на рівні 2011,22 год.

Результати моделювання показали, що запропонована модель дозволяє оцінити трудомісткість проекту з більшою точністю порівняно з традиційним підходом.

Цей приклад є лише базовим шаблоном, який може бути адаптований під інший тип проекту з можливістю зміни вихідних даних в процесі оцінки трудомісткості проекту. Оптимальний вибір підходу залежить від конкретної ситуації. Для невеликих проектів з простими вимогами може бути достатньо традиційного підходу. Для великих і складних проектів більш доцільно використовувати запропоновану математичну модель. Можна комбінувати обидва підходи.

Застосування цих методів і підходів допоможе керівникам та командам розробки ІТ-проектів формувати вимоги, які відповідають потребам користувачів і можуть бути успішно реалізовані під час формування трудомісткості проектів та складання бюджетів вищого рівня.

5 РЕЗУЛЬТАТИ

Для демонстрації ефективності запропонованої моделі було проведено дослідження на прикладі розробки мобільного додатку відеоредактора Filmu. Маючи набір вихідних змінних по проекту та критерій оцінки, в статті автором було проведено порівняння двох варіантів оцінки трудомісткості робіт. Перший – традиційний, шляхом прямого підрахунку витрат годин на розробку від команди проекту та застосування до них коригуючих коефіцієнтів, що сформовані експертним шляхом на основі попереднього досвіду, T (формула 1). Другий – на основі математичної моделі, яка виражена формулою 5. Відобразимо та проаналізуємо отримані результати в обох варіантах (табл. 9) та проведемо обґрунтування доцільності застосування запропонованої модулі в системі управління ІТ-проектами на етапі.

В таблиці 10 наведені основні нормативи розподілу годин трудомісткості між етапами розробки ІТ-проекту, які використовувались для визначення загального значення показника $T_i(f)$.

Наведені в таблиці 9 та 10 дані є достатньо зрозумілими для аналізу та визначення трудомісткості проекту. Для розрахунку нормативи визначені експертним шляхом на основі досвіду роботи команд над іншими проектами. Однак така простота криє за собою ряд недоліків, зокрема відволікання розробників від безпосередньої їх роботи для проведення оцінки трудомісткості фіч, суб'єктивізм в оцінці, що впливає з набутого досвіду фахівців та

неможливості врахувати всі залежності в проекті та передбачити вплив такого фактору як зміна вимог користувача. Усунути зазначені недоліки допоможе запропонована в цій статті математична модель, яка є більш складною і сучасною, оскільки базується на методах use case та враховує зміни вимог користувачів та інші ризики розробки продукту. Використовуючи формулу (5) отримуємо такі результати (табл. 11).

Таблиця 9 – Трудомісткість робіт, $T_i(f)$ за стадіями розробки мобільного додатку відеоредактора Filmu (традиційний підхід)

Фіча, F	$T_i(f)$, год.	UI	Dev	QA	PM, BA	CR
Обрізка відео	273,6	32	120	45,6	22,8	53,2
Додавання відео із галереї	172,8	16	80	28,8	14,4	33,6
Додавання музики до відео	273,6	32	120	45,6	22,8	53,2
Додавання нового проекту	100,8	16	40	16,8	8,4	19,6
Додавання тексту	151,2	24	60	25,2	12,6	29,4
Зміна розміру кадру (4:3 або 16:9)	172,8	16	80	28,8	14,4	33,6
Експорт в різній якості	122,4	8	60	20,4	10,2	23,8
Публікація відео в соціальні мережі	136,8	16	60	22,8	11,4	26,6
Адміністрування системи	453,6	32	220	75,6	37,8	88,2
Хмарне сховище для файлів	151,2	24	60	25,2	12,6	29,4
Платіжна система	28,8		16	4,8	2,4	5,6
Налаштування телефону	356,4	48	150	59,4	29,7	69,3
Всього, T	2394	264	1066	399	199,5	465,5

Таблиця 10 – Нормативу розподілу трудомісткості за етапами розробки проекту

Етап	% від часу на розробку
Стабілізація системи	0,25
Тестування системи	0,3
Час менеджера	0,15
Комунікація в команді	0,1
Поставка готового коду	пів дня – день
Ризики	0,2

На основі математичних розрахунків отримали результати, які відображені в таблиці 11. Варто відзначити, що всі зазначені коефіцієнти вже були попередньо розраховані та усувають існуючі недоліки в традиційному підході оцінки трудомісткості проекту. Зокрема, K_{ACI} та K_{UCCF} – це коефіцієнти для опису функціональності системи, які враховують фактори зміни кількості акторів та use case та є змінними для T^* , G – комплексний показник, що враховує вплив організаційних та технічних факторів ризику на проект. В основу розрахунку покладено базовий показник продуктивності розробки на один

прийняття для окремої User story – це умови, які повинні бути виконані, щоб вважати історію користувача завершеною. Вони допомагають визначити, чи відповідає історія користувача потребам користувача і чи відповідає вона очікуванням команди розробки. Отже, фіча може складатися з однієї або декількох User Story, залежно від її масштабів. Це полегшує роботу команди та відповідний контроль над виконанням завдань проекту. Критерії прийняття можна уточнювати як для кожної User story, так і цілої фічі. Вибір залежить від масштабу, складності та деталізації завдання. Така формалізація та структура User story є достатньо простою і зрозумілою для всіх стейкхолдерів на стадіях user requirements та system requirements проекту.

Однак, проведений аналіз літературних джерел не виявив досліджень, які б застосовували такий формат опису вимог на рівні presale. Це пояснюється високою трудомісткістю та ресурсоемністю такої оцінки для компанії на даному етапі роботи із замовником. Тому очевидним стає застосування інших підходів для вирішення поставленої задачі. В нашому дослідженні вибір було зроблено на методі use case, який формалізовано в математичний шаблон для оцінки трудомісткості проекту на етапі пресеїлу.

Запропонований формалізм відкриває широкі перспективи для розробки більш точних і об'єктивних методів оцінки та математичних моделей, які дозволяють швидко, точно та ефективно визначати сценарії трудомісткості витрат проекту, особливо з урахуванням ризиків розробки та зміни умов користувача.

Отримані результати дослідження демонструють потенціал запропонованої математичної моделі для більш точної оцінки трудомісткості розробки програмного забезпечення, зокрема мобільних додатків. Порівняння з традиційним методом, заснованим на експертних оцінках, виявило ряд переваг запропонованого підходу, зокрема:

- математична модель усуває суб'єктивність, властиву експертним оцінкам, за рахунок використання кількісних показників та алгоритмів розрахунку;

- такий підхід до визначення вимог та трудомісткості проекту на етапі попередньої оцінки дозволяє враховувати широкий спектр факторів, що впливають на трудомісткість проекту, таких як зміна вимог, технологічні ризики та організаційна складність розробки функціоналу продукту;

- результати, отримані за допомогою запропонованої моделі на основі методу use case, демонструють більш високу точність порівняно з традиційним підходом;

- запропонована модель може бути легко застосована до інших проектів, забезпечуючи стабільність результатів.

Незважаючи на переваги, запропонована модель має певні обмеження. Наприклад, точність © Свінцицька О. М., Пулеко І. В., Граф М. С., Петросян Р. В., 2025
DOI 10.15588/1607-3274-2025-1-17

прогнозування залежить від якості вхідних даних та правильності калібрування моделі. Для більш точного прогнозування необхідно проводити додаткові дослідження та збирати більші обсяги даних про реальні проекти.

Перспективними напрямками подальших досліджень є:

- верифікація моделі на більшій вибірці проектів за типами та рівнями складності;

- перевірка стійкості моделі до різних типів проектів та масштабів, залежно від ризиків.

За проаналізованими факторами ризику, що наводяться в даній статті, проведено експертну оцінку їх впливу на трудомісткість проекту. На рисунку 5б зображені відповідна динаміка потенційних втрат годин трудомісткості.

Згідно з наведеним графіком, втрати часу через різні фактори ризику суттєво впливають на загальну трудомісткість проекту. Це означає, що непередбачені ситуації та проблеми, які виникають під час реалізації проекту, можуть призводити до значного збільшення часу, необхідного для його завершення. Оскільки в нашому дослідженні є сталий показник G на рівні 11067, в подальшому пропонується підходити до його розрахунку диференційовано в залежності від ймовірності настання в кожному окремому випадку і типу проекту.

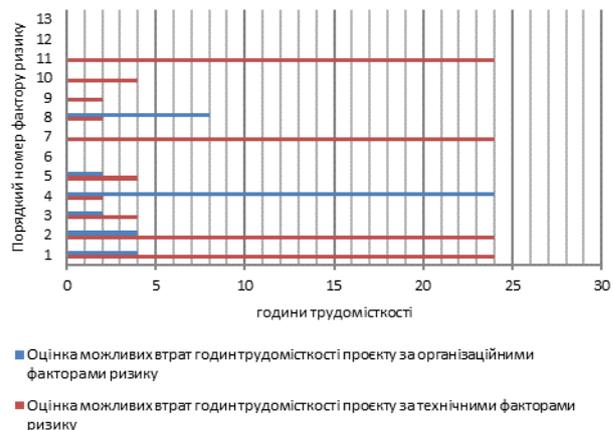


Рисунок 5 – Показники можливих втрат годин трудомісткості за факторами ризиків проекту

Для більш глибокого розуміння впливу різних факторів ризику на трудомісткість проекту необхідно проводити подальші дослідження та створювати кількісні моделі, які дозволять прогнозувати їх рівень та ймовірність настання. Це дозволить більш точно підходити до планування проектів, гнучко реагувати на запити користувачів та витратити менше ресурсів на їх оцінку.

ВИСНОВКИ

Вирішується актуальна проблема формування та оцінки вимог користувача в управлінні IT-проектами на основі методу Use Case в методології Agile.

Наукова новизна отриманих результатів полягає у наступному:

1) вперше запропоновано комплексний підхід до проблеми формування та оцінки вимог до Agile-проєкту на етапі планування трудомісткості проєкту, що базується на врахуванні факторів зовнішнього та внутрішнього впливу та поєднанні графічних, аналітичних та математичних інструментів, зокрема, діаграми use case та коригуючих коефіцієнтів, що враховують складність системи та вплив організаційних та технічних ризиків.

2) дістав подальший розвиток метод Use Case шляхом деталізації за визначеними критеріями акторів та варіантів використання (use case) на прості, середні та складні, яка виражена через коефіцієнти складності актора та use case з метою їх застосування під час розрахунку одного з показників ефективності проєкту – трудомісткості.

3) вперше запропоновано математичну модель формування вимог Agile проєкту на основі методу use case, результат якої виражено показником загальної трудомісткості, що дозволяє робити точні розрахунки тривалості проєкту та будувати різні сценарії бюджету за рахунок використання кількісних показників та алгоритмів розрахунку, які враховують такі фактори як: зміна вимог замовника, технологічні ризики та організаційну складність розробки функціоналу продукту.

Практична значимість дослідження полягає в тому, що авторський підхід до вирішення проблеми формування вимог в управлінні ІТ-проєктом, що базується на використанні математичної моделі, дозволяє отримати швидкі і більш точні результати оцінки трудомісткості порівняно з традиційним підходом, а також легко адаптуватись до інших типів проєктів, забезпечуючи стабільність результатів. Проведений розрахунок підтвердив доцільність її застосування на прикладі мобільного додатку на 16% економії трудовитрат, а отже здатність досягати більшої ефективності.

Перспективними напрямками подальших досліджень є перевірка стійкості моделі до різних типів проєктів та масштабів, залежно від впливу ризиків.

ПОДЯКА

Робота підтримана державним університетом «Житомирська політехніка». Госпдоговірна тема: «Рекомендації щодо удосконалення веб-сторінок сайту deps.ua» (реєстраційний номер: 0123U102487).

ЛІТЕРАТУРА

1. The Impact of Incomplete or Changing Requirements on IT Project Success. [Electronic resource]. – Access mode: <https://www.boardroommetrics.com/blog/the-impact-of-incomplete-or-changing-requirements-on-it-project-success-20131222.htm>
2. The state-of-practice in requirements specification: an extended interview study at 12 companies / [X. Franch,

- C. Palomares, C. Quer et al.] // Requirements Engineering. – 2023. – No. 28 (3). – P. 377-409. DOI:10.1007/s00766-023-00399-7
3. Solutions for software requirement risks using artificial intelligence techniques / [V.K.R. Reddy, U. Rahamathunnisa, P. Subhashini et al.] // Handbook of Research on Data Science and Cybersecurity Innovations in Industry 4.0 Technologies. – 2023. – P. 45–64. DOI: 10.4018/978-1-6684-8145-5
4. IEEE Standard Glossary for Software Engineering Terminology, ANSI/IEEE Std 610.12-1990. [Electronic resource]. – Access mode: <https://ieeexplore.ieee.org/document/159342>
5. Козак О. Л. Опорний конспект лекцій з курсу «Аналіз вимог до програмного забезпечення» для студентів напрямку підготовки «Програмна інженерія» / О. Л. Козак. – Тернопіль, 2011. – 56 с. [Електронний ресурс]. – Режим доступу: http://dspace.wunu.edu.ua/retrieve/14135/FCIT_kKN_sP_ZS_dAVPZ_%20LEC.pdf
6. Грицюк Ю. І. Особливості формулювання вимог до програмного забезпечення. [Електронний ресурс] / Ю. І. Грицюк, О. А. Немова. – Режим доступу: file:///C:/Users/%D0%9E%D0%BB%D0%B5%D0%BA%D1%81%D0%B0%D0%BD%D0%B4%D1%80%D0%B0/Downloads/Osoblivosti_formuluvanna_vimog_do_programnogo_zabe.pdf
7. ISO/IEC 25062. Software Engineering – Software product Quality Requirements and Evaluation (SQuaRE) – Common Industry Format (CIF) for Usability Test Reports (2006). [Electronic resource]. – Access mode: [Electronic resource]. – Access mode: <https://www.iso.org/standard/43046.html>
8. ISO 9241-210:2019. Ergonomics of human-system interaction – Human-centred design for interactive systems (2019). [Electronic resource]. – Access mode: <https://www.iso.org/standard/77520.html>
9. ISO/IEC/IEEE 29148. Systems and software engineering – Life cycle processes – Requirements engineering (2011). [Electronic resource]. – Access mode: <https://www.iso.org/standard/72089.html>
10. Kotonya G. Requirements Engineering: Processes and Techniques (1st ed.). / G. Kotonya, I. Sommerville // Wiley Publishing. – 1998. – P. 294.
11. Geis T. Specifying Usability Requirements and Test Criteria for Interactive Systems: Consequences for New Releases of Software-related Standards Within the ISO 9241 Series / T. Geis, W. Dzida, W. Redtenbacher // Federal Institute for Occupational Safety and Health. – 2004. – P.63.
12. Maguire M. User Requirements Analysis: A Review of Supporting Methods / M. Maguire, N. Bevan // In Proceedings of the IFIP 17th World Computer Congress – TC13 Stream on Usability: Gaining a Competitive Edge. – 2002. – P. 133–148. DOI:10.1007/978-0-387-35610-5_9
13. Hernande R. Requirements management in DevOps environments: a multivocal mapping study. / R. Hernandez, B. Moros, J. Nicolas // Requirements Engineering. – 2023.– 28 (3). – P. 317–346. Cited 1 time. DOI:10.1007/s00766-023-00396-w

14. Sathe C. A. An Empirical Study of Project Management Constraints in Agile Software Development: Multigroup Analysis between Scrum and Kanban / C. A. Sathe, C. Panse // *Brazilian Journal of Operations and Production Management*. – 2023. – Vol. 20, No. – P. 1–17. 3 special edition, <https://doi.org/10.14488/BJOPM.1796.2023>
15. Silvius G. Exploring the Project Owner's Behaviour of Addressing Sustainability in Project Assignment and Governance / G. Silvius, R. Ursem, J. Magano // *Sustainability*. – 2023. – No. 15(19). – P. 14294; <https://doi.org/10.3390/su151914294>

Стаття надійшла до редакції 05.12.2023
Після доробки 17.12.2024.

UDC 004.9

USE CASE METHOD IN IT PROJECT MANAGEMENT BASED ON AGILE METHODOLOGY

Svintsycka O. M. – PhD, Associate professor, Associate professor of the Educational Department of Computer Sciences, Zhytomyr Polytechnic State University, Zhytomyr, Ukraine.

Puleko I. V. – PhD, Associate professor, Associate professor of the Educational Department of Computer Engineering and Cyber Security, Zhytomyr Polytechnic State University, Zhytomyr, Ukraine.

Graf M. S. – PhD, Head of the Educational Department of Computer Sciences, Zhytomyr Polytechnic State University, Zhytomyr, Ukraine.

Petrosian R. V. – Senior Lecturer, Computer Sciences department, Zhytomyr Polytechnic State University, Zhytomyr, Ukraine.

ABSTRACT

Context. The article considers the role and process of forming user requirements based on the Use Case method in assessing the complexity of an Agile project at the stage of preliminary assessment by the company's management. Since the mid-70s, it has been known that errors in requirements are the most numerous, expensive, and time-consuming to correct in projects. In this regard, the importance of requirements management in IT projects using modern technologies and methods for their formation and evaluation is increasing.

