p-ISSN 1607-3274 Радіоелектроніка, інформатика, управління. 2025. № 2
e-ISSN 2313-688X Radio Electronics, Computer Science, Control. 2025. № 2

UDC 004.932

COMBINED METRIC FOR EVALUATING THE QUALITY OF SYNTHESIZED BIOMEDICAL IMAGES

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

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

Dombrovskyi M. O. – Master of Computer Engineering, West Ukrainian National University, Ternopil, Ukraine.

Liashchynskyi P. B. – PhD, Lecturer of the Department of Computer Engineering, West Ukrainian National University, Ternopil, Ukraine.

Melnyk G. M. – PhD, Associate Professor of the Department of Computer Engineering, West Ukrainian National University, Ternopil, Ukraine.

ABSTRACT

Context. This study addresses the problem of developing a new metric for evaluating the quality of synthesized images. The relevance of this problem is explained by the need for assessing the quality of artificially generated images. Additionally, the study highlights the potential of biomedical image synthesis based on diffusion models. The research results can be applied for biomedical image generation and quantitative quality assessment of synthesized images.

Objective. The aim of this study is to develop a combined metric and an algorithm for biomedical image synthesis to assess the quality of synthesized images.

Method. A combined metric M_C for evaluating the quality of synthesized images is proposed. This metric is based on two existing metrics: M_{IS} and M_{FID}. Additionally, an algorithm for histopathological image synthesis using diffusion models has been developed.

Results. To study the M_{IS}, M_{FID}, and M_C metrics, histopathological images available on the Zenodo platform were used. This dataset contains three classes of histopathological images G1, G2, and G3, representing pathological conditions of breast tissue. Based on the developed image synthesis algorithm, three classes of artificial histopathological images were generated.

Using the M_{IS}, M_{FID}, and M_C metrics, quality assessments of the synthesized histopathological images were obtained. The developed metric will form the basis of a software module for image quality assessment using metrics. This software module will be integrated into CAD systems.

Conclusions. A combined metric for evaluating the quality of synthesized images has been developed, along with a proposed algorithm for biomedical image synthesis. The software implementation of the combined metric and image synthesis algorithm has been integrated into an image quality assessment module.

KEYWORDS: metric, IS metric, FID metric, histopathological images, deep neural networks, diffusion models, Stable Diffusion.

ABBREVIATIONS

GAN is a generative adversarial network;

CNN is a convolutional neural network;

VT is a Vision-Transformer;

Zenodo is a general-purpose open repository developed under the European OpenAIRE program and operated by CERN;

G1 is a class of images of non-proliferative mastopathy;

G2 is a class of images of proliferative mastopathy;

G3 is a class of images of fibrocystic mastopathy;

IS is a distance based on the Google Inception V3 image classification model;

FID is a Fréchet inception distance;

HD is a Hausdorff Distance;

HI is a histopathological image;

U-Net is a network with U-shaped structure;

UF is an uncertainty Fréchet metric;

CIVIs is a clustering internal validity Indices;

VQ-VAE is a vector Quantized-variational autoencoder;

FAED is a Fréchet AutoEncoder distance;

MMD is a maximum mean discrepancy;

MMD GAN is a Maximum Mean Discrepancy for generative adversarial network;

KID is a Kernel Inception Distance;

QMC Quasi-Monte Carlo method;

PyTorch is an open-source machine learning framework Python Torch;

TensorFlow is an open-source platform and framework for machine learning, which includes libraries and tools based on Python and Java;

Clean-FID is a metrics with pre-processing of images;

CLIP is a Contrastive Language-Image Pre-training

CLIP-MMD is a Maximum Mean Discrepancy metric based on CLIP model;

CMMD is a metric based on CLIP embeddings and the maximum mean discrepancy distance with the Gaussian kernel;

ACCURACY is a metric that determines the proportion of correct predictions made by a model out of the total number of predictions.

NotImageNet32 is a synthetic dataset;

CIS is a Conditional Inception Score;

CFID is a Conditional Fréchet Inception Distance;

Wasserstein-1 is an improved of Wasserstein metric;

DOI 10.15588/1607-3274-2025-2-15

SSIM is a structural similarity index measure;

PSNR is a peak signal-to-noise ratio is ratio between the maximum possible power of a signal and the power of corrupting noise;

AUC is an area under the curve;

MS-SSIM is a Multi-Scale SSIM;

PRECISION is an accuracy of a model in predicting the positive class.