Objective. Formation and evaluation of user requirements in IT project management based on the Use Case method and their impact on one of the project performance indicators at the planning stage, particularly labor intensity.

Method. The article proposes a new author's approach to the formation and evaluation of user requirements in Agile projects, taking into account the impact of risks and system complexity assessment based on the Use Case method, and as a result of the study and proposals to achieve this goal, a mathematical model for estimating project complexity is proposed. The mathematical template of the model allows us to consider additional variables that may affect the project, such as the number of user levels, available functionality, and technical and organizational risks. It is flexible and can be adapted to the different needs of a particular project, which aligns with the principles of the Agile methodology. The number of components in the formula can be changed to take into account the importance of different variables or expanded to take into account additional variables that may affect the project.

Results. A mathematical model for estimating project complexity based on the use case method has been developed and tested using the example of a mobile application, which contains a set of initial data for product development and constraints on changing user requirements and organizational and technical risks. The proposed mathematical model allows you to quickly, accurately, and efficiently determine scenarios of project labor intensity of various types and levels of complexity and can serve as an effective tool for making management decisions. A mathematical model for estimating project complexity based on the use case method has been developed and tested using the example of a mobile application, which contains a set of initial data for product development and constraints on changing user requirements and organizational and technical risks. The proposed mathematical model allows you to quickly, accurately, and efficiently determine scenarios of project labor intensity of various types and levels of complexity and can serve as an effective tool for making management decisions.

Conclusions. The general findings obtained after analyzing the methods of forming and evaluating user requirements in Agile management are as follows. At the work planning stage, based on an expert assessment of each functional requirement, the primary project evaluation model has been replaced by a more modern and complex one based on the use case method and considering changes in user requirements and other product development risks. The new model uses graphical, analytical, and mathematical tools, including a use case diagram, adjustment factors considering the complexity of the actor and use case, and factors considering organizational and technical risks. As a result, we get a mathematical format for calculating the project's complexity. This approach allows us to adapt to different types of projects quickly. With the correct initial data definition, the model will enable us to obtain reasonably accurate estimates early in project planning. The practical results of the study demonstrate the potential of the proposed mathematical model, which can be logically continued by verifying the model on a larger sample and assessing its resilience to different types of projects and risks.

KEYWORDS: user requirements, IT product, Agile project, User Story, Use Case, mobile application.

REFERENCES

1. The Impact of Incomplete or Changing Requirements on IT Project Success. [Electronic resource]. Access mode: <https://www.boardroommetrics.com/blog/the-impact-of-incomplete-or-changing-requirements-on-it-project-success-20131222.htm>
2. Franch X., Palomares C., Quer C., Chatzipetrou P., Gorschek T. The state-of-practice in requirements specification: an extended interview study at 12 companies, *Requirements Engineering*, 2023, No. 28 (3), pp. 377–409. DOI:10.1007/s00766-023-00399-7
3. Reddy V. K. R., Rahamathunnisa U., Subhashini P., Aancy H. M., Meenakshi S., Boopathi S. Solutions for software requirement risks using artificial intelligence techniques, *Handbook of Research on Data Science and Cybersecurity Innovations in Industry 4.0 Technologies*, 2023, pp. 45–64. DOI: 10.4018/978-1-6684-8145-5
4. IEEE Standard Glossary for Software Engineering Terminology, ANSI/IEEE Std 610.12-1990. [Electronic resource]. Access mode: <https://ieeexplore.ieee.org/document/159342>
5. Kozak O. L. Opornyi konspekt lektsii z kursu «Analiz vymoh do prohramnoho zabezpechennia» dlia studentiv napriamku pidhotovky «Prohramna inzheneriia». Ternopil, 2011, 56 p. [Elektronnyi resurs]. Rezhym dostupu: http://dspace.wunu.edu.ua/retrieve/14135/FCIT_kKN_sP_ZS_dAVPZ_%20LEC.pdf
6. Hrytsiuk Yu.I., Niemova O.A. Osoblyvosti formuluvannia vymoh do prohramnoho zabezpechennia. [Elektronnyi resurs]. Rezhym dostupu: file:///C:/Users/%D0%9E%D0%BB%D0%B5%D0%BA%D1%81%D0%B0%D0%BD%D0%B4%D1%80%D0%B0/Downloads/Osoblyvosti_formuluvanna_vimog_do_prohramnogo_zabe.pdf
7. ISO/IEC 25062. Software Engineering – Software product Quality Requirements and Evaluation (SQuaRE) – Common Industry Format (CIF) for Usability Test Reports (2006). [Electronic resource]. Access mode: [Electronic resource]. Access mode: <https://www.iso.org/standard/43046.html>
8. ISO 9241-210:2019. Ergonomics of human-system interaction – Human-centred design for interactive systems (2019). [Electronic resource]. Access mode: <https://www.iso.org/standard/77520.html>
9. ISO/IEC/IEEE 29148. Systems and software engineering – Life cycle processes – Requirements engineering (2011). [Electronic resource]. Access mode: <https://www.iso.org/standard/72089.html>
10. Kotonya G., Sommerville I. Requirements Engineering: Processes and Techniques (1st ed.). Wiley Publishing, 1998, P. 294.
11. Geis T., Dzida W., Redtenbacher W. Specifying Usability Requirements and Test Criteria for Interactive Systems: Consequences for New Releases of Software-related Standards Within the ISO 9241 Series, *Federal Institute for Occupational Safety and Health*, 2004, P.63.
12. Maguire M., Bevan N. User Requirements Analysis: A Review of Supporting Methods, *In Proceedings of the IFIP 17th World Computer Congress – TC13 Stream on Usability: Gaining a Competitive Edge*, 2002, pp. 133–148. DOI:10.1007/978-0-387-35610-5_9
13. Hernande R., Moros B., Nicolas J. Requirements management in DevOps environments: a multivocal mapping study, *Requirements Engineering*, 2023, No. 28 (3), pp. 317–346. Cited 1 time. DOI:10.1007/s00766-023-00396-w
14. Sathe C. A., Panse C. An Empirical Study of Project Management Constraints in Agile Software Development: Multigroup Analysis between Scrum and Kanban, *Brazilian Journal of Operations and Production Management*, 2023, Vol. 20, No, pp. 1–17. 3 special edition, <https://doi.org/10.14488/BJOPM.1796.2023>.
15. Silvius G., Ursem R., Magano J. Exploring the Project Owner’s Behaviour of Addressing Sustainability in Project Assignment and Governance, *Sustainability*, 2023, No. 15(19), P. 14294; <https://doi.org/10.3390/su151914294>

ЗАСТОСУВАННЯ БІНАРНОГО ДЕРЕВА ПОШУКУ З ФІКСОВАНОЮ ВИСОТОЮ ДЛЯ ПРИСКОРЕННЯ ОБРОБКИ ОДНОВИМІРНИХ МАСИВІВ

Шпортко О. В. – канд. техн. наук, доцент, доцент кафедри інформаційних систем та обчислювальних методів ПВНЗ «Міжнародний економіко-гуманітарний університет імені академіка Степана Дем'янчука», Рівне, Україна.

Бомба А. Я. – д-р техн. наук, професор, професор кафедри комп'ютерних наук та прикладної математики Національного університету водного господарства та природокористування, Рівне, Україна.

АНОТАЦІЯ

Актуальність. На сьогодні для прискорення пошуку, сортування та відбору елементів масивів широко використовуються бінарні дерева пошуку. Але обчислювальна складність пошуку з використанням бінарного дерева пропорційна його висоті, яка, в свою чергу, залежить від послідовності опрацювання елементів масиву. Для зменшення висоти дерева періодично виконують його балансування, яке є тривалим процесом, тому розробка альтернативних способів контролю за висотою бінарного дерева є на сьогодні актуальним науковим завданням.

Мета. Розробка принципів та відповідних алгоритмів формування та використання бінарного дерева з фіксованою висотою для прискорення пошуку елемента в масиві та визначення довільної i -ї порядкової статистики, зокрема, медіани масиву.

Метод. В дослідженні запропоновано встановлювати фіксовану висоту бінарного дерева пошуку на одиницю більшою від мінімально можливої висоти бінарного дерева для розміщення всіх елементів масиву, адже збільшення фіксованої висоти призводить до зайвих витрат оперативної пам'яті, а зменшення – сповільнює модифікації дерева. Формування таких дерев подібне до балансування дерев, але, на відміну від нього, рекурсивне переміщення вузлів у них виконується лише тоді, коли відповідне піддерево заповнене повністю. Для бінарного дерева пошуку з фіксованою висотою оперативна пам'ять виділяється один раз при його створенні – відразу під всі можливі вузли бінарного дерева заданої висоти. Це дає змогу уникнути операцій виділення та звільнення пам'яті під кожен вузол дерева та зберігати значення вузлів в одновимірному масиві без використання вказівок.

Результати. Наші експерименти показали, що для прискорення пошуку елементів та визначення i -тих порядкових статистик часто змінюваних неупорядкованих масивів доцільно додатково формувати бінарне дерево пошуку з фіксованою висотою. Для ініціалізації цього дерева доцільно використати відсортовану копію ключів елементів масиву, а не вставляти їх по чергово. Використання бінарного дерева з фіксованою висотою прискорює, наприклад, пошук медіан таких масивів більш ніж в 7 разів відносно методу двох бінарних пірамід та додатково прискорює перерозподіл стиснутих даних між модифікованими DEFLATE-блоками в процесі прогресуючого ієрархічного стиснення без втрат зображень набору АСТ в середньому на 2,92%.

Висновки. Для визначення медіан чи i -тих порядкових статистик окремих непов'язаних масивів та підмасивів замість відомих методів сортування доцільно використовувати розбиття Hoare з обміном на великих відстанях, оскільки воно переставляє лише окремі елементи, а не впорядковує весь масив повністю. З метою визначення медіан послідовності вкладених підмасивів, впорядкованих за зростанням їх довжини, варто застосовувати метод двох бінарних пірамід, адже вони орієнтовані на швидке доповнення новими елементами. Для знаходження медіан чи i -тих порядкових статистик після змін чи вилучень елементів неупорядкованого масиву доцільно використати бінарне дерево пошуку ключів елементів масиву з фіксованою висотою, оскільки таке фіксування запобігає неконтрольованому зростанню кількості операцій порівняння та дає змогу обробляти дерево без використання вказівок.

КЛЮЧОВІ СЛОВА: сортування масивів, медіани масиву, метод двох бінарних пірамід, бінарне дерево пошуку з фіксованою висотою.

АБРЕВІАТУРИ

АСТ – Archive Comparison Test, тестовий набір зображень;

BST – Binary Search Tree, бінарне дерево пошуку;

RAM – Random Access Memory, оперативна пам'ять;

БД – База Даних.

НОМЕНКЛАТУРА

$countHBlock$ – початкова кількість результуючих блоків кожного модифікованого DEFLATE-блоку;

$countLevelTree$ – кількість рівнів BST з фіксованою висотою;

$countLeft_i$ – кількість заповнених вузлів зліва для i -го вузла BST;

$countRight_i$ – кількість заповнених вузлів справа для i -го вузла BST;

$countSub_i$ – кількість створених підпорядкованих вузлів зліва і справа для i -го вузла BST;

$indexArray$ – індекс елемента в масиві для вузла BST;

$indexDel$ – індекс елемента для вилучення з відсортованого масиву ключів;

$indexInsert$ – індекс елемента для вставки у відсортований масив ключів;

$\log(N)$ – логарифм числа N за основою 2;

N – кількість елементів вхідного масиву;

$tree_i$ – i -й вузол BST з фіксованою висотою;

$indexLeft_i$ – індекс підпорядкованого вузла зліва для i -го вузла BST;
 $valueArray$ – значення елемента з масиву для вузла BST;
 $valueNode$ – структура вузла BST;
 X – вхідний масив для сортування;
 x_i – i -й елемент масиву X для сортування;
 Y_i – i -й елемент послідовності неперервних підмасивів, впорядкованих за зростанням довжини, який містить елементи з x_0 по елемент x_i ;
 Z_i – i -й елемент послідовності неперервних підмасивів, впорядкованих за спаданням довжини, який містить елементи з x_0 по елемент x_i .

ВСТУП

Як відомо, на сьогодні для прискорення пошуку, сортування та відбору елементів масивів широко використовуються бінарні дерева пошуку (BST) [1]. Нагадаємо, що таке дерево – це пов'язаний ациклічний орієнтований граф, який складається з кореневого вузла та лівого і правого піддерев, що не перетинаються між собою [2]. В кожному вузлі дерева міститься ключ, який складається зі значення одного з елементів масиву та індекса цього елемента в масиві. Зліва від кожного вузла BST містяться вузли з меншими, а справа – з більшими значеннями ключів. Таке дерево дає змогу реалізувати аналог бінарного пошуку замість лінійного, адже після порівняння ключа пошуку з ключем чергового вузла цей процес рекурсивно продовжується в одному з двох напрямків (зліва чи справа), а вузли в іншому напрямку відкидаються [2]. Принципи обробки елементів масиву за допомогою BST широко використовуються в індексах таблиць БД [3; 4].

Обчислювальна складність пошуку з використанням BST пропорційна його висоті [1], яка, в свою чергу, залежить від послідовності опрацювання елементів масиву при формуванні і вставці вузлів. Тому для зменшення висоти дерева періодично виконують його балансування [2], яке є тривалим процесом. Отже, розробка алгоритмів для автоматичного підтримання висоти дерева, близької до мінімальної, є на сьогодні актуальним науковим завданням.

Бінарне дерево пошуку також дає змогу швидко віднайти найменше чи найбільше значення в масиві, на які вказують, відповідно, найлівіший чи найправіший вузол дерева відносно його кореня. Поряд з цим, бінарне дерево пошуку несуттєво прискорює пошук довільної i -ї порядкової статистики, зокрема медіани масиву [5], адже це дерево впорядковує елементи, але не вказує місце кожного елемента у відсортованому масиві. Значення зазначеного недоліку BST не варто недооцінювати, адже на сьогодні в статистиці та соціології для аналізу ряду величин замість середнього арифметичного, мінімального чи максимального значення серед всіх його елементів все ширше використовують саме його медіану (величину, що розташова-

на всередині ряду величин, розміщених у зростаючому або спадному порядку). Медіана ділить ряд значень ознаки на дві рівні частини [6; 7], тому для масивів з непарною кількістю елементів – це значення центрального елемента, а з парною – середнє арифметичне двох центральних елементів, якщо елементи масиву відсортувати. Елементи масиву можуть змінюватися з часом і щоразу формувати відсортовану копію масиву чи проходитися по частині його елементів для знаходження медіани або i -ї порядкової статистики недоцільно. Саме тому прискорення визначень i -ї порядкової статистики загалом і медіани масиву зокрема, в тому числі за допомогою BST, теж є на сьогодні актуальним науковим завданням.

Отже, **об'єктом** цього дослідження є механізми формування та використання бінарного дерева пошуку для прискорення обробки одновимірних масивів. **Предметом дослідження** є засоби реалізації та використання бінарного дерева пошуку з фіксованою висотою. **Мета дослідження** – розробка принципів та відповідних алгоритмів формування та використання бінарного дерева з фіксованою висотою для прискорення пошуку елемента в масиві та визначення i -ї порядкової статистики, зокрема, медіани масиву.

1 ПОСТАНОВКА ЗАДАЧІ

Нехай потрібно віднайти медіану вхідного масиву $X = \langle x_0, x_1, \dots, x_{N-1} \rangle$ з N елементів. Очевидно, що в процесі визначення медіани доведеться використовувати додаткову пам'ять для сортування елементів копії масиву чи зберігання динамічних структур (пірамід, бінарного дерева) щоб не спотворити початковий масив. Зрозуміло також, що для визначення медіани масиву можливо використати стандартний метод сортування оболонки мови програмування (наприклад, для C# [8] – це метод `Array.Sort()`) чи один з швидких методів сортування [9] та обрати елементи, які після цього опиняться посередині. Але таке визначення буде тривалим, оскільки під час сортування будуть впорядковуватися всі елементи вхідного масиву, а нам потрібно лише дізнатися, які значення опиняться після сортування посередині. Ми проаналізуємо ефективність різних методів визначення медіани масиву не лише однократно, а й після послідовної зміни кожного елемента.

Дослідимо також ефективність різних методів для визначення медіан впорядкованої за зростанням довжини послідовності неперервних підмасивів $Y_i = \langle x_0, x_1, \dots, x_{i-1}, x_i \rangle$, $i = \overline{0, N-1}$, де наступний підмасив отримується з попереднього підмасиву за допомогою доповнення його новим елементом. Очевидно, що визначення медіани чергового підмасиву прискориться, якщо використати впорядковані дані попереднього підмасиву, доповнивши їх новим елементом, і тут ефективними можуть виявитися інші методи, ніж для визначення медіани масиву X .