RECALL is a metric, that reflects the model's ability to identify all actual positive cases.

F1-score is an average of precision and recall;

LPIPS is a Learned Perceptual Image Patch Similarity;

GMM-GAN is a Graph-based Manifold Matching GAN;

MAE is a Mean Absolute Error;

UMMC is a University Malaya Medical Centre;

WSI is a whole slide image;

H&E is a hematoxylin and eosin;

IHC is an immunohistochemistry;

ER-IHC is an Estrogen Receptor Immunohistochemistry;

ER is an estrogen receptor;

PGR is a progesterone receptor;

HER2 is a human epidermal growth factor receptor 2;

PR is a progesterone receptor;

KI67 is a marker of proliferation Ki-67;

TIFF is a file format for storing raster graphics images;

JPG is a joint photographic experts' group;

MIRAX is a Carl Zeiss MIRAX slide scanner system file format;

CPU is a central processing unit

GPU is a graphics processing unit;

CUDA is a Compute Unified Device Architecture;

CAD stands for Computer-Aided Diagnosis;

RAM is a Random Access Memory;

VRAM is a Virtual Random Access Memory;

GB is a gigabyte;

GOOGLE COLAB is a free cloud-based platform provided by GOOGLE that allows users to write and execute Python code directly in their browser.

NOMENCLATURE

I_r is a set of real histopathological images;

I_l is a training set of histopathological images;

I_{test} is a test set of histopathological images;

I_c is a set of images based on diffusion models,

M_{IS} is a metric based on inception score;

M_{FID} is a metric based on Fréchet inception distance;

M_C is a a combined metric;

t is a time step index; t is selected from the range $t \in [0,T]$;

Z_{0C} is a set of histopathological images into the latent space at timestep t=0;

α_t is a coefficient that determines the noise rate at timestep t;

T is a total number of timesteps;

ε_t is a value of random Gaussian noise at timestep t;

$\hat{\varepsilon}_t$ is an estimated noise value at timestep t;

$\bar{\alpha}_t$ is a coefficient that determines the noise level at the previous timestep t;

β_t is a coefficient controlling the rate of noise reduction at timestep t;

Z_{1C} set of histopathological images into a latent space at timestep $t = T$;

I_{CD} is a set of images after encoder transforms Z_{1C};

$M_1, \ldots, M_n$ are metrics;

$X_1, \ldots, X_n$ are sets;

$N(E, D)$ is a normal distribution;

E is a mean;

D is a variance;

$M(x, y)$ is a given metric that measures the distance or dissimilarity between points x and y in a certain space;

s is a scaling factor applied to the metric;

m is a truncation parameter that limits the metric values;

a is a constant;

v is a translation parameter;

p is a fixed element from the set X;

μ and β are scaling coefficients;

μ_r, μ_G are mean feature vectors for real and generated images;

Σ_r, Σ_G are covariance matrixs for real and generated images.

Tr is a trace of the matrix;

E_x is an expected value over the distribution of x;

$p(x \mid y)$ is a probability distribution of class y for image x;

$p(y)$ is a marginal probability distribution over all classes;

D_{KL} is a Kullback-Leibler divergence.

$M_{FID_{\min}}$ is a minimum value of the M_{FID} metric;

$M_{FID_{\max}}$ is a maximum value of the M_{FID} metric;

$M_{IS_{\min}}$ is a minimum value of the M_{IS} metric;

$M_{IS_{\max}}$ is a maximum value of the M_{IS} metric;

M_{IS}^{N} is a normalized value of the M_{IS};

M_{FID}^{N} is a normalized value of the M_{FID};

α is a weight of a metric, with $\alpha \in [0,1]$.

INTRODUCTION

Biomedical images are structural and functional representations of human and animal organs designed for disease diagnosis and the study of the anatomical and physiological state of the body. Medical images used for disease diagnosis are obtained from digital radiology, computed tomography, nuclear magnetic resonance, ultrasound, microscopy, etc. For the diagnosis of oncological diseases, microscopic color images are used, including cytological, histopathological, and immunohistochemical images. Oncological diseases are a global issue, signifi-

DOI 10.15588/1607-3274-2025-2-15

cantly affecting countries such as the European Union, the United States, and Ukraine [1].

This article examines HIs of breast cancer, which has the highest mortality rate among women [2].

Existing datasets of HIs are limited due to objective reasons. Currently, deep neural networks of the CNN (Convolutional Neural Network) type are used for diagnosis [3]. Training neural networks requires large datasets. A solution to this problem is the synthesis of artificial HIs. The synthesis of artificial images is based on real HIs.

In a short period, the following approaches have been used for medical image synthesis:

1. Autoencoders and fully convolutional networks [4].
2. U-Net and GANs architectures [5].
3. VT architectures.
4. Diffusion models [6].

Thus, at present, the use of diffusion models is a relevant direction in medical image synthesis.

The generated images must be quantitatively evaluated. For this purpose, metrics are used to assess the similarity between the generated image and the real one. Another important task is to evaluate the diversity of the generated images. The most widely used metrics for these tasks are Inception Score (IS) and Fréchet Inception Distance (FID) [8, 9].

The subject of the research is the process of HI generation and quality assessment based on metrics.

The object of the research is metrics for evaluating the quality of synthesized images.

The purpose of the research is to develop a combined metric for assessing the quality of synthesized HIs.

1 PROBLEM STATEMENT

Let a set of real HIs I_r be given. We divide this set into two subsets: I_l and I_{test}, and

$$I_r = I_l \cup I_{test} .$$

To expand the training set I_l, we generate a set of images I_c based on diffusion models, such that:

$$I_c \gg I_l .$$

For image quality assessment, two metrics are used: M_{IS}, M_{FID}.

Each of these metrics has its own advantages and disadvantages. Therefore, it is necessary to develop a combined metric M_C that incorporates both M_{IS} and M_{FID}.

Thus, the objectives of this study are:

1. Development of the combined metric M_C.
2. Development of an algorithm for synthesizing HIs based-on diffusion models.
3. Conducting computational experiments to evaluate the quality of generated images using M_{IS} and M_{FID} metrics and comparing the experimental results with the M_C metric.

2 REVIEW OF THE LITERATURE

The study [9] addresses the problem of evaluating CIVIs for synthetic and real datasets. The novel UF index enables the assessment of clustering quality while considering uncertainty. The authors improved the accuracy and adaptability of clustering metrics, particularly for datasets with non-convex structures or high noise levels.

The authors of [10] focus on evaluating the quality of generated images in GANs. They point out that existing metrics, which are trained on ImageNet for feature extraction, may be inadequate for other data domains. Additionally, these metrics can be sensitive to noise and image distortions and do not always reflect human perception of image quality. To address these issues, the researchers proposed the FAED method, which incorporates VQ-VAE to achieve better clustering and local feature representation compared to traditional evaluation methods.

In [11], the authors introduce the FID metric as an alternative to the IS, arguing its advantages in measuring the similarity between the distributions of real and generated images. The study demonstrates that FID correlates better with human perception and is more sensitive to changes in image quality than IS.

The study [12] examines the training and performance of GANs using the MMD metric, leading to the development of MMD-GAN. The authors compare IS, FID, and their newly proposed KID. They acknowledge that FID is widely used as the primary evaluation tool, as it better captures the distribution of real and generated images compared to IS. However, they highlight FID's bias in small sample sizes, which can lead to inaccurate model comparisons. As an alternative, they propose KID, which serves as an unbiased estimator and is less dependent on feature distribution shifts.

The researchers [13] demonstrate that the IS and FID are biased metrics. The values of these metrics, when computed on finite sample sizes, differ from their true values, which can only be obtained with an infinite sample size. To address this issue, the authors propose the use of the QMC method to improve the estimation of FID and IS on finite datasets.

The study [14] investigates the impact of low-level image processing on the evaluation of generative models, particularly the influence of image resizing and compression techniques on the calculation of FID and IS metrics. The authors show that standard FID implementations in PyTorch and TensorFlow introduce artifacts due to incorrect smoothing during downscaling, affecting the accuracy of the metric. As a solution, the authors propose the Clean-FID evaluation, which eliminates these artifacts by ensuring consistent pre-smoothing and avoiding quantization errors, leading to more reliable quality assessments of generated images.

The authors of [15] argue that the FID metric has significant limitations, including incorrect assumptions about the normality of feature distributions, low efficiency with small sample sizes, and poor alignment with human perception of image quality. As an alternative, they propose

DOI 10.15588/1607-3274-2025-2-15

the CMMD metric, which is based on CLIP embeddings and the MMD method.

The key issue addressed in [16] is the volatility of FID and IS metrics and their limited ability to accurately compare similar generative models. The authors introduce a new evaluation protocol that relies on the creation of a synthetic dataset NotImageNet32 and the use of KL divergence for more objective model comparisons.