Крім цього, проаналізуємо ефективність різних методів прискорення обробки елементів масивів для

подібної до Y_i послідовності вкладених підмасивів $Z_i = \langle x_0, x_1, \dots, x_{i-1}, x_i \rangle$, $i = \overline{N-1, 0}$, впорядкованої за спаданням їх довжини. Тут наступний підмасив отримується з попереднього за допомогою вилучення його останнього елемента. Такі послідовності використовуються нами в процесі прогресуючого ієрархічного стиснення зображень без втрат [10], під час якого після кожного вилучення максимального елемента масиву ще й змінюються значення двох інших його елементів, що призводить до модифікації пов'язаних динамічних структур.

2 ОГЛЯД ЛІТЕРАТУРИ

На практиці для визначення медіани масиву найчастіше застосовують розбиття С. А. R. Hoare [11]. Цей метод діє за принципом «Розділяй і володарюй»: серед елементів копії масиву обирається опорний елемент і всі елементи, не більші за нього, переміщуються в масиві лівіше від цього елемента, а не менші – правіше. Якщо після перестановок опорний елемент опинився посередині масиву (для масивів з непарною довжиною) чи в одній з центральних позицій (для масивів з парною довжиною), то обчислюють медіану масиву з участю цього опорного елемента. Інакше, якщо опорний елемент виявився правіше від центру, то аналогічно продовжують розбиття лівіше від нього. В протилежному випадку виконують розбиття елементів, розміщених справа від опорного елемента [1]. Елементи, розміщені з іншого боку від опорного елемента, на відміну від швидких алгоритмів сортування [11], надалі не аналізуються, тому середня обчислювальна складність такого алгоритму пошуку медіан менше $O(N \times \log(N))$. На сьогодні також розроблені інші методи пошуку медіани масиву, які мають теоретичну обчислювальну складність порядку $O(N)$ [1], але на практиці їх реалізації виявляються набагато складнішими від розбиття Ноаге. Всі ці методи орієнтовані на однократний пошук медіани і результати їх роботи несуттєво впливають на прискорення наступних пошуків медіан після зміни значень елементів масиву.

З іншого боку, розбиття Ноаге не завжди є найшвидшим варіантом для пошуку медіан підмасивів. Наприклад, для впорядкованої за зростанням довжини послідовності неперервних підмасивів Y_i , які на початку містять однакові елементи, найефективнішим є метод двох бінарних пірамід [12]. Цей метод послідовно формує дві піраміди однакового розміру, причому перша з них – незростаюча, містить менші елементи вхідного масиву, а друга – неспадна, містить більші елементи. Тому медіани вкладених підмасивів обчислюються тут лише за допомогою верхин цих пірамід і, можливо, чергового елемента, якщо він ще не має пари і не включений в піраміди. Метод двох бінарних пірамід орієнтований суто на пошук медіан підмасивів з послідовним збільшенням кількості їх елементів і неефективний при частих змінах значень елементів

масиву чи для визначення довільної i -ї порядкової статистики.

Для визначення послідовності медіан чи i -тих порядкових статистики вхідного масиву X після кожної зміни значень його елементів на практиці використовують відсортований масив ключів елементів масиву, в який синхронно вносять зміни з використанням бінарного пошуку [9] за таким алгоритмом [12]:

1. Віднайти у відсортованому масиві ключів бінарним пошуком індекс ключа попереднього значення елемента вхідного масиву та записати його у змінну *indexDel*;
2. Віднайти у відсортованому масиві ключів бінарним пошуком індекс ключа, більшого за ключ нового значення елемента вхідного масиву та записати його у змінну *indexInsert*;
3. Якщо *indexInsert* > *indexDel*, то перемістити ключі з позиції *indexDel*+1 по позицію *indexInsert*-1 на один елемент вліво та вставити ключ нового значення елемента у відсортований масив ключів в позицію *indexInsert*-1;
4. Інакше перемістити ключі з позиції *indexDel*-1 по позицію *indexInsert* на один елемент вправо та вставити ключ нового значення елемента у відсортований масив ключів в позицію *indexInsert*.

Приклад синхронних змін масиву ключів для вхідного масиву X наведено на рис. 1. Для спрощення в масиві ключів тут наведені лише відсортовані значення елементів без їх індексів у вхідному масиві, оскільки вони унікальні. У випадку заміни елемента 18 на 60 у відсортованому масиві ключів спочатку шукається індекс попереднього значення ключа (змінна *indexDel*), потім – індекс більшого ключа від ключа нового значення (змінна *indexInsert*), далі ключі з позиції *indexDel*+1 по позицію *indexInsert*-1 переміщуються на один елемент вліво, після чого в позицію *indexInsert*-1 вставляється нове значення ключа (60).

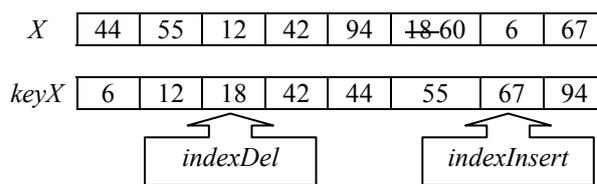


Рисунок 1 – Зміни у відсортованому масиві ключів для масиву $X = \langle 44, 55, 12, 42, 94, 18, 6, 67 \rangle$ після заміни елемента 18 на 60

Відсортований масив ключів дає змогу відразу визначити будь-яку i -ту порядкову статистику, хоча й вимагає повсякчасного переміщення ключів між позиціями видалення та вставки. Як зазначалося в [12] і буде підтверджено в цій статті, переміщення елементів лише між цими позиціями, а не двічі від кожної з цих позицій до кінця масиву, суттєво прискорює визначення порядкових статистик.

У цій статті ми деталізуємо новий метод визначення i -х порядкових статистик масивів та підмасивів за допомогою BST з фіксованою висотою та досліджуємо

мо його ефективність для визначення медіан. Вперше цей метод був представлений на конференції [13]. В цій праці ми використаємо даний метод також для прискорення обробки елементів одновимірних вкладених впорядкованих за спаданням довжини масивів Z_i в процесі прогресуючого ієрархічного стиснення зображень без втрат [10]: в процесі перерозподілу стиснутих даних між модифікованими DEFLATE-блоками [14] виконується ітеративний вибір та поєднання двох суміжних результуючих блоків, які максимально зменшують довжину стиснутих даних. Після поєднання результуючих блоків до кожного з них окремо застосовується контекстно-незалежне кодування [15]. Чим більше результуючих блоків згенеровано на початку – тим триваліше буде ітеративне поєднання. Під час кожної такої ітерації необхідно:

1. Віднайти два суміжні результуючі блоки, які максимально зменшують довжину стиснутих даних (максимально економлять біти для зберігання стиснутих даних);
2. Поєднати ці блоки в єдиний результуючий блок;
3. Перерахувати економії бітів від поєднання отриманого блоку з попереднім та наступним блоками.

Розміри результуючих блоків до поєднання	5007	6580	6020	4840	3000	7000	
Економії від поєднання суміжних блоків	...	12	42	94	18	6	...

Рисунок 2 – Фрагмент розмірів результуючих блоків та прогнозованих зменшень (економій) від поєднання суміжних блоків умовного вхідного блоку на початку чергової ітерації, бітів

Розміри результуючих блоків після поєднання	5007	6580	10766	3000	7000	
Економії від поєднання суміжних блоків	...	12	62	8	6	...

Рисунок 3 – Фрагмент розмірів результуючих блоків та прогнозованих зменшень (економій) від поєднання суміжних блоків умовного вхідного блоку в кінці чергової ітерації, бітів

Після кожної ітерації кількість результуючих блоків зменшується на 1 за умови, що існує ще хоча б одна невід’ємна економія бітів від поєднання суміжних блоків. Нехай початкова кількість результуючих блоків Z_i рівна $countHBlock$. Тоді максимальна кількість таких ітерацій становить $countHBlock - 1$. Якщо під час кожної ітерації виконується лінійний пошук максимальної економії від поєднання суміжних блоків, то загальна кількість порівнянь для таких пошуків у найгіршому випадку становитиме $(countHBlock - 2) \times (countHBlock - 1) / 2$, тобто маємо квадратичну обчислювальну складність. Чим менші результуючі блоки генеруються під час початкового розбиття вхідного блоку – тим більша їх однорідність, але й тим більше їх буде, і тому довше триватиме їх ітеративне поєднання. Отже, в цій статті розглянемо також варіанти прискорення пошуку максимального зменшення розміру коду від поєднання суміжних результуючих блоків та модифікації цих зменшень внаслідок такого поєднання за допомогою як відсортованого масиву ключів і бінарних пошуків в ньому, так і BST з фіксованою висотою.

Розглянемо, наприклад, характеристики фрагменту результуючих блоків для умовного вхідного блоку на початку і в кінці чергової ітерації (рис. 2, 3). Зменшення від поєднання суміжних блоків на цих рисунках схематично виведено між суміжними блоками, до яких воно стосується. На початку чергової ітерації максимальне зменшення розміру стиснутих даних (максимальна економія) від поєднання суміжних результуючих блоків (94 біти, виведені на рис. 2 на темно-сірому фоні) може бути досягнуто від поєднання третього та четвертого блоків цього фрагменту (виведені на світло-сірому фоні). Після поєднання цих результуючих блоків (рис. 3) їх загальна кількість зменшилася на один, розмір блоку поєднання зменшився на 94 біти відносно суми розмірів суміжних поєднаних блоків ($10766=6020+4840-94$, виведений на світло-сірому фоні), але ще й змінилися значення економій від поєднань блоку поєднання з суміжними блоками (42 на 62 та 18 на 8). Причому, як бачимо, значення економій бітів може як збільшитися, якщо суміжні результуючі блоки мають близькі нерівномірності розподілів, так і зменшитися, якщо ці нерівномірності розподілів суттєво різняться.

3 МАТЕРІАЛИ І МЕТОДИ

Розглянемо особливості реалізації операцій вставки та видалення для BST з фіксованою висотою, пов’язаного з одновимірним масивом, адже пошук та сортування з використанням цього дерева виконуються так само, як і для класичного BST [1]. В кожному вузлі бінарного дерева для його зв’язку з масивом збережемо не лише значення одного з елементів масиву $valueArray$, а й поле $indexArray$, яке вказує на індекс цього елемента в масиві. Ці два поля формують структуру складеного ключа вузла дерева:

```
struct valueNode
{int valueArray;
 int indexArray;};
```

Відсутність значення в корені дерева позначимо індексом -1 . Порівняння ключів двох вузлів виконаємо за відповідними значеннями елементів масиву, а коли вони однакові – за індексами цих елементів в масиві. Мовою C# [8] така функція порівняння ключів може бути записана так:

```
int cmpNode(valueNode a, valueNode b)
{if (a.valueArray < b.valueArray ||
    (a.valueArray == b.valueArray &&
     a.indexArray < b.indexArray)) return -1;
if (a.valueArray > b.valueArray ||
    (a.valueArray == b.valueArray &&
     a.indexArray > b.indexArray)) return 1;
return 0; }
```

Оскільки в ключ вузла входить його індекс в масиві, то різні елементи масиву ніколи не будуть породжувати однакові ключі, що дає змогу однозначно ідентифікувати їх по вузлах дерева.

Вставляти вузли з елементів масиву чи підмасиву та їх індексів в бінарне дерево будемо, як правило, традиційно [2] – перший вузол записуємо в корінь дерева, а всі інші опрацюємо послідовно та рекурсивно: якщо ключ чергового вузла більший значення ключа з поточного вузла дерева, то вставку рекурсивно продовжуємо в праве піддерево, коли ж ключ чергового вузла менший за ключ поточного вузла, то вставку рекурсивно продовжуємо в ліве піддерево. Рекурсивні виклики виконуємо до тих пір, поки поточний вузол дерева не виявиться порожнім, куди й включаємо черговий вузол. Приклад бінарного дерева, сформованого для вхідного масиву $X = \langle 44, 55, 12, 42, 94, 18, 6, 67 \rangle$, наведено на рис. 4. У вузлах дерева тут і надалі наводяться лише значення елементів масиву, оскільки кожне з них є унікальним і дає змогу однозначно визначити індекс цього елемента в масиві. Бачимо, що перший елемент масиву обов'язково потрапляє у вершину дерева, а положення інших елементів залежить від значень попередніх елементів масиву. З елементів цього масиву сформувалося чотирьохрівневе дерево, хоча, якби елементи масиву були впорядковані за зростанням чи спаданням, то кількість рівнів сягнула б восьми.

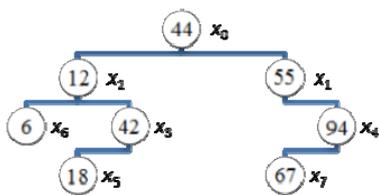


Рисунок 4 – Бінарне дерево, послідовно заповнене елементами масиву $X = \langle 44, 55, 12, 42, 94, 18, 6, 67 \rangle$

Оскільки обчислювальна складність основних операцій над BST пропорційна його висоті [1], яка, в свою чергу, залежить від послідовності опрацювання елементів масиву при формуванні і вставці вузлів, то для прискорення пошуку, вставки і видалення вузлів з бінарного дерева додатково обмежимо його висоту значенням $countLevelTree$ (саме тому таке дерево називається бінарним деревом пошуку з фіксованою висотою). Більші значення висоти дерева прискорюють вставку елементів в нього (оскільки з'являється більше вільних вузлів), хоча й збільшують обсяги оперативної пам'яті для його зберігання.

© Шпортко О. В., Бомба А. Я., 2025
DOI 10.15588/1607-3274-2025-1-18

Змінну $countLevelTree$ визначатимемо перед використанням BST з фіксованою висотою. Оскільки в бінарне дерево з висотою $countLevelTree$ можна поміститися максимум $2^{countLevelTree} - 1$ вузли, то значення цієї змінної підберемо так, щоб загальна кількість вузлів дерева була не меншою кількості елементів масиву ($2^{countLevelTree} - 1 \geq N$). Ми встановлювали змінну $countLevelTree$ на одиницю більшою від мінімального можливого значення:

$$countLevelTree = \lceil \log(N + 1) \rceil + 1. \quad (1)$$

З метою уникнення зайвих операцій з виділення та звільнення пам'яті для окремих елементів відразу відведемо пам'ять під всі можливі вузли бінарного дерева висоти $countLevelTree$, тобто під всі $2^{countLevelTree} - 1$ вузли, більшість з яких залишатимуться вільними. Тоді зберігати значення вузлів бінарного дерева можна в масиві: зліва від елемента $tree_i$ завжди буде міститися елемент $tree_{2i+1}$, а справа – $tree_{2i+2}$. Зрозуміло, що ключ вершини дерева буде міститися в $tree_0$. Змінну $countLevelTree$ та кількість вузлів дерева можна збільшувати в процесі експлуатації, якщо BST виявиться заповненим. Для прискорення обробки дерева створимо також масив $indexLeft$, в якому для індекса кожного вузла збережемо індекс підпорядкованого вузла зліва: $indexLeft_i = 2 * i + 1$. Використовуючи цей масив, індекс підпорядкованого вузла справа буде більшим від індекса підпорядкованого вузла зліва на одиницю.

Бінарне дерево пошуку з фіксованою висотою подібне до збалансованого дерева [2], але в ньому відразу зарезервовано місце під всі можливі вузли, а операція балансування (обернення дерева) зводиться до переміщення значень вузлів без модифікацій вказівок на піддерева.

Приклад чотирьохрівневого BST з фіксованою висотою, заповненого елементами вхідного масиву $X = \langle 44, 55, 12, 42, 94, 18, 6, 67 \rangle$, наведено на рис. 5.

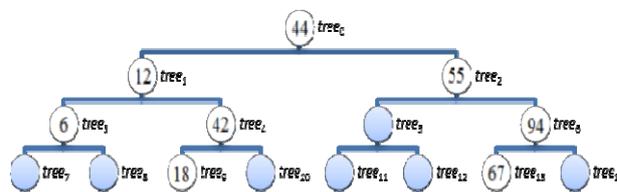


Рисунок 5 – Чотирьохрівневе бінарне дерево з фіксованою висотою, послідовно заповнене елементами масиву $X = \langle 44, 55, 12, 42, 94, 18, 6, 67 \rangle$

Вставка нових значень в бінарне дерево з фіксованою висотою сповільнюється, якщо в обраному напрямку (вліво чи вправо) відсутні вільні вузли. Таке буває, якщо ключ нового значення дерева з чергового елемента масиву менший ключа з поточного вузла елемента, а зліва вільні вузли відсутні, або коли ключ нового значення більший ключа з поточного вузла, а

справа немає незайнятих вузлів. Тоді у першому з цих випадків звільнимо поточний вузол, вставивши його значення вправо, та порівняємо ключ нового значення з найбільшим значенням ключа зліва (міститься зліва направо). Якщо ключ нового значення більший від найбільшого значення ключа зліва, то вставимо в поточний вузол черговий елемент, інакше ж перенесемо в поточний вузол елемент з максимальним значенням ключа зліва та рекурсивно продовжимо вставку чергового елемента відносно вузла зліва (це можливо, бо в цьому напрямку ми тільки що звільнили один вузол). Для другого випадку аналогічні дії виконуються дзеркально. Припустимо, наприклад, що в масиві X наступним елементом (під індексом 8) є число 17 і його потрібно вставити в бінарне дерево з фіксованою висотою, наведене на рис. 5. Тоді, рухаючись від вершини бінарного дерева, цей ключ буде послідовно порівнюватися зі значеннями вузлів 44, 12 та 42. Число 17 відносно 42 менше, тому мало б рекурсивно вставлятися відносно вузла зліва (там міститься 18), але зліва вільні вузли відсутні. Тому значення 42 рекурсивно вставляється відносно вузла справа, тим самим звільнюючи поточний вузол, в поточний вузол вставляється найбільше значення зліва (тобто, 18), а зліва рекурсивно вставляється черговий елемент 17 (рис. 6).