The study [17] focuses on the evaluation of generative models in the case of conditional image generation. The authors propose two new metrics: CIS and CFID, which decompose the evaluation into within-class and between-class components, allowing for a more refined assessment of generative models.

The research conducted by [18] examines the training of generative models for unsupervised learning, specifically the challenges in measuring the closeness between real and generated data distributions. The authors introduce a new metric, Wasserstein-1, which addresses training instability and mode collapse issues in GANs, providing a more stable and theoretically grounded approach to generative model evaluation.

The study [19] analyzes various deep learning-based generative models, including CNNs, GANs, Transformers, and Diffusion models, highlighting their applications, architectural differences, and effectiveness in medical image synthesis. The performance evaluation focuses on medical image quality metrics, such as SSIM, PSNR, and segmentation accuracy in downstream tasks.

The article [20] investigates the problem of automatic metastasis detection in breast HIs, examining the impact of color variation on CNN performance. The study's primary objective is to minimize the effect of staining variability and improve classification accuracy through ensemble learning and color normalization techniques. The evaluation metrics used include Accuracy, AUC, Sensitivity, Specificity, and Kappa coefficient.

The research [21] aims to address the insufficient and imbalanced volume of medical data for breast tumor classification using ultrasound and mammography. The use of GANs allows for the generation of synthetic images, which helps reduce overfitting and enhance the performance of deep neural networks in diagnostic tasks. The evaluation of synthetic images is conducted using metrics such as KID, SSIM, and MS-SSIM.

The article [22] focuses on automated classification of histopathological medical images for early breast cancer diagnosis using deep learning. The primary evaluation metrics utilized by the authors include Accuracy, Sensitivity, Specificity, F1-score, and AUC.

The authors of [23] investigate the limited availability of large-scale annotated medical image datasets, which is a critical factor in training deep models for medical imaging. They propose a method for generating ultrasound images using semantic information to enhance realism. In addition to FID, the authors apply PSNR, MS-SSIM, and LPIPS to more accurately measure the similarity of texture features between synthetic and real images.

The study [24] addresses the challenge of cross-modality medical image generation, which is crucial for diagnostics when certain image types are unavailable. The authors introduce the GMM-GAN, which leverages graph neural networks to improve the quality of synthetic images and ensure their alignment with real medical data. The evaluation is conducted using MAE, PSNR, SSIM, and MS-SSIM.

The article [25] focuses on the automated synthesis of medical images in the absence of paired data, which is critical when obtaining multimodal images is restricted. Besides FID, the authors utilize Dice Score, HD, MAE, PSNR, and SSIM to assess the quality of the generated images.

The review study [26] analyzes the application of GANs in digital pathology and HI processing. The article discusses key challenges such as color normalization, resolution enhancement, and artifact removal. Additionally, the study highlights the need to improve consistency across images obtained from different laboratories. The evaluation is based on SSIM, PSNR, Dice Score, and HD, focusing on the quality assessment of synthetic images used by pathologists.

The study [27] applies CNNs and pre-trained models for the classification of oral cancer images. The authors evaluate the performance of their models using standard classification metrics, including Accuracy, Recall, Precision, F1-score, and AUC.

The study [28] explores various methods for evaluating image segmentation quality. The authors review traditional and modern approaches to segmentation assessment, emphasizing objective and subjective evaluation criteria. The paper highlights the importance of quantitative metrics such as Precision, Recall, and SSIM, which are commonly used to validate the performance of segmentation algorithms. The study contributes to the ongoing development of reliable evaluation techniques in image processing, particularly in medical imaging applications.

The article [29] investigates GANs for biomedical image synthesis, proposing a method for automatically searching for optimal GAN architectures. The authors discuss the challenges associated with GAN architecture selection and present an approach that optimizes models for high-quality image generation. The study also evaluates the performance of different architectures using FID, SSIM, and IS to ensure the realism and diversity of the generated images.

The research [30] introduces a deep learning-based method and software for biomedical image generation and classification in small sample settings. The authors address the data scarcity problem in medical imaging by developing a model that enhances image synthesis and classification accuracy. The paper focuses on GAN-based data augmentation, improving classification performance using CNNs and Transformer models. The evaluation is conducted using Accuracy, F1-score, AUC and SSIM to measure the effectiveness of both synthetic image generation and classification accuracy. The results demonstrate

DOI 10.15588/1607-3274-2025-2-15

significant improvements in small-sample learning scenarios, making the approach highly relevant for biomedical applications.