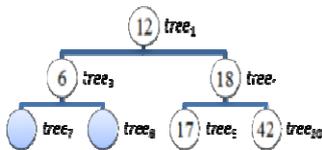


Рисунок 6 – Ліва частина чотирьохрівневого бінарного дерева, сформованого для масиву $X = \langle 44, 55, 12, 42, 94, 18, 6, 67 \rangle$ після доповнення його наступним елементом 17

Такі переміщення значень вузлів бінарного дерева забезпечують дотримання основного принципу його побудови: вузли справа від поточного вузла містять більші, а зліва – менші значення ключів.

З метою ідентифікації випадків відсутності вільних вузлів збережемо для кожного вузла BST з фіксованою висотою в цілочисельних масивах $countLeft$ та $countRight$ відповідно кількості заповнених вузлів зліва і справа, а в масиві $countSub$ – кількість створених підпорядкованих вузлів. Зрозуміло, що значення елементів цих масивів для листків дерева рівне нулю. Співставлення $countLeft_i$ та $countRight_i$ з $countSub_i$ відразу дає змогу визначити можливість вставки значень вузлів без переміщень в обраному напрямку.

Вилучення елементів з BST з фіксованою висотою виконується також традиційно – якщо ключ вузла для вилучення менше ключа чергового вузла, то вилучення виконується рекурсивно відносно підпорядкованого вузла зліва, якщо більше – то відносно вузла справа. Якщо ж вони рівні, то значення вузла замінюється найменшим значенням серед вузлів справа, а при його відсутності – найбільшим серед вузлів зліва.

Побудовані таким чином процедури вилучення і вставки вузлів дають змогу не перевищувати задану висоту BST, а підрахунок кількостей зайнятих вузлів зліва і справа дозволяють швидко не лише вставляти і вилучати елементи, а й визначати будь-яку його i -ту порядкову статистику [5]. Функцію для такого визначення (її перший аргумент – номер порядкової статистики, а другий не вказується) мовою C# подамо так:

```
valueNode valueTree(int index, int startIndex=0)
{if (index < countLeft[startIndex]) // значення зліва
  return valueTree(index, indexLeft[startIndex]);
  index-=countLeft[startIndex]; // пропускаємо зліва
  if (index == 0) // якщо значення з поточного вузла
    return tree[startIndex];
  // переходимо до вузла справа
  return valueTree(index-1, indexLeft[startIndex]+1);}
```

Таким чином, для визначення довільної порядкової статистики, пошуку та вилучення елемента в BST з фіксованою висотою потрібно не більше $2 * countLevelTree(1)$ порівнянь, а кількість порівнянь при вставці залежить від розміщення вільних вузлів в дереві і може мати значно більшу обчислювальну складність. Крім цього, при вставці вузлів в дерево потрібно щоразу корегувати кількості зайнятих вузлів зліва чи справа. Отже, найтривалішим етапом при використанні такого дерева є його початкове заповнення. Тому у випадках, коли всі елементи початкового одновимірного масиву відомі до внесення у нього змін, відсортований масив ключів з елементів та індексів цього масиву доцільно відразу рекурсивно записати в дерево пошуку з фіксованою висотою за принципом: в поточний вузол дерева заносимо ключ, що міститься посередині масиву чи підмасиву; ключі підмасиву зліва від цього елемента рекурсивно записуємо в дерево відносно вузла зліва; ключі підмасиву справа від цього елемента рекурсивно записуємо в дерево відносно вузла справа. Такий підхід дає змогу відразу вказати кількості зайнятих вузлів дерева зліва і справа та не виконувати зайвих порівнянь, хоча й вимагає попереднього сортування ключів елементів початкового масиву. Для подальших корегувань початкового масиву потрібно буде лише дописувати та вилучати з дерева відповідні ключі. Зауважимо також, що вилучення попередніх та вставку нових ключів під час зміни елементів вхідного масиву можливо виконувати разом, адже якщо ці два ключі одночасно більші від ключа поточного вузла, то цю операцію можливо рекурсивно продовжити відносно вузла справа, якщо менші – то відносно вузла зліва і при цьому кількості зайнятих вузлів зліва і справа відносно поточного вузла змінювати не потрібно.

4 ЕКСПЕРИМЕНТИ

Перевірку ефективності різних методів однократного визначення медіан масивів ми виконували над масивами X з 10 млн. дійсних чисел. Ці експерименти

дають змогу оцінити швидкість **побудови** додаткових масивів чи динамічних структур (пірамід або дерев) для знаходження цього показника.

Для визначення послідовності медіан ми також послідовно обробляли 10 тис. вкладених підмасивів Y_i , впорядкованих за зростанням довжини. Такі тести дозволяють співставити швидкості **доповнення** додаткових структур при реалізації різних методів визначення медіан масивів.

Швидкості **модифікацій** додаткових масивів чи динамічних структур з метою визначення медіан масивів після частих змін їх елементів ми оцінювали на масивах X з 5 тис. елементів, послідовно вносячи в них 5 тис. змін. Тобто після ініціалізації вхідного масиву ми послідовно змінювали значення кожного його елемента і після кожної зміни визначали медіану масиву.

Для оцінки ефективності **видалення** та модифікацій додаткових масивів та дерев, що використовуються для прискорення обробки елементів одновимірних масивів, ми порівняли час поєднання результуючих блоків з використанням впорядкованих за спаданням довжини масивів економії бітів Z_i в процесі прогресуючого ієрархічного стиснення зображень відомого

тестового набору АСТ. Завантажити ці зображення можна, наприклад, з <http://www.compression.ca/act/act-files.html>. Даний набір містить як синтезовані (№№ 1, 2, 7) так і фотореалістичні (решта) зображення. Вибір саме цього тестового набору зумовлений різноплановістю його зображень та наявністю у відкритих джерелах результатів тестувань на ньому алгоритмів інших дослідників.

Алгоритми розглянутих методів ми реалізували в оболонці Microsoft Visual Studio мовою програмування C# [8]. Для реалізації швидкого сортування ми використали стандартний метод оболонки програмування *Array.Sort()*. Тестування проводилося на комп'ютері з процесором Intel Pentium 4 з тактовою частотою 3 GHz та розміром RAM 4Gb.

5 РЕЗУЛЬТАТИ

Результати тестування алгоритмів різних методів визначення медіан для різних варіантів масивів наведені в табл. 1–3. Тривалість визначення медіан з використанням принципів сортування для масивів, відсортованих за спаданням, як правило більша, ніж для масивів, відсортованих за зростанням, оскільки ці алгоритми орієнтовані на сортування за зростанням.

Таблиця 1 – Тривалості визначень медіан масивів з 10 млн. дійсних чисел алгоритмами різних методів, мс

Метод визначення медіани	Варіант масиву			
	Згенерований випадковим чином	Відсортований за зростанням	Відсортований за спаданням	З однакових елементів
Швидке сортування	3906	1328	2153	2086
Розбиття Hoare	625	273	328	602
Бінарні включення по масиву ключів	> 22.8 млн.	6992	> 24 млн.	6953
Дві бінарні піраміди	5718	9521	8833	5030
Бінарне дерево з фіксованою висотою	> 59 млн.	> 345 млн.	> 345 млн.	> 148 млн.
Бінарне дерево по відсортованому масиву	5146	3151	3827	2669

Таблиця 2 – Тривалості знаходжень медіан 10 тис. вкладених підмасивів дійсних чисел алгоритмами різних методів, мс

Метод визначення медіани	Використання попередніх впорядкованих даних	Варіант масиву			
		Згенерований випадковим чином	Відсортований за зростанням	Відсортований за спаданням	З однакових елементів
Швидке сортування	Ні	11414	4714	6605	6145
	Так	4868	3877	6233	5498
Розбиття Hoare	Ні	4118	2678	2477	2516
	Так	1070	1063	1102	828
Бінарні включення по масиву ключів	Так	250	16	469	23
Дві бінарні піраміди	Так	23	16	16	16
Бінарне дерево з фіксованою висотою	Так	510	3420	3704	35

Таблиця 3 – Тривалості визначень медіан масивів дійсних чисел з 5 тис. елементів після послідовної модифікації кожного елемента алгоритмами різних методів, мс

Метод визначення медіани	Використання попередніх впорядкованих даних	Варіант масиву			
		Згенерований випадковим чином	Відсортований за зростанням	Відсортований за спаданням	З однакових елементів
Швидке сортування	Ні	6296	3882	4799	4540
	Так	3496	2792	2958	2606
Розбиття Hoare	Ні	2400	502	674	709
	Так	642	332	432	498
Бінарні включення по масиву ключів	Так	340	235	329	126
Дві бінарні піраміди	Так	268	469	470	14
Бінарне дерево з послідовними вилученнями та вставками елементів	Так	38	2198	2341	16
Бінарне дерево з одночасними вилученнями і вставками елементів	Так	36	2216	2331	13

Таблиця 4 – Час перерозподілу стиснутих даних між модифікованими DEFLATE-блоками в процесі прогресуючого ієрархічного стиснення зображень набору АСТ з використанням різних варіантів алгоритмів прискорення цього процесу, мс

Варіант алгоритму прискорення перерозподілу стиснутих даних між модифікованими DEFLATE-блоками	№ файла								Середній час
	1	2	3	4	5	6	7	8	
Лінійний пошук максимальної економії бітів	330	120	330	470	340	640	30	480	343
Бінарні включення по масиву ключів змінених економій бітів з переміщенням елементів від позицій вилучення та вставки до кінця масиву	260	80	200	220	160	330	30	270	194
Бінарні включення по масиву ключів змінених економій бітів з переміщенням елементів між позиціями вилучення та вставки	240	80	170	220	140	300	20	230	175
Бінарне дерево пошуку з фіксованою висотою та початковою ініціалізацією вставками	250	80	190	250	130	330	30	220	185
Бінарне дерево пошуку з фіксованою висотою та початковою ініціалізацією з відсортованого масиву	220	80	160	200	140	280	20	220	165
Бінарне дерево пошуку з фіксованою висотою та поєднанням вилучень і вставок економій бітів	230	80	170	230	160	270	20	230	174

Час перерозподілу стиснутих даних між модифікованими DEFLATE-блоками в процесі прогресуючого ієрархічного стиснення зображень тестового набору АСТ з використанням різних методів прискорення обробки одновимірних масивів наведено в табл. 4.

6 ОБГОВОРЕННЯ

Проаналізуємо спочатку тривалості визначень однократних медіан алгоритмами різних методів (табл. 1). Бачимо, що, як і передбачалося, для однократних визначень медіан окремих великих масивів доцільно використовувати розбиття Hoare. Воно швидше за алгоритми сортування, оскільки не впорядковує елементи масиву повністю, та швидше за алгоритми методів двох бінарних пірамід [12] і описаного в цій статті і в [13] бінарного дерева пошуку з фіксованою висотою, бо не формує ієрархічні структури, а лише переміщує окремі елементи. Формування описаного тут BST по невпорядкованому масиву з 10 млн. дійсних чисел взагалі триває понад 16 год., оскільки вимагає повсякчасних переміщень елементів для забезпечення його фіксованої висоти. Тому початкове бінарне дерево слід формувати по відсортованому масиву, а це не раціонально для однократного визначення медіани, адже після сортування масиву цю характеристику можна визначити відразу. Цікаво, що визначення медіани після бінарних включень ключів для масиву, відсортованого за зростання, триває в 3432 рази швидше, ніж для масиву, відсортованого за спаданням, адже для першого з цих масивів бінарне включення кожного ключа не призводить до переміщення всіх попередніх елементів масиву, а для другого – переміщує всі попередні елементи на один вправо.

Співставимо тепер тривалості визначень медіан 10 тис. підмасивів дійсних чисел, впорядкованих за зростанням їх довжини, алгоритмами різних методів (табл. 2). Бачимо, що для визначення медіан вкладених підмасивів Y_i доцільно використати впорядковані дані попередніх підмасивів, а не опрацьовувати їх спочатку. Формування описаного тут бінарного дерева пошуку для підмасивів, впорядкованих за зростанням чи спаданням триває довше, ніж для підмасивів, згенерованих випадковим чином, оскільки супроводжується постійними переміщеннями елементів для забезпечення його фіксованої висоти. Як і прогнозувалося, найефективнішим та найстабільнішим для визначення медіан таких підмасивів виявився метод

двох бінарних пірамід, оскільки для визначення всіх медіан він в середньому використовує лише $N \log N$ порівнянь та стільки ж присвоєнь.

Розглянемо також тривалості знаходжень 5 тис. медіан масивів дійсних чисел такого ж розміру після модифікації кожного елемента (табл. 3). Як і для вкладених підмасивів бачимо, що для визначення медіани масиву після зміни чергового елемента варто відкоригувати попередні впорядковані дані (для алгоритмів сортування – його відсортований аналог, для розбиття Hoare – результати перестановок попереднього масиву, для двох бінарних пірамід та дерева пошуку з фіксованою висотою – ці самі ієрархічні структури для попередніх масивів), а не обробляти елементи щоразу спочатку. Алгоритм методу двох бінарних пірамід показує для невпорядкованих масивів не найкращі результати, оскільки при пошуку елемента для вилучення він може переглядати практично повністю одну з двох пірамід (половину масиву). Найефективнішим методом для визначення медіан після зміни кожного елемента масиву, згенерованого випадковим чином, як і для масиву з однаковими елементами, виявився метод BST з фіксованою висотою та одночасними вилученнями і вставками ключів елементів, адже він в середньому виконує лише $2 \times N \times \text{countLevelTree}$ порівнянь. Побудова такого дерева по відсортованому масиву ключів прискорює визначення медіан більш ніж в 7 разів відносно методу двох бінарних пірамід. З іншого боку, для масивів, впорядкованих за зростанням чи спаданням, найшвидше визначення медіан продемонстрував метод бінарного включення по масиву ключів, адже він в середньому використовує лише $2 \times N \times \log N$ порівнянь і переміщує елементи між позиціями вилучення та вставки без складнішого корегування ієрархічних структур.

На завершення співставимо тривалості перерозподілу стиснутих даних між модифікованими DEFLATE-блоками в процесі прогресуючого ієрархічного стиснення зображень без втрат (табл. 4). Бачимо, що використання бінарного включення ключів змінених економій бітів в масивах Z_i замість їх лінійного пошуку дало змогу прискорити перерозподіл в середньому на 48.98% (перший та третій рядки табл. 4). З них 5.54% досягається за допомогою переміщення ключів елементів між позиціями вилучення і

вставки, а не до кінця масиву (другий рядок цієї таблиці). Бінарне дерево пошуку з фіксованою висотою дає змогу додатково прискорити перерозподіл ще на 2.92%, але за умови його початкової ініціалізації з відсортованого масиву (п'ятий рядок табл. 4). Програмне поєднання операцій вилучень і вставки модифікованих економій бітів (шостий рядок даної таблиці) не принесло очікуваного ефекту, оскільки воно призводить до додаткових порівнянь, а відповідні значення найчастіше потрапляють в різні гілки дерева. Зважаючи на ефективність BST з фіксованою висотою, ми рекомендуємо використовувати на практиці саме цю динамічну структуру для визначення i -тих порядкових статистик після змін окремих елементів невпорядкованих масивів.

ВИСНОВКИ

1. Наші дослідження вчоргове підтвердили недоцільність пошуку «універсальних методів» визначення медіан чи i -тих порядкових статистик, ефективних для всіх послідовностей масивів чи підмасивів. Зокрема, при виборі такого методу слід зважати на кількості елементів, якими відрізняються масиви.

2. Для визначення медіан чи i -тих порядкових статистик окремих непов'язаних масивів та підмасивів найефективнішим, як і очікувалося, виявилось розбиття Ноаге з обміном на великих відстанях, оскільки воно переставляє лише окремі елементи, а не впорядковує масиви повністю.

3. З метою визначення медіан послідовності з N вкладених підмасивів, впорядкованих за зростанням їх довжин, варто застосовувати метод двох бінарних пірамід, адже його реалізації для всіх таких підмасивів в середньому виконують лише $N \log N$ порівнянь та стільки ж присвоєнь, хоча цей метод не придатний для визначення довільних i -тих порядкових статистик. Знаходження цих статистик для вкладених підмасивів варто виконувати за допомогою бінарних включень по масиву ключів.

4. Для знаходження медіан чи i -тих порядкових статистик після змін чи вилучення елементів невпорядкованого масиву доцільно використовувати бінарне дерево ключів з фіксованою висотою. З цією метою слід спочатку відсортувати ключі елементів масиву, побудувати по них це дерево і після цього послідовно виконувати пов'язані модифікації елементів вхідного масиву та дерева пошуку з фіксованою висотою.

Надалі ми плануємо дослідити ефективність застосування BST з фіксованою висотою в процесі прогресуючого ієрархічного стиснення зображень без втрат після застосування різницевої кольірних моделей [16].

ЛІТЕРАТУРА

1. Introduction to Algorithms, Fourth Edition / [T. H. Cormen, C. E. Leiserson, R. L. Rivest, C. Stein]. – Cambridge : MIT Press, 2022. – 1312 p. – Access mode: <http://mitpress.mit.edu/9780262046305/introduction-to-algorithms>.
2. Шаховська Н. Б. Алгоритми і структури даних : Навчальний посібник / Н. Б. Шаховська, Р. О. Голошук. – Львів : Магнолія-2006, 2020. – 214 с.

3. Silberschatz A. Database system concepts, Seventh edition / A. Silberschatz, H. F. Korth, S. Sudarshan. – New York : McGraw-Hill Education, 2020. ISBN 978-1-260-08450-4.
4. Kerttu P-M. B+-trees [Electronic resource] / P-M. Kerttu. – The Department of Computer Science University of Helsinki. – 9 p. – Access mode: <https://www.cs.helsinki.fi/u/mluukkai/tirak2010/B-tree.pdf>.
5. Gentle J. E. Computational Statistics / J. E. Gentle. – Springer, 2009. – 749 p. ISBN 9780387981444. DOI: 10.1007/978-0-387-98144-4.
6. Ageel M. I. The Mean-Median-Mode Inequality for Discrete Unimodal Probability Measure / M. I. Ageel // Far East Journal of Mathematical Sciences. – 2000. – № 2 (2). – P. 187–192.
7. Abadir K. M. The mean-median-mode inequality: counter examples / K. M. Abadir // Econometric Theory. – 2005. – № 21 (02). – P. 477–482. DOI: 10.1017/S0266466605050267.
8. C# 8.0 draft specification [Electronic resource]. – Microsoft Corporation, 2024. 3769 p. – Access mode: <https://learn.microsoft.com/en-us/dotnet/csharp/language-reference/language-specification/introduction>.
9. Knuth D. E. The Art of Computer Programming, Vol. 3. Sorting and Searching, Second edition / D. E. Knuth. – Massachusetts : Addison Wesley Longman, 1998. – 791 p.
10. Bomba A. Ya. Redistribution of the Compressed Data Between Modified DEFLATE-Blocks in the Image Compression Process Without Lossless / A. Ya. Bomba, A. V. Shportko, V. A. Postolatii // Computational Linguistics and Intelligent Systems (COLINS 2024) : Proceedings of the 8th International Conference (Lviv, 12–13 april, 2024). Volume II: Modeling, Optimization, and Controlling in Information and Technology Systems Workshop (MOCITSW). – CEUR Workshop Proceedings, 2024. – Vol. 2604. – P. 145–156. Access mode: <https://ceur-ws.org/Vol-3668/paper11.pdf>.
11. Hoare C. A. R. Quicksort / C. A. R. Hoare // Computer Journal. – 1962. – № 5 (1). – P. 10–16. DOI: 10.1093/comjnl/5.1.10.
12. Shportko A. The Acceleration of the Determination of the Median of Nested Subarrays Using Two Binary Pyramids / A. Shportko, V. Shportko // Computational Linguistics and Intelligent Systems (COLINS 2020) : Proceedings of the 4th International Conference (Lviv, Ukraine, 23–24 april, 2020). – CEUR Workshop Proceedings, 2020. – Vol. 2604. – P. 1102–1116. Access mode: <http://ceur-ws.org/Vol-2604/paper72.pdf>.
13. Shportko O. The Use of a Fixed Height Binary Tree to Accelerate the Calculation of the Medians of Subarrays / O. Shportko, L. Shportko // Computer Science and Information Technologies (CSIT 2020) : Proceedings of the XVth International Scientific and Technical Conference (Zbarazh, Ukraine, 23–26 septembr, 2020). – Springer Cham, 2020. – Vol. 2. – P. 46–49. Access mode: <https://ieeexplore.ieee.org/document/9321921>.
14. Deutsch P. DEFLATE Compressed Data Format Specification version 1.3. RFC 1951 / P. Deutsch. – Alladin enterprises, 1996. – 15 p. Access mode: <https://www.rfc-editor.org/rfc/rfc1951>. DOI: 10.17487/rfc1951.
15. Kotha H. D. Review on Lossless Compression Techniques / H. D. Kotha, M. Tummanapally, V. K. Upadhyay // Journal of Physics. – 2019. – Vol. 1228. DOI: 10.1088/1742-6596/1228/1/012007.
16. Shportko A. V. Programming the Formation of Difference Color Models for Lossless Image Compression / A. V. Shportko, A. Ya. Bomba, V. A. Postolatii // Computational Linguistics and Intelligent Systems (COLINS 2023) : Proceedings of the 7th International Conference (Kharkiv, Ukraine, 20–21 april, 2023). – CEUR Workshop Proceedings, 2023. – Vol. 3. – P. 53–68. Access mode: <http://ceur-ws.org/Vol-3403/paper5.pdf>.

Received 12.08.2024.
Accepted 05.12.2024.

APPLICATION OF BINARY SEARCH TREE WITH FIXED HEIGHT TO ACCELERATE PROCESSING OF ONE-DIMENSIONAL ARRAYS

Shportko A. V. – PhD, Associate Professor, Associate Professor of the Department of Information Systems and Computing Methods of the Private Higher Educational Institution “International University of Economics and Humanities named after Academician Stepan Demianchuk”, Rivne, Ukraine.

Bomba A. Ya. – Dr. Sc., Professor, Professor of the Department of Computer Sciences and Applied Mathematics of the National University of Water and Environmental Engineering, Rivne, Ukraine.

ABSTRACT

Topicality. Nowadays, binary search trees are widely used to speed up searching, sorting, and selecting array elements. But the computational complexity of searching using a binary tree is proportional to its height, which depends on the sequence of processing the elements of the array. In order to reduce the height of a tree, its balancing is periodically carried out, which is a long process, thus, the development of alternative methods of controlling the height of a binary tree is currently an actual scientific task.

Objective. Development of algorithms for the formation and use of a binary tree with a fixed height to accelerate the search for an element in an array and to determine arbitrary i -th order statistics, in particular, the median of the array.

Method. In this study, it is proposed to set the fixed height of the binary search tree by one greater than the minimum possible height of the binary tree to accommodate all the elements of the array because increasing the fixed height leads to excessive RAM consumption, and decreasing it slows down tree modifications. The formation of such trees is similar to the balancing of trees but, unlike it, the recursive movement of nodes in them is performed only when the corresponding subtree is completely filled. For a binary search tree with a fixed height, RAM is allocated once when it is created, immediately under all possible nodes of a binary tree with a given height. This allows to avoid allocating and freeing memory for each node of the tree and store the values of the nodes in a one-dimensional array without using pointers.

The results. Our experiments showed that in order to speed up the search of elements and to determine the i -th order statistics of frequently changing unordered arrays, it is advisable to additionally form a binary search tree with a fixed height. To initialize this tree, it is advisable to use a sorted copy of the keys of the array elements, and not to insert them one by one. For example, the use of a binary tree with a fixed height accelerates the search of medians of such arrays by more than 7 times compared to the method of two binary pyramids and additionally accelerates the redistribution of compressed data between modified DEFLATE-blocks in the process of progressive hierarchical lossless compression of images of the ACT set by an average of 2.92%.

Conclusions. To determine medians or i -th order statistics of individual unrelated arrays and subarrays, instead of known sorting methods, it is advisable to use Hoare partitioning with exchange over long distances as it rearranges only individual elements and does not order the entire array completely. In order to determine the medians of the sequence of nested subarrays, ordered by the growth of their length, it is worth using the method of two binary pyramids because they are oriented to rapid addition of new elements. To find medians or i -th order statistics after changes or removal of elements of an unordered array, it is advisable to use a binary search tree for the keys of array elements with a fixed height as such fixing prevents uncontrolled growth of the number of comparison operations and makes it possible to process the tree without using instructions.

KEYWORDS: array sorting, array medians, method of two binary pyramids, binary search tree with fixed height.

REFERENCES

1. Cormen T. H., Leiserson C. E., Rivest R. L., Stein C. . Introduction to Algorithms, Fourth Edition. Cambridge, MIT Press, 2022, 1312 p. Access mode: <http://mitpress.mit.edu/9780262046305/introduction-to-algorithms>.
2. Shakhovska N. B., Holoshchuk R. O. Alhorytmy i struktury danykh : Navchalnyi posibnyk. Lviv, Mahnoliia-2006, 2020, 214 p.
3. Silberschatz A., Korth H. F., Sudarshan S. Database system concepts, Seventh edition. New York, McGraw-Hill Education, 2020. ISBN 978-1-260-08450-4.
4. Keritu P-M. B+-trees [Electronic resource] / P-M. Keritu. – The Department of Computer Science University of Helsinki, 9 p. Access mode: <https://www.cs.helsinki.fi/u/mluukkai/tirak2010/B-tree.pdf>.
5. Gentle J. E. Computational Statistics. Springer, 2009, 749 p. ISBN 9780387981444. DOI: 10.1007/978-0-387-98144-4.
6. Ageel M. I. The Mean-Median-Mode Inequality for Discrete Unimodal Probability Measure, *Far East Journal of Mathematical Sciences*, 2000, № 2 (2), pp. 187–192.
7. Abadir K. M. The mean-median-mode inequality: counter examples, *Econometric Theory*, 2005, № 21 (02), pp. 477–482. DOI: 10.1017/S0266466605050267.
8. C# 8.0 draft specification [Electronic resource]. – Microsoft Corporation, 2024, 3769 p. Access mode: <https://learn.microsoft.com/en-us/dotnet/csharp/language-reference/language-specification/introduction>.
9. Knuth D. E. The Art of Computer Programming, Vol. 3. Sorting and Searching, Second edition. Massachusetts, Addison Wesley Longman, 1998, 791 p.
10. Bomba A. Ya., Shportko A. V., Postolatii V. A. Redistribution of the Compressed Data Between Modified DEFLATE-Blocks in the Image Compression Process Without Lossless, *Computational Linguistics and Intelligent Systems (COLINS 2024) : Proceedings of the 8th International Conference (Lviv, 12–13 april, 2024)*. Volume II: Modeling, Optimization, and Controlling in Information and Technology Systems Workshop (MOCITSW). CEUR Workshop Proceedings, 2024, Vol. 2604, pp. 145–156. Access mode: <https://ceur-ws.org/Vol-3668/paper11.pdf>.
11. Hoare C. A. R. Quicksort, *Computer Journal*, 1962, № 5 (1), pp. 10–16. DOI: 10.1093/comjnl/5.1.10.
12. Shportko A., Shportko V. The Acceleration of the Determination of the Median of Nested Subarrays Using Two Binary Pyramids, *Computational Linguistics and Intelligent Systems (COLINS 2020) : Proceedings of the 4th International Conference (Lviv, Ukraine, 23–24 april, 2020)*, CEUR Workshop Proceedings, 2020, Vol. 2604, pp. 1102–1116. Access mode: <http://ceur-ws.org/Vol-2604/paper72.pdf>.
13. Shportko O., Shportko L. The Use of a Fixed Height Binary Tree to Accelerate the Calculation of the Medians of Subarrays, *Computer Science and Information Technologies (CSIT 2020) : Proceedings of the XVth International Scientific and Technical Conference (Zbarazh, Ukraine, 23–26 september, 2020)*. Springer Cham, 2020, Vol. 2, pp. 46–49. Access mode: <https://ieeexplore.ieee.org/document/9321921>.
14. Deutsch P. DEFLATE Compressed Data Format Specification version 1.3. RFC 1951. Alladin enterprises, 1996, 15 p. Access mode: <https://www.rfc-editor.org/rfc/rfc1951>. DOI: 10.17487/rfc1951.
15. Kotha H. D., Tummanapally M., Upadhyay V. K. Review on Lossless Compression Techniques, *Journal of Physics*, 2019, Vol. 1228. DOI: 10.1088/1742-6596/1228/1/012007.
16. Shportko A. V., Bomba A. Ya., Postolatii V. A. Programming the Formation of Difference Color Models for Lossless Image Compression, *Computational Linguistics and Intelligent Systems (COLINS 2023) : Proceedings of the 7th International Conference (Kharkiv, Ukraine, 20–21 april, 2023)*. CEUR Workshop Proceedings, 2023, Vol. 3, pp. 53–68. Access mode: <http://ceur-ws.org/Vol-3403/paper5.pdf>.

УПРАВЛІННЯ У ТЕХНІЧНИХ СИСТЕМАХ

CONTROL IN TECHNICAL SYSTEMS

UDC 51-74:004

APPLICATION OF SINGULAR SPECTRAL ANALYSIS IN CONTROL SYSTEMS OF TECHNOLOGICAL PROCESSES AND EXPLOSION SAFETY CONTROL OF FACILITIES

Holinko O. V. – Postgraduate student of the Department of Computer Systems Software, Dnipro University of Technology, Dnipro, Ukraine.

Alekseev M. O. – Dr. Sc., Professor, Head of the Department of Computer Systems Software, Dnipro University of Technology, Dnipro, Ukraine.

Holinko V. I. – Dr. Sc., Professor, Head of the Department of Labor Protection and Civil Safety, Dnipro University of Technology, Dnipro, Ukraine.

Zabelina V. A. – Postgraduate student of the Department of Labor Protection and Civil Safety, Dnipro University of Technology, Dnipro, Ukraine.

ABSTRACT

Context. The question of increasing the productivity of technological processes of extraction, processing and preparation of raw materials, improving product quality, reducing energy consumption, as well as creating safe working conditions during technological processes and preventing accidents is always quite relevant and requires the implementation of modern control and management systems. For the effective operation of such systems, it is important to pre-process and filter the data received from the sensors for monitoring the grinding processes and the explosive status of objects. One of the possible ways to increase the informativeness of data is the use of singular spectral analysis.

Objective. Increasing the efficiency of technological process control systems and the reliability of explosive control systems of coal mines and oil and fuel complex facilities by processing and pre-filtering data received from sensors for monitoring grinding processes and the state of facilities.

Method. To analyze the output signals of sensors used in control and management systems, the method of singular spectral analysis is used, which allows revealing hidden structures and regularities in time series by pre-filtering and data processing of acoustic, thermocatalytic, and semiconductor sensors.

Results. A new approach to the management of technological processes of grinding raw materials in jet mills and control of the explosiveness of coal mines and objects of the oil and fuel complex is proposed, based on methods that allow to speed up the processing speed of sensor output data and improve the quality of information. It is shown that one of the promising methods that can be used for the pre-processing of time series of output data of sensors in control and control systems is the method of singular spectral analysis, the use of which allows filtering data, revealing hidden structures and regularities, and forecasting changes based on the analysis of previous information, identify anomalies and unusual situations, make more informed decisions and improve the processes of managing technological processes.

Conclusions. The conducted experiments have confirmed the proposed software operability and allow recommending it for use in advancing both theoretical and practical aspects of process control systems through an enhanced singular spectral analysis (SSA) method for time series processing. This improved approach has been successfully demonstrated in real-world applications, including grinding processes in jet mills and explosion monitoring in coal mines and oil and fuel facilities. The implementation demonstrates a significant increase in data processing speed and information quality, which makes it particularly valuable for use in safety-critical industrial facilities.

KEYWORDS: control systems, explosion control, sensors, singular spectral analysis, time series.

ABBREVIATIONS

SSA is a singular spectrum analysis.

NOMENCLATURE

$\tilde{F}^{(k)}$ is an output row;

F is a material time series;

f_i is an element of the time series;

F_n is a restored matrix;

g_k is an intermediate result after diagonal averaging;

i is a i -th index;

j is a j -th index;

K is a sequence of multidimensional vectors obtained through nesting;

L is a length of the window;

m is a number of non-intersecting subsets;
 n is a maximum non-zero number from the eigenvalues of the matrix S ;
 N is a quantity;
 S is a result of the singular decomposition of the trajectory matrix;
 U_i is an orthonormal system of matrix eigenvectors S ;
 V_i is an intermediate result after singular decomposition;
 x_{ij} is an element of the matrix;
 Y is a $L \times K$ matrix;
 λ_i is an eigenvalue of matrix S ;
 X is a matrix composed of nesting vectors.

INTRODUCTION

Object control and management systems are extremely widely used in industry [1]. They are used to automate production processes, optimize energy consumption, monitor and diagnose equipment, automate warehouses and logistics, ensure workplace safety, control product quality, and manage production resources. These systems help businesses increase productivity, reduce costs and ensure the reliability of production processes, making them more competitive in the market. Application in industry covers a wide range of areas and tasks [2].