3 MATERIALS AND METHODS

The development of the combined metric and the evaluation of generated HIs consist of the following stages:

1. Preparation of the dataset of HIs of breast pathology.
2. Development of an algorithm for generating HIs based-on diffusion models.
3. Design of a combined metric based on FID and IS metrics.
4. Conducting computational experiments using HIs.

This section describes steps 1, 2, and 3.

For machine learning purposes, researchers utilize public biomedical image datasets.

The ACROBAT dataset [31] consists of 4212 WSI from 1153 patients diagnosed with primary breast cancer. The images were obtained from tissue samples stained with H&E, as well IHC staining for ER, PGR, HER2, and KI67 markers. Each patient has one H&E image and between one to four IHC images. The images were digitized at 10x magnification using Hamamatsu NanoZoomer S360 scanners.

The dataset was collected as part of the "Cancer Histopathology Image Epidemiology Project" (Sweden) and is used for the development of computational pathology methods. It is divided into: training set: 3,406 images, validation set: 200 images, test set: 606 images. The dataset is available in TIFF format and includes metadata with detailed information about each image.

The UMMC ER-IHC Breast Histopathology Whole Slide Image and Allred Score dataset [32] contains 37 whole-slide histological images, obtained in collaboration with the UMMC. The images were stained using IHC for ER and digitized using a 3DHistech Pannoramic Desk scanner at 20x magnification, with a resolution of 80000 × 200000 pixels.

Each image was annotated by pathologists, and the dataset is available in MIRAX format, accompanied by thumbnail images with annotation overlays.

For machine learning in image classification, the authors have created and published a dataset of cytological and histological images [33]. This dataset consists of: 73 cytological images, 68 histological images.

The images were obtained from tissue samples stained with H&E. Additionally, IHC images for ER, HER2, KI67, and PR markers were acquired. Each patient has one H&E image and between one to two IHC images.

The images were digitized at 10x and 20x magnification using a Nikon Eclipse Si laboratory-grade microscope. The dataset, along with diagnostic information, was collected at the Interdepartmental Educational and Research Laboratory of Ternopil National Medical University named after I.Y. Horbachevsky (Ukraine, Ternopil). The dataset is available in JPG format, with metadata including: patient age, cancer stage. The cytological images are categorized into three classes: melanoma, non-proliferative fibrous mastopathy, cystic mastopathy. The histological images belong to the class of invasive ductal carcinoma.

Mastopathy is one of the most common pathologies of the mammary glands in women of reproductive age. While this condition is generally benign, in certain cases, it is associated with an increased risk of malignant tumor development. This characteristic makes mastopathy a critical subject for early diagnosis and regular monitoring.

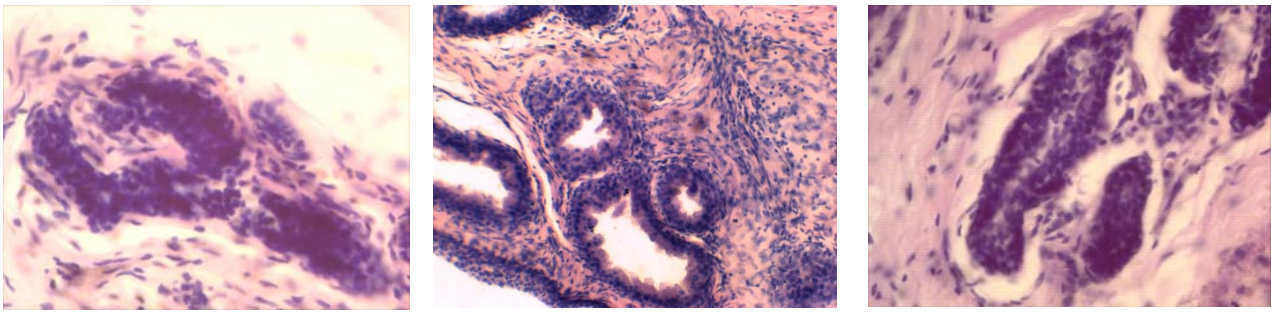
The dataset contains histological images with a resolution of 3664 × 2748 pixels, stained with H&E. The total number of images is 369. The images in the dataset (Fig. 1) belong to three classes: non-proliferative mastopathy, proliferative mastopathy, fibrocystic mastopathy.