There are many types of control and management systems, each with its own advantages and disadvantages, and the choice of a specific type depends on specific tasks and requirements. The main elements of control and management systems are sensors [3]. Nowadays, sensors have become an integral part of our everyday life, surrounding us in various forms and fields of application. Their variety and popularity are explained by their unique capabilities and advantages. Today's sensors are capable of measuring various parameters such as temperature, pressure, humidity, light, speed, sound level, biometrics and much more. They can also operate in a variety of environments, including extreme temperatures, high pressures, and aggressive chemical environments, making them versatile tools in a variety of industries [4].

In terms of speed, sensors are able to transmit information in real time, which is critical in many situations. In industrial processes, the speed of operation of sensors allows you to quickly adjust production parameters and minimize losses. Accuracy of measurements also plays a fundamental role [4]. Sensors in today's world are not only ordinary devices, but also key components that ensure safety, efficiency and convenience in various fields of human activity. Their speed and measurement accuracy contribute to the improvement of our lives and production, making them an integral part of technological progress.

The speed and accuracy of sensors play an important role in many areas of human activity, from industrial processes and medical diagnostics to scientific research and security. These two parameters are fundamental characteristics that determine the usefulness and efficiency of sensors [5]. Sensor speed is critical for automatic control and safety systems. The accuracy of the

sensors is also of great importance, especially in areas where small errors can have serious consequences. Combining high accuracy with speed can give the best result, especially in areas where it is necessary to respond quickly to changes in the environment and at the same time ensure high accuracy of measurements. These sensor characteristics play a key role in increasing productivity, ensuring safety, and improving the quality of life in today's world [6].

Sensors monitor various parameters of technological processes, which allows to increase the productivity of the equipment and the quality of products. Accurate measurements of the concentration of explosive gases and vapors are important to ensure the safety of personnel, they can quickly detect gas and fuel leaks and warn of danger, allowing immediate measures to be taken to prevent accidents [7,8]. In these areas, the speed and accuracy of sensors save lives, ensure product quality, and contribute to maintaining the overall level of safety. Considering this, ensuring high accuracy and speed of measurement is an urgent task, the solution of which depends on the final result.

1 PROBLEM STATEMENT

Let the output signal of the sensor represent a time series of numerical data $F = (f_0, \dots, f_{N-1})$, for example, a chaotic set of acoustic noises that arise during the mechanical interaction of high-energy gas jets with material particles. In order to control the object, it is necessary to select the information components $F^{(k)} = (f_0^{(k)}, \dots, f_{N-1}^{(k)})$ that characterize its mode of operation from the input numerical series. The task is to find effective methods of processing the output signals of the sensors, which allow in real time to carry out preliminary filtering of the data received from the sensors, to isolate the information components necessary for the management of the control objects, to reveal hidden structures and regularities in the time series under the conditions of execution requirements for ensuring the speed of technological process control systems and explosion protection.

2 REVIEW OF THE LITERATURE

Various types of sensors are used in technological process control systems, the output signal of which requires pre-processing to identify the information components that are used to control the process. For example, acoustic sensors are used to control technological processes of grinding raw materials in jet mills [9, 10]. The output signal of these sensors is a chaotic set of acoustic noises that arise during the mechanical interaction of high-energy gas jets with particles of the material to be further crushed. Directly, without preliminary processing of such a signal, it is almost impossible to detect informational signs that characterize the mode of operation of the mill. Therefore, various methods and algorithms for analyzing the signal received from the acoustic sensor are used to analyze the

output signal. One of the approaches is the use of fuzzy logic [11, 12]. The use of fuzzy logic algorithms for analysis allows you to get a fairly accurate result with a small error of 10%, however, in this approach, the fuzzy logic model is very sensitive to both the quality of the input raw materials and the parameters of the stirrup mill. Thus, the use of a trained model of fuzzy logic allows obtaining a clear result only with optimized input parameters [13]. In [14], the use of the discrete fractional Fourier transform of the discrete type is proposed for the analysis of parameters, the article considers various methods of using this analysis. When they are used on different parts of the signal, these methods show a fairly high accuracy of parameter estimation, but the disadvantage of this method is the need for experimental selection of the method of using the discrete fractional Fourier transform both for different signals and for different parts of the same signal.

Thermocatalytic sensors are widely used in explosiveness monitoring systems [15]. This is due to the high selectivity of such sensors, their low sensitivity to changes in the composition of gases, air humidity, the presence of dust, temperature fluctuations and other external factors. A feature of thermocatalytic sensors is the ability to regulate their operation by changing the power supply parameters, which, together with modern microprocessor devices, allows for computer diagnostics of the performance of control devices and protective shutdown systems. One of the tasks of diagnostics is to identify cases of significant contamination of gas diffusion filters of sensors and intentional restriction of access to them of the controlled environment. Such diagnostics is possible by analyzing the pulsations of the sensor output signal caused by the turbulent mode of movement of the medium – by changing the frequency corresponding to the maximum of the spectral characteristic of the sensor output signal [15], or by evaluating the nature of the transient process after a short-term decrease in the sensor supply voltage [16]. In the first case, the calculation of the frequency requires a preliminary determination of the characteristics of the sensor in the city of installation and subsequent adjustment of the control system, and during the analysis of the transient process, the explosiveness control process is temporarily disrupted.

Semiconductor (metal oxide) sensors are most often used in air purification systems in homes, offices, in ventilation and air conditioning systems [17], as well as in portable electronic devices. Their advantage is high sensitivity, long service life and low cost. However, these sensors are not stable enough, which limits the possibility of their use in explosive control systems and monitoring of working conditions. The instability of the sensors is mainly caused by an incomplete recovery process after exposure to high concentrations of vapors or gases, which leads to a gradual drift of its sensitivity. To reduce the impact, the processing of signals received from the matrix of gas sensors responding to different concentrations of vapors by the method of partial least squares is proposed

[18]. However, this significantly complicates control systems, increases their cost and reduces reliability.

The singular spectral analysis method is a methodology for time series analysis that combines elements of classical time series research, analysis of multivariate statistical processes, and data processing in the context of complex systems. All aspects of the SSA methodology, as well as its individual components, can be used for data analysis, depending on the goals of the analysis. Thus, in [19], singular spectral analysis is used for data reconstruction, which made it possible to more accurately and quickly determine the initial components and overall global trends of the initial data of motor learning. In [20], the advantages of using processed SSA data instead of raw data in time series modeling and analysis are considered. It is also proposed to detect data anomalies by analyzing the time derivative of the SSA signal, this is especially relevant when the time derivative is taken from the original data containing noise. The use of a dual methodology to assess both the level and the dynamics of the SSA signal changes helps to identify abnormal situations. In [21], the analysis of the SSA eigenvalue function is used to find the correlation between the components of the input data, which allows finding regularities in them.

3 MATERIALS AND METHODS

Increasing the efficiency of technological process control systems and the reliability of explosive control systems of coal mines and oil and fuel complex facilities is possible by processing and pre-filtering the data received from the sensors for monitoring the grinding processes and the condition of the facilities.

To analyze the output signals of sensors used in control and management systems, the method of singular spectral analysis is used, which allows to reveal hidden structures and regularities in time series by pre-filtering and data processing of acoustic, thermocatalytic and semiconductor sensors of explosive gases and vapors.

Solving the problem of data complexity in object control and management systems requires a comprehensive approach, including the use of modern technologies, algorithms and methods of data analysis, as well as training and development of competencies in the field of data analysis within the organization. by pre-filtering and data processing. To analyze the output signals of sensors used in control and management systems, the method of singular spectral analysis can be used, which allows to reveal hidden structures and regularities in time series.

Singular spectral analysis refers to non-parametric methods of time series analysis. The purpose of the method is to decompose the watch series into interpreted additive components.

Let $N > 2$. Consider a material time series $F = (f_0, \dots, f_{N-1})$ of length N . Assuming that the series F is nonzero, that is, there is at least on $f_i = f(i\Delta)$ that is different from zero.

The basic algorithm consists of two successive, complementary stages: decomposition and restoration. The first stage is decomposition, which in turn is divided into embedding and singular decomposition.

The nesting procedure converts the original time series into a sequence of multidimensional vectors. Let L be some integer (window length), $1 < L < N$. The nesting procedure forms $K = L - N + 1$ nesting vectors:

$$X_i = (f_{i-1}, \dots, f_{i+L-2})^T.$$

When $1 \leq i \leq K$ have dimension L . These vectors are called L -embedding vectors. We denote by X the matrix composed of embedding vectors:

$$X = [X_1 : \dots : X_K].$$

In other words, the trajectory matrix has the form:

$$X = (x_{ij})_{ij=1}^{L,K} = \begin{pmatrix} f_0 & f_1 & f_2 & \dots & f_{K-1} \\ f_1 & f_2 & f_3 & \dots & f_K \\ f_2 & f_3 & f_4 & \dots & f_{K+1} \\ \dots & \dots & \dots & \dots & \dots \\ f_{L-1} & f_L & f_{L+1} & \dots & f_{N-1} \end{pmatrix}.$$

Nesting is a standard procedure in time series analysis. After that, this method uses the singular expansion of the trajectory matrix of the series.

Let $S = XX^T$. Denote by $\lambda_1, \dots, \lambda_L$ the eigenvalues of the matrix S , taken in irreducible order ($\lambda_1 \geq \dots \geq \lambda_L$). Due to the symmetry of the matrix S , it has a real spectrum. We denote by U_1, \dots, U_L – the orthonormal system of eigenvectors of the matrix S corresponding to the corresponding eigenvalues

Let $n = \max \{i : \lambda_i > 0\}$. If mark:

$$V_i = \frac{X^T U_i}{\sqrt{\lambda_i}}, i = 1, \dots, d,$$

then the singular expansion of the matrix X can be written as:

$$X = X_1 + \dots + X_d,$$

where $X_i = \sqrt{\lambda_i} U_i V_i^T$. Each of the matrices has rank 1. Therefore, they can be called elementary matrices.

We will call the set $(\sqrt{\lambda_i} U_i V_i^T)$ the proper triple of the singular expansion.

The decomposition step is followed by reconstruction consisting of clustering and diagonal averaging.

The clustering process divides the entire set of indices $\{1, \dots, d\}$ into m disjoint subsets based on decomposition I_1, \dots, I_m based on decomposition. Let $I = \{i_1, \dots, i_p\}$. Then the resulting matrix X_i , corresponding to the group I , is defined as

$$X_i = X_{i_1} + \dots + X_{i_p}.$$

Such matrices are calculated for $I = I_1, \dots, I_m$, thus the expansion can be written in grouped form

$$X_i = X_{i_1} + \dots + X_{i_m}.$$

The procedure for selecting sets I_1, \dots, I_m is called the grouping of proper triples.

The last step of singular spectral analysis is diagonal averaging. On it, each matrix of the grouped decomposition is transformed into a new series of length N .

Let Y – be some matrix $L * K$ with elements y_{ij} where $1 \leq i \leq L$, $1 \leq j \leq K$. Suppose, $L^* = \min(L, K)$, $K^* = \max(L, K)$ and $N = L + K - 1$. Let $y_{ij}^* = y_{ij}$ if $L < K$, and $y_{ij}^* = y_{ji}$ otherwise. Diagonal averaging translates the matrix Y to a row g_0, \dots, g_{N-1} – according to the formula:

$$g_k = \begin{cases} \frac{1}{k+1} \sum_{m=1}^{k+1} y_{m, k-m+2}^*; 0 \leq k \leq L^* - 1 \\ \frac{1}{L^*} \sum_{m=1}^{L^*} y_{m, k-m+2}^*; L^* - 1 \leq k \leq K^* \\ \frac{1}{N-k} \sum_{m=k-K^*+2}^{N-k+1} y_{m, k-m+2}^*; K^* \leq k \leq N \end{cases}.$$

Applying diagonal averaging to the resulting matrices X_{I_k} , we get $\tilde{F}^{(k)} = (\tilde{f}_0^{(k)}, \dots, \tilde{f}_{N-1}^{(k)})$ and, therefore, the original series (f_0, \dots, f_{N-1}) the original series is decomposed into the sum of m series:

$$f_n = \sum_{k=1}^m \tilde{f}_n^{(k)}.$$

The application of singular spectral analysis can help solve the following problems in data processing:

1. Data dimensionality reduction: SSA allows you to decompose a complex time series into its basic components (singular numbers and singular vectors), which reduces the data dimensionality. This simplifies the

analysis and visualization of the data, and also helps to identify the main components of the time series.

2. Noise and anomaly filtering: SSA can extract the main signal components while ignoring noise and anomalies in the data. This is especially useful if the data contains random variations or artifacts.

3. Highlighting trends and cycles: SSA helps to highlight trends, cycles and seasonal variations in data, which can be important for understanding temporal patterns and forecasting.

4. Time Series Forecasting: SSA-based analysis can be used to forecast future values of a time series based on selected components.

5. Anomaly detection: SSA allows you to compare actual data with model predictions, which can help identify anomalous events or changes in data.

6. Signal and Variability Analysis: SSA can be used for signal analysis, including the processing of temporal data from various sources.

7. Decomposing complex data into components: SSA divides the raw data into a set of components, each representing a different aspect of the time series. This allows for a deeper understanding of the data structure and the identification of key patterns.

8. Exploring data structure: SSA can be used to identify underlying patterns and structure in data, which facilitates data analysis and interpretation.

The application of SSA can significantly simplify the analysis of time data and the processing of complex time series, allowing to explore their structure, highlight key components and reduce the influence of noise.

4 EXPERIMENTS

In order to evaluate the possibility of using singular spectral analysis of acoustic signals, acoustic signals were analyzed for different modes of internal loading of the jet mill. Data from acoustic monitoring of the slag grinding process were used. The following operating modes of the jet mill were analyzed: loading, working mode and unloading.

The standard SSA algorithm assumes a gradual change in the global trend of time series. However, the acoustic signals for the different modes of the jet mill tend to have an initial steep slope due to the fact that most of the learning occurs at the beginning of the adaptation process. In this scenario, we observed that the standard SSA algorithm often fails to accurately capture the initial steepness of the time series. To solve this problem, we applied the Overlap SSA algorithm [22], where the standard SSA algorithm is separately applied to successive overlapping segments of the same length. Considering the computational costs and the requirement for a sufficient time series length in the spectral analysis used in the clustering step, we settled on using only three overlapping segments for the adaptive time series.

Based on the condition of separation of components, we will choose the optimal length of the window $L = 4000$. Based on the ordered series of eigenvalues of the covariance matrices, we estimate the number of main components necessary to restore the range of the useful signal. In this case, we leave the first ten significant components each, which explain 91.35% of the variance of the output series for the state of jet mill loading (Fig. 1), 93.96% – for the operating mode (Fig. 2), 98.4% – for unloading (Fig. 3).

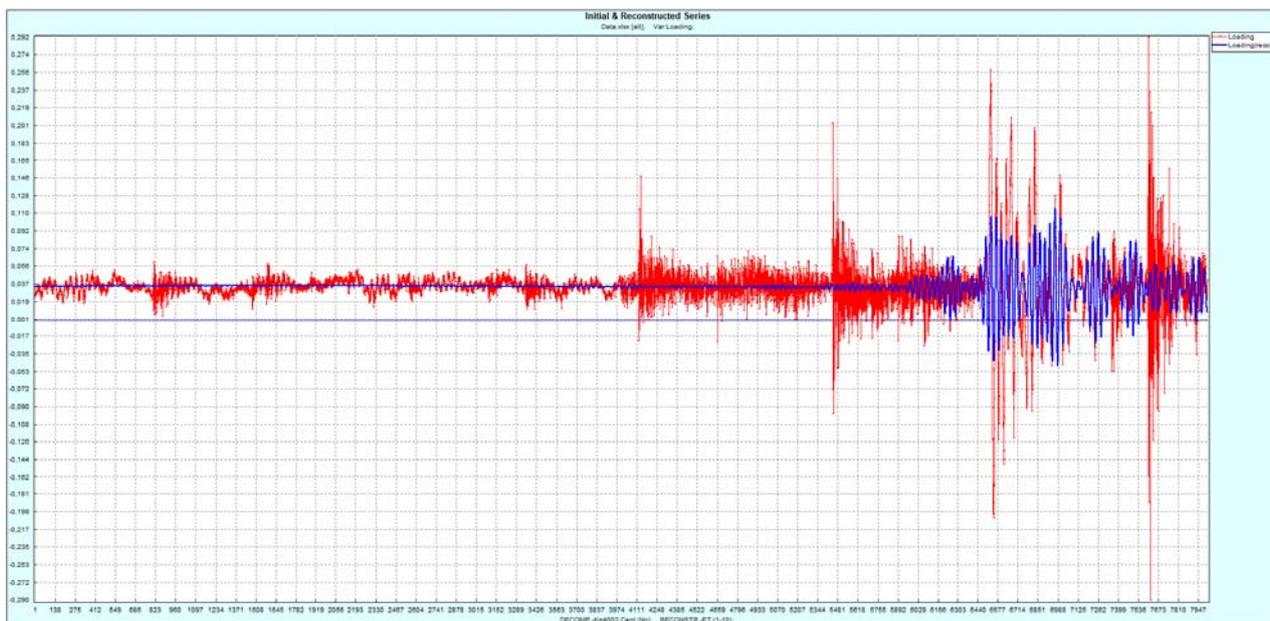


Figure 1 – Mill loading acoustic noise reconstructed by the first ten significant components

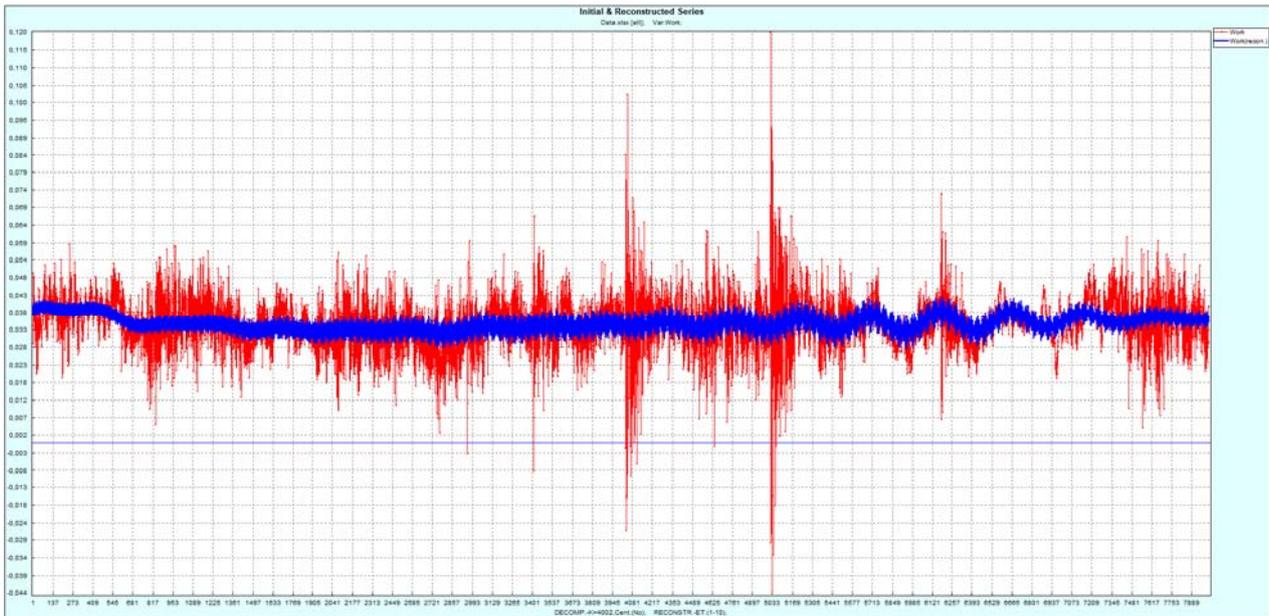


Figure 2 – Acoustic noise in the operating mode reconstructed by the first ten significant components

Based on the fact that more than 90% of the variances of the output series are contained in their first ten significant components at different operating modes of jet mills, it is advisable to use the definition of the main

significant components that carry the main informativeness of acoustic signals for the classification of acoustic signals arising during the operation of jet mills, and control their loading.

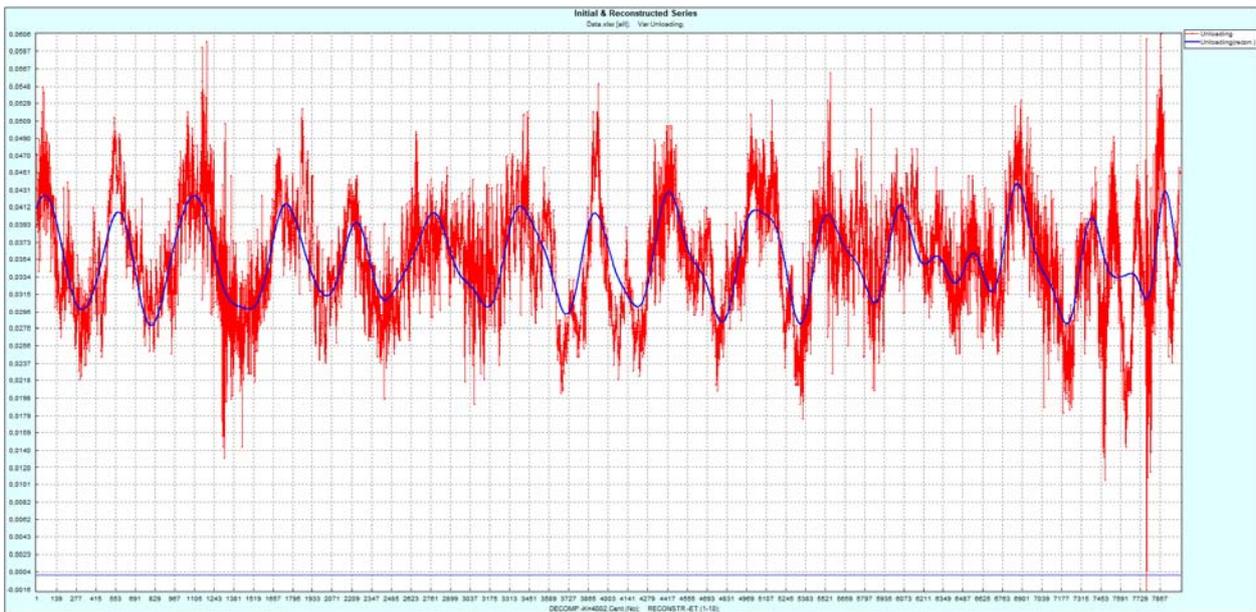


Figure 3 – Acoustic noise during mill unloading reconstructed by the first ten significant components

In order to assess the possibility of using singular spectral analysis in explosiveness control systems for filtering noise and anomalies, as well as to detect cases of significant contamination of gas diffusion filters of sensors and limiting access to them of the controlled environment, the monitoring data of the output signals of the sensors in different modes of their operation were analyzed.

The actual raw output signal of the sensors (Fig. 4) always contains noise and anomalies due to the influence of industrial sources of electromagnetic disturbances. In order to simplify the processing of the information coming from the sensors and offload the computing resources of the explosion monitoring systems, it is advisable to isolate the main signal components, ignoring noise and anomalies in the data. SSA can extract the main signal components while filtering out noise and

anomalies. In Fig. 5 shows the results of SSA processing of the sensor output signal in the absence of directional movement of the controlled medium. The sensor signal reconstructed by the first two significant components reflects the dynamics of changes in the concentration of the explosive component in the air. At the same time, the signal is filtered from uninformative noises and anomalies.

The output signal of the sensor is actual and restored according to the first two significant components in the

presence of directional movement of the controlled environment (Fig. 6) also reflects the dynamics of changes in the concentration of the explosive component in the air and is filtered from noise and anomalies. In its turn, reconstructed according to the first seven significant components (Fig. 7), the same sensor signal contains a harmonic component due to the presence of macroturbulent disturbances in the controlled environment.

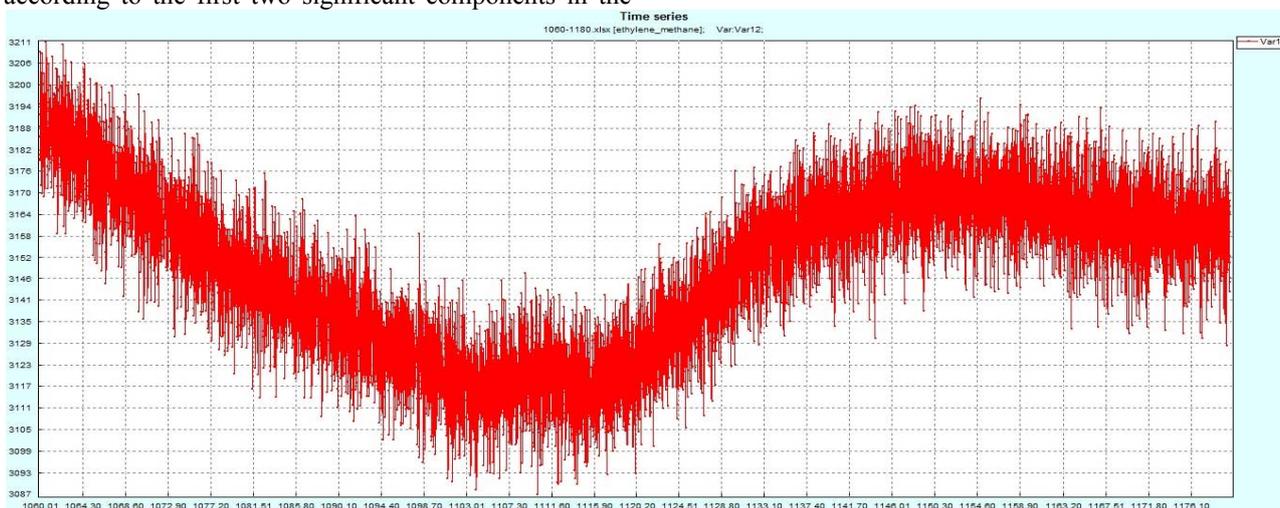


Figure 4 – Actual raw sensor output

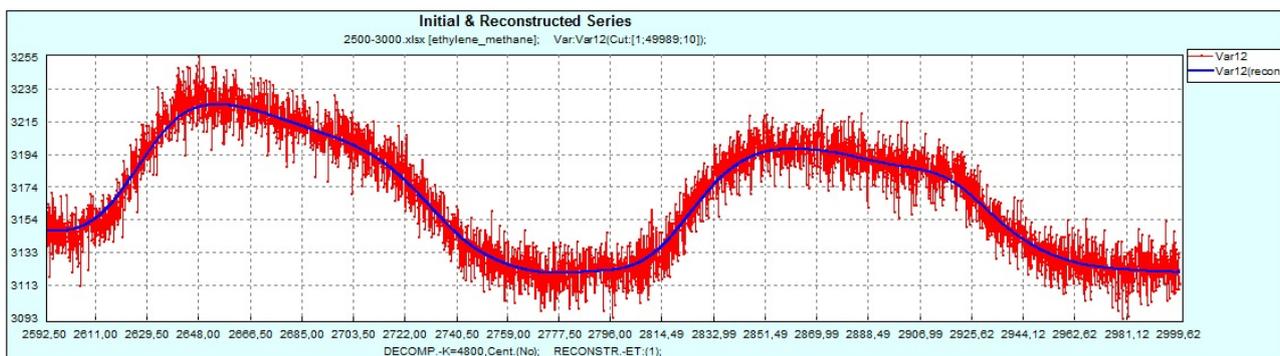


Figure 5 – Sensor output signal actual and restored by the first two significant components in the absence of directional motion of the monitored medium

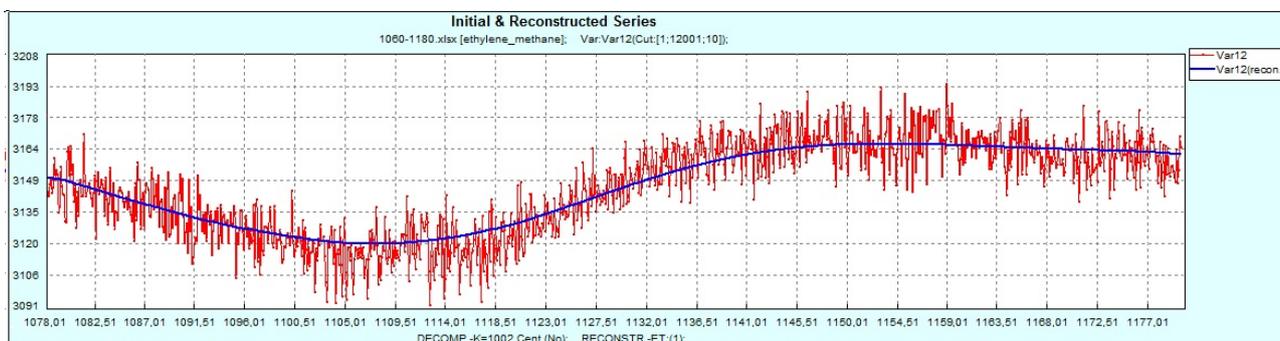


Figure 6 – Sensor output signal actual and restored by the first two significant components in the presence of directional motion of the controlled medium

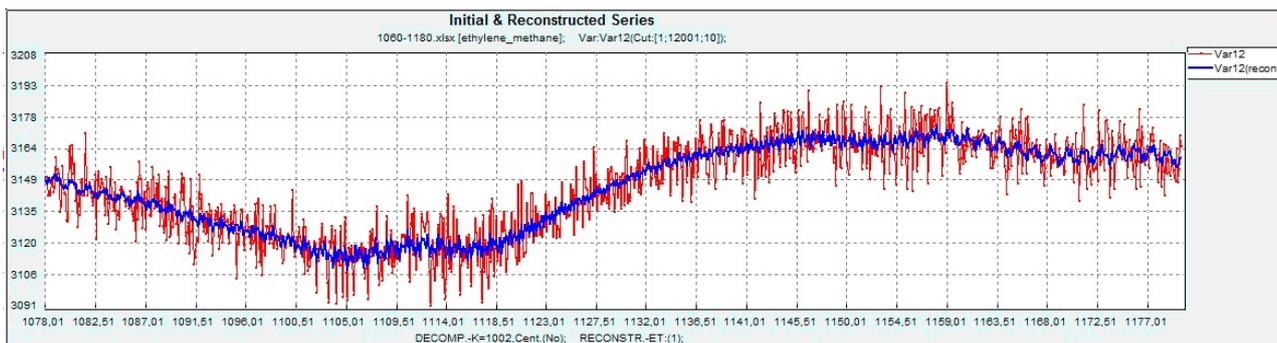


Figure 7 – Sensor output actual and restored by the first seven significant components in the presence of controlled medium motion

5 RESULTS

Thus, by decomposing difficult-to-analyze signals into components, SSA divides the raw data into a set of components, each representing a different aspect of the time series. In this case, the appearance of a harmonic component when restoring the signal by the first seven significant components indicates the presence of a directional movement of the medium. The frequency of macroturbulent disturbances at constant geometric dimensions of the flow depends on the speed of the medium, and the relative amplitude of the harmonic component depends on the diffusion conductivity of the sensor filters [14]. In the future, the use of SSA in explosion control systems allows to significantly expand their functions, including detecting significant pollution of gas diffusion filters of sensors, cases of restricted access to them of the controlled environment and changes in location, detecting fuel-air emissions at the facilities of the oil and fuel complex and evaluating them volumes All this determines the need for further research in this direction.

The proposed approach to managing technological processes of grinding raw materials in jet mills and controlling the explosiveness of coal mines and objects of the oil and fuel complex makes it possible to speed up the processing speed of sensor output data and improve the quality of information. One of the promising methods that can be used for preliminary processing of time series of output data from sensors in control and control systems is the method of singular spectral analysis.

6 DISCUSSION

The use of singular spectral analysis in technological process control systems and control of the explosiveness of objects allows to improve such criteria as:

1. Management of technological processes. Singular spectral analysis makes it possible to distinguish the main components in complex systems, which contributes to more accurate and efficient management of production processes. This is especially important in the conditions of modern high-tech industries, where the complexity and scale of technological processes require new approaches to management.

2. Safety of explosive objects. In the context of industrial safety, singular spectral analysis has been applied to the analysis and control of parameters related

© Holinko O. V., Alekseev M. O., Holinko V. I., Zabelina V. A., 2025
DOI 10.15588/1607-3274-2025-1-19

to explosiveness. Accurate and timely data on changes in technological processes allow you to warn about potential dangers or other factors that can cause accidents.

3. Automation and monitoring. The use of singular spectral analysis in control systems allows for the automation of monitoring and control processes, which reduces the probability of human errors and ensures a higher sensitivity of the system to changes in the production environment.

4. Reduction of risks and economic benefits. Early detection of potential problems in technological processes and quick response to them reduces the risk of accidents, increases overall safety and economic efficiency of production.

CONCLUSIONS

The performed studies showed the relevance of the use of singular spectral analysis for the analysis of time series obtained in technological process control systems. The application of singular spectral analysis is a powerful tool for processing and analyzing time data. Decomposing complex time series into basic components, such as singular numbers and vectors, using SSA provides effective data dimensionality reduction. This greatly simplifies the task of data analysis and visualization, allowing to highlight the main structural elements of the time series. SSA also excels in its ability to filter out noise and anomalies, which is a valuable aspect when dealing with real-world data prone to random variations. In addition, the application of SSA expands the analytical possibilities for highlighting trends, cycles and seasonal fluctuations in the data. This method is useful for forecasting future values of time series and contributes to a deeper understanding of temporal patterns. SSA opens up new possibilities in the analysis of temporal data by providing tools for extracting key components, exploring data structure, and reducing the influence of noise, which together make it a valuable technique in process control systems.