The number of images in each class:

– non-proliferative mastopathy: 37;
– proliferative mastopathy: 45;
– fibrocystic mastopathy: 23.

The preprocessing of HIs is aimed at enhancing image quality, reducing variations, and removing artifacts, thereby improving the accuracy and reliability of analysis results.

Initially, color normalization was performed to minimize variations caused by different staining and scanning conditions. This was achieved through color standardization and intensity normalization methods. Next, noise filtering was applied, specifically using a median filter to reduce unwanted artifacts.



a b c

Figure 1 – Image Types in the Dataset: a – non-proliferative mastopathy; b – proliferative mastopathy; c – fibrocystic mastopathy

DOI 10.15588/1607-3274-2025-2-15

Since the initial dataset is relatively small, the next step involves expanding the sample for all three classes using the Stable Diffusion model.

For the synthesis of HIs, diffusion models were utilized within the Stable Diffusion environment [34]. The Stable Diffusion framework consists of two deep neural networks: Primary Network and Hypernetwork.

The Primary Network is trained on a large dataset of images and stores the parameters (weights) for the general dataset. The training process using HI datasets takes place within the Hypernetwork environment, which adjusts the weights of the base model.

The image generation algorithm in both networks is based on diffusion models and consists of the following steps:

1) training on the HI dataset within the Hypernetwork environment;

2) noise addition process (diffusion) applied to the initial dataset I_c;

3) denoising process (reverse diffusion) to generate high-quality synthetic images.

Let's detail the steps. We perform the transformation of the set of HIs into the latent space:

$$I_c \rightarrow Z_{0C}.$$

Based on Z_{0C}, the noise value is computed at each timestep t as follows:

$$Z_t = \sqrt{\alpha_t} Z_{0C} + \sqrt{1-\alpha_t}\,\varepsilon_t,$$

where: $\varepsilon_t \sim N(0, 1)$, $N(0, 1)$ is the normal distribution with mean $E = 0$ and variance $D = 1$.

The denoising value is computed as follows:

$$Z_{t-1} = \frac{1}{\sqrt{\alpha_t}}\left(Z_t - \frac{\beta_t}{\sqrt{1-\bar{\alpha}_t}}\hat{\varepsilon}_t\right).$$

After completing the denoising process (after $t = T$ timesteps), the vector Z_{1C} is formed in the latent space. Then, the encoder transforms Z_{1C} into the set of images I_{CD}, where $I_{CD} \gg I_C$. The quality of the generated images is evaluated using M_{IS} and M_{FID} metrics.

There are various ways to derive a new metric from an existing one. Let M_1 be a given metric on a set X. Then, a new metric can be constructed based on M_1 through transformations.

Alternatively, given multiple metrics $M_1, \ldots, M_n$, defined on sets $X_1, \ldots, X_n$, we can obtain a new metric on the extended set $X_1, \ldots, X_n$.

Consider two metrics M_1 and M_2 on the set X.

A metric transformation is a new metric obtained as a function of the existing metrics M_1 and M_2 on X.

The general matrix transformations are as follows [35]:

1) Scaled metric:

$$s\,M(x, y),\ s > 0.$$

2) Truncated metric:

$$\min\{mM(x, y)\}.$$

3) Discrete metric transformation:

$$\max\{a, M(x, y)\},\ x \neq y.$$

4) Translation metric:

$$M(x, y) + v.$$

5) Normalized metric transformation:

$$\frac{M(x, y)}{1 + M(x, y)}.$$

6) Biotope metric:

$$M_p(x, y) = \frac{M(x, y)}{M(x, p) + M(y, p) + M(x, y)}.$$

7) Maximum metric transformation:

$$\max\{M_1(x, y), M_2(x, y)\}.$$

8) Metric cone transformation:

$$\mu M_1(x, y) + \beta M_2(x, y).$$

In the article, the 8th transformation is used.

To evaluate the quality of generated images, two common metrics are utilized: M_{FID} and M_{IS}.

The distance calculation between real and generated images is performed using the following formula:

$$M_{FID} = \|\mu_r - \mu_G\|^2 + Tr\left(\Sigma_r + \Sigma_G - 2\left(\Sigma_r\Sigma_G\right)^{\frac{1}{2}}\right).$$

The M_{IS} metric is defined by the following formula:

$$M_{IS} = \exp\left(E_x\left(D_{KL}\left(p(x\,|y) \,\|\, p(y)\right)\right)\right).$$

The Kullback-Leibler divergence D_{KL} is calculated as follows:

$$D_{KL}\left(p(x\,|y) \,\|\, p(y)\right) = \sum_y p(y\,|x) \log \frac{p(y\,|x)}{p(y)}.$$

The goal is to combine these two metrics.

The M_{FID} and M_{IS} metrics have different ranges and interpretations. A lower M_{FID} value indicates that the gen-

DOI 10.15588/1607-3274-2025-2-15

erated image is more similar to the real one. A higher M_{IS} value suggests that the generated images are better represented across different classes.

To combine these metrics, normalization must be performed.

The normalized value of the M_{FID} metric is computed as follows:

$$M_{FID}^{N} = \frac{M_{FID} - M_{FID_{\min}}}{M_{FID_{\max}} - M_{FID_{\min}}}.$$

For the M_{IS} metric, the normalization formula is:

$$M_{IS}^{N} = \frac{M_{IS} - M_{IS_{\min}}}{M_{IS_{\max}} - M_{IS_{\min}}}.$$

Thus, the combined metric is defined as:

$$M_C = (1-\alpha)M_{FID}^{N} + \alpha M_{IS}^{N}.$$

4 EXPERIMENTS

In this paper, two types of experiments were conducted: 1. Synthesis of HIs of three classes based on real images in the Stable Diffusion environment. 2. Evaluation of the quality of the generated images using the IS, FID metrics and the combined metric.

For conducting computer experiments on HI synthesis, the Stable Diffusion Web UI software tool by Jarvis Labs was used. The computational resources utilized for this process were: GPU: 1 x A6000 Ampere (CUDA 12.3), CPU: 7 cores, RAM: 32 GB, VRAM: 48 GB.

This configuration ensures high performance for generating synthetic images.

Stable Diffusion version specifications: version: v1.9.4; python: 3.10.14; torch: 2.1.2+cu121; xformers: 0.0.23; post1; gradio: 3.41.2; checkpoint: 5493a0ec49; model: v1-5-pruned-emaonly.safetensors.

Before starting the experiment, small real datasets were expanded using affine transformation algorithms, implemented via the Rudi library [36] in Python.

The Rudi library is used for automated augmentation (generating variations) of cropped images. The following augmentation operations were performed: transformation probability, rotations, scaling, distortion intensity.

During the experiment, three extended datasets representing different image classes were used: histo_fibrocystic_cystic_mastopathy, non-proliferative, proliferative. After expansion, the number of images per class ranged from 119 to 165.

Training was conducted for 10000 steps per dataset, corresponding to 60 to 84 epochs, depending on the number of images in the dataset. Total training time per class: 43 minutes. Time per step: 0.26 seconds. Epoch duration: 31 to 42.9 seconds, varying due to dataset size differences.

The generation of synthetic images was performed using a specialized sampling approach, ensuring high detail and realism in the results. Synthesis of 4000 images per class took approximately 3 hours and 20 minutes.

Figure 2 presents original and generated images for the first class (histo_fibrocystic_cystic_mastopathy), demonstrating the quality of the obtained results.

Figure 3 presents examples of real and generated images for the class (proliferative).

Figure 4 presents examples of real and generated images for the class (non-proliferative).

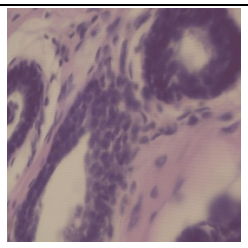
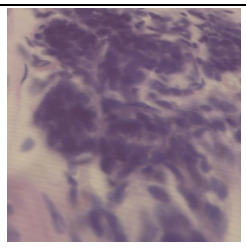
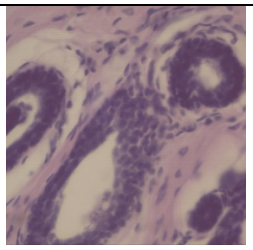
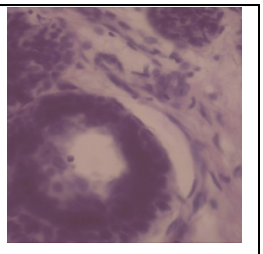
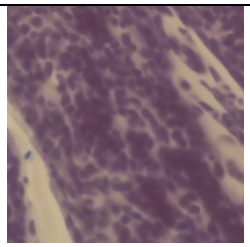