ACKNOWLEDGEMENTS

The authors express their sincere thanks to N. C. Pryadko. for the provided numerical data of signals of acoustic sensors characterizing different modes of operation of jet mills.



REFERENCES

1. Aldhafeeri T., Tran M. K., Vrolyk R. et al. A Review of Methane Gas Detection Sensors: Recent Developments and Future Perspectives, *Inventions*, 2020, No. 5(3), P. 28. doi.org/10.3390/inventions5030028
2. Nagahage I. S. P., Nagahage E. A. A. D., Fujino T. Assessment of the applicability of a low-cost sensor-based methane monitoring system for continuous multi-channel sampling, *Environmental Monitoring and Assessment*, 2021, No. 193, P. 509. doi.org/10.1007/s10661-021-09290-w
3. Okafor N. U., Alghorani Y., Delaney D. T. Improving Data Quality of Low-cost IoT Sensors in Environmental Monitoring Networks Using Data Fusion and Machine Learning Approach, *ICT Express*, 2020, No. 6(3), pp. 220–228. doi.org/10.1016/j.ict.2020.06.004
4. Manuel J., Lizaranzu M., Cuesta F. Improving Electronic Sensor Reliability by Robust Outlier Screening, *Sensors*, 2013, No. 13(10), pp. 13521–13542. doi.org/10.3390/s131013521
5. Martini A., Sorce A. Gross Error Detection Based on Serial Elimination Applications to an Industrial Gas Turbine, *ASME Turbo Expo*, 2014. http://dx.doi.org/10.1115/GT2014-26746
6. Hema P., Sathish E., Maheswari M. et al. Robust soft sensor systems for industry: Evaluated through real-time case study, *Measurement: Sensors*, 2022, P. 24. doi.org/10.1016/j.measen.2022.100542
7. Castaño F., Strzelczak S., Villalonga A. et al. Sensor Reliability in Cyber-Physical Systems Using Internet-of-Things Data A Review and Case Study, *Remote Sens*, 2019, No. 11(19), P. 2252. doi.org/10.3390/rs11192252
8. Teh H. Y., Liehr A. W. K., Wang K. K. Sensor data quality: a systematic review, *Journal of Big Data*, 2020, No. 7(11). doi.org/10.1186/s40537-020-0285-1
9. Pryadko N., Ternova K. Acoustic monitoring of jet grinding: monograph. Kyiv, VD “Akademperiodika”, 2020, P. 192. doi.org/10.15407/akademperiodyka.409.192 (In Ukrainian).
10. Bublikov A. V., Pilov P. I., Pryadko N. S. Automatic control of jet grinding on the basis of acoustic monitoring of mill operating zones, *Naukovyi Visnyk Natsionalnoho Hirnychoho Universytetu*, 2021, No. (4), pp. 29–34. www.scopus.com/record/display.uri?eid=2-s2.0-85113670976&origin=resultlist
11. Krzywanski J., Urbaniak D., Otwinowski H. et al. Fluidized Bed Jet Milling Process Optimized for Mass and Particle Size with a Fuzzy Logic Approach, *Materials*, 2020, No. 13(15), P. 3303. doi.org/10.3390/ma13153303
12. Krzywanski J., Wesolowska M., Blaszcuk A. et al. Fuzzy logic and bed-to-wall heat transfer in a large-scale CFBC, *International Journal of Numerical Methods for Heat & Fluid Flow*, 2018, No. 28, pp. 254–266. doi.org/10.1108/HFF-09-2017-0357
13. Ross T. J. Fuzzy Logic with Engineering Applications. New Mexico, John Wiley & Sons Chichester, 2010, 628 p.
14. Deng B., Luan J., Cui S. Analysis of parameter estimation using the sampling-type algorithm of discrete fractional Fourier transform, *Defence Technology*, 2014, No. 10(4), pp. 321–327. doi.org/10.1016/j.dt.2014.06.011
15. Golinko V., Kotlyarov A. Control of explosive environment in mining and equipment of coal mines. Dnipropetrovsk, “Lyra” publishing house, 2010, 368 p. discovery.kpi.ua/Record/000351517 (In Ukrainian).
16. Golinko V. I., Golinko O. V. Methods and algorithms for monitoring the state of the gas diffusion filter of methane analyzer, *Naukovi pratsi DonNTU Seriya: «Informatyka, kibernetyka ta obchyslyval'na tekhnika»*, 2022, №2–35, pp. 22–29. typeset.io/pdf/methods-and-algorithms-for-monitoring-the-state-of-the-gas-2u81q6go.pdf (In Ukrainian).
17. Koval Y. New semiconductor MEMS gas sensors from FIGARO ENGINEERING, *CHIP NEWS Ukraine*, 2012, No. 3(113), pp. 52–55. storage.sea.com.ua/files/links/mems-datciki-gaza-figarocno320121504002845.pdf
18. Laref R., Ahmadou D., Losson E. et al. Orthogonal Signal Correction to Improve Stability Regression Model in Gas Sensor Systems, *Hindawi Journal of Sensors*, 2017, No. 9851406, 8 p. doi.org/10.1155/2017/9851406
19. Swart S. B., Otter A. R., Lamoth C. J. C. Singular Spectrum Analysis as a data-driven approach to the analysis of motor adaptation time series, *Biomedical Signal Processing and Control*, 2022, No. 71, P. 103068. doi.org/10.1016/j.bspc.2021.103068
20. Espinosa F., Bartolomé A. B., Hernández P. V. et al. Contribution of Singular Spectral Analysis to Forecasting and Anomalies Detection of Indoors Air Quality, *Sensors*, 2022, No. 22(8), P. 3054. doi.org/10.3390/s22083054
21. Kumar S., Natarajan V. Singular Spectral Analysis (Ssa) of Solid Earth Tide (Set) – Implications to Identify Earthquake Precursors and Earthquakes in the Himalayan Region ($M \geq 6$) During 1991–2021, *Research Square*, 2022. doi.org/10.21203/rs.3.rs-1939978/v1
22. Leles M. C. R., Sansão J. P. H., Mozelli L. A. et al. A new algorithm in singular spectrum analysis framework: The Overlap-SSA (ov-SSA), *SoftwareX*, 2018, No. 8, pp. 26–32. doi.org/10.1016/j.softx.2017.11.001

Received 10.12.2025.
Accepted 21.01.2025.

ЗАСТОСУВАННЯ СИНГУЛЯРНОГО СПЕКТРАЛЬНОГО АНАЛІЗУ В СИСТЕМАХ КЕРУВАННЯ ТЕХНОЛОГІЧНИМИ ПРОЦЕСАМИ ТА КОНТРОЛЮ ВИБУХОНЕБЕЗПЕЧНОСТІ ОБ'ЄКТІВ

Голінько О. В. – аспірант кафедри програмного забезпечення комп'ютерних систем, Національний технічний університет «Дніпровська Політехніка», Дніпро, Україна.

Алексєєв М. О. – д-р техн. наук, професор, завідувач кафедри програмного забезпечення комп'ютерних систем, Національний технічний університет «Дніпровська Політехніка», Дніпро, Україна.

Голінько В. І. – д-р техн. наук, професор, завідувач кафедри охорони праці та цивільної безпеки, Національний технічний університет «Дніпровська Політехніка», Дніпро, Україна.

Забєліна В. А. – аспірант кафедри охорони праці та цивільної безпеки, Національний технічний університет «Дніпровська Політехніка», Дніпро, Україна.

АНОТАЦІЯ

Актуальність. Питання підвищення продуктивності технологічних процесів видобутку, переробки та підготовки сировини, поліпшення якості продукції, зниження енерговитрат, а також створення безпечних умов праці при веденні технологічних процесів та попередження аварій завжди є досить актуальним і потребує впровадження сучасних систем контролю і управління. Для ефективної роботи таких систем важливою є попередня обробка та фільтрація даних, отриманих від датчиків контролю процесів подрібнення та стану вибухонебезпечності об'єктів. Одним із можливих шляхів підвищення інформативності даних є застосування сингулярного спектрального аналізу.

Мета. Підвищення ефективності систем керування технологічними процесами та надійності систем контролю вибухонебезпечності вугільних шахт та об'єктів нафто-паливного комплексу шляхом обробки та попередньої фільтрації даних, отриманих від датчиків контролю процесів подрібнення та стану об'єктів.

Метод. Для аналізу вихідних сигналів датчиків, які використовуються в системах контролю та управління використаний метод сингулярного спектрального аналізу, який дозволяє виявити сховані структури та закономірності в часових рядах шляхом попередньої фільтрації та обробки даних акустичних, термокаталітичних та напівпровідникових датчиків.

Результати. Запропоновано новий підхід до керування технологічними процесами подрібнення сировини в струминних млинах та контролю вибухонебезпечності вугільних шахт і об'єктів нафто-паливного комплексу на основі методів, що дозволяють пришвидшити швидкість обробки вихідних даних датчиків та підвищити якість інформації. Показано, що одним із перспективних методів який можливо використати для попередньої обробки часових рядів вихідних даних датчиків в системах керування та контролю є метод сингулярного спектрального аналізу, використання якого дозволяє здійснити фільтрацію даних, виявити сховані структури та закономірності, здійснювати прогнозування змін на основі аналізу попередньої інформації, виявити аномалії і нештатні ситуації, приймати більш обгрунтовані рішення та поліпшити процеси керування технологічними процесами.

Висновки. Проведені експерименти підтвердили працездатність запропонованого програмного забезпечення та дозволили рекомендувати його для використання в розвитку як теоретичних, так і практичних аспектів систем керування технологічними процесами за допомогою вдосконаленого методу сингулярного спектрального аналізу (SSA) для обробки часових рядів. Цей удосконалений підхід був успішно продемонстрований у реальних застосуваннях, включаючи процеси подрібнення на струминних млинах і моніторингу вибухонебезпечності на вугільних шахтах та підприємствах нафто-паливного комплексу. Впровадження демонструє значне підвищення швидкості обробки даних і якості інформації, що робить його особливо цінним для застосування на критично важливих для безпеки промислових об'єктах.

КЛЮЧОВІ СЛОВА: системи керування, контроль вибухонебезпечності, датчики, сингулярний спектральний аналіз, часові ряди.

ЛІТЕРАТУРА

1. A Review of Methane Gas Detection Sensors: Recent Developments and Future Perspectives / [T. Aldhafeeri, M.K. Tran, R. Vrolyk et al.] // *Inventions*. – 2020. – No. 5(3). – P. 28. doi.org/10.3390/inventions5030028
2. Nagahage I. S. P. Assessment of the applicability of a low-cost sensor-based methane monitoring system for continuous multi-channel sampling / I. S. P. Nagahage, E. A. A. D. Nagahage, T. Fujino // *Environmental Monitoring and Assessment*. – 2021. – No. 193. – P. 509. doi.org/10.1007/s10661-021-09290-w
3. Okafor N. U. Improving Data Quality of Low-cost IoT Sensors in Environmental Monitoring Networks Using Data Fusion and Machine Learning Approach / N. U. Okafor, Y. Alghorani, D. T. Delaney // *ICT Express*. – 2020. – No. 6(3). – P. 220–228. doi.org/10.1016/j.ict.2020.06.004
4. Manuel J. Improving Electronic Sensor Reliability by Robust Outlier Screening / J. Manuel, M. Lizaranzu, F. Cuesta // *Sensors*. – 2013. – No. 13(10). – P. 13521–13542. doi.org/10.3390/s131013521
5. Martini A. Gross Error Detection Based on Serial Elimination Applications to an Industrial Gas Turbine / A. Martini, A. Sorce // *ASME Turbo Expo*. – 2014. http://dx.doi.org/10.1115/GT2014-26746
6. Robust soft sensor systems for industry: Evaluated through real-time case study / [P. Hema, E. Sathish, M. Maheswari et al.] // *Measurement: Sensors*. – 2022. – No. 24. doi.org/10.1016/j.measen.2022.100542

7. Sensor Reliability in Cyber-Physical Systems Using Internet-of-Things Data A Review and Case Study / [F. Castaño, S. Strzelczak, A. Villalonga et al.] // *Remote Sens.* – 2019. – No. 11(19). – P. 2252. doi.org/10.3390/rs11192252
8. Teh H. Y. Sensor data quality: a systematic review / H. Y. Teh, A. W. K. Liehr, K. K. Wang // *Journal of Big Data.* – 2020. – No. 7(11). doi.org/10.1186/s40537-020-0285-1
9. Pryadko N. Acoustic monitoring of jet grinding: monograph / N. Pryadko, K. Ternova. – Kyiv, VD “Akadempriodika”, 2020. – P. 192. doi.org/10.15407/akadempriodyka.409.192 (In Ukrainian).
10. Bublikov A. V. Automatic control of jet grinding on the basis of acoustic monitoring of mill operating zones. / A. V. Bublikov, P. I. Pilov, N. S. Pryadko // *Naukovyi Visnyk Natsionalnoho Hirnychoho Universytetu* – 2021. – No. (4). – P. 29–34. www.scopus.com/record/display.uri?eid=2-s2.0-85113670976&origin=resultlist
11. Fluidized Bed Jet Milling Process Optimized for Mass and Particle Size with a Fuzzy Logic Approach / [J. Krzywanski, D. Urbaniak, H. Otwinowski et al.] // *Materials.* – 2020. – No. 13(15). – P. 3303. doi.org/10.3390/ma13153303
12. Fuzzy logic and bed-to-wall heat transfer in a large-scale CFBC / [J. Krzywanski, M. Wesolowska, A. Blaszcuk et al.] // *International Journal of Numerical Methods for Heat & Fluid Flow.* – 2018. – No. 28. – P. 254–266. doi.org/10.1108/HFF-09-2017-0357
13. Ross T. J. *Fuzzy Logic with Engineering Applications* / T. J. Ross. – New Mexico : John Wiley & Sons Chichester, 2010. – 628 p.
14. Deng B. Analysis of parameter estimation using the sampling-type algorithm of discrete fractional Fourier transform / B. Deng, J. Luan, S. Cui // *Defence Technology.* – 2014. – No. 10(4). – P. 321–327. doi.org/10.1016/j.dt.2014.06.011
15. Golinko V. Control of explosive environment in mining and equipment of coal mines / V. Golinko, A. Kotlyarov. – Dnipropetrovsk : “Lyra” publishing house, 2010. – 368 p. discovery.kpi.ua/Record/000351517 (In Ukrainian).
16. Golinko V. I. Methods and algorithms for monitoring the state of the gas diffusion filter of methane analyzer / V. I. Golinko, O. V. Golinko // *Naukovi pratsi DonNTU Seriya: «Informatyka, kibernetyka ta obchyslyuval’na tekhnika».* – 2022. – № 2–35. – P. 22–29. typeset.io/pdf/methods-and-algorithms-for-monitoring-the-state-of-the-gas-2u81q6go.pdf (In Ukrainian).
17. Koval Y. New semiconductor MEMS gas sensors from FIGARO ENGINEERING / Y. Koval // *CHIP NEWS Ukraine.* – 2012. – No. 3(113). – P. 52–55. storage.sea.com.ua/files/links/mems-datciki-gaza-figarocno320121504002845.pdf
18. Orthogonal Signal Correction to Improve Stability Regression Model in Gas Sensor Systems / [R. Laref, D. Ahmadou, E. Losson et al.] // *Hindawi Journal of Sensors.* – 2017. – No. 9851406. – 8 p. doi.org/10.1155/2017/9851406
19. Swart S. B. Singular Spectrum Analysis as a data-driven approach to the analysis of motor adaptation time series / S. B. Swart, A. R. Otter, C. J. C. Lamoth // *Biomedical Signal Processing and Control.* – 2022. – No. 71. – P. 103068. doi.org/10.1016/j.bspc.2021.103068
20. Contribution of Singular Spectral Analysis to Forecasting and Anomalies Detection of Indoors Air Quality / [F. Espinosa, A. B. Bartolomé, P. V. Hernández et al.] // *Sensors.* – 2022. – No. 22(8). – P. 3054. doi.org/10.3390/s22083054
21. Kumar S. Singular Spectral Analysis (Ssa) of Solid Earth Tide (Set) – Implications to Identify Earthquake Precursors and Earthquakes in the Himalayan Region ($M \geq 6$) During 1991–2021 / S. Kumar, V. Natarajan // *Research Square.* – 2022. doi.org/10.21203/rs.3.rs-1939978/v1
22. A new algorithm in singular spectrum analysis framework: The Overlap-SSA (ov-SSA) / [M. C. R. Leles, J. P. H. Sansão, L. A. Mozelli et al.] // *SoftwareX.* – 2018. – No. 8. – P. 26–32. doi.org/10.1016/j.softx.2017.11.001

Наукове видання

**Радіоелектроніка,
інформатика,
управління**

№ 1/2025

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 19.02.2025. Формат 60×84/8.

Папір офс. Різогр. друк. Ум. друк. арк. 25,58.

Тираж 300 прим. Зам. № 171.

69063, м. Запоріжжя, НУ «Запорізька політехніка», друкарня, вул. Жуковського, 64

Свідоцтво суб'єкта видавничої справи

ДК № 6952 від 22.10.2019.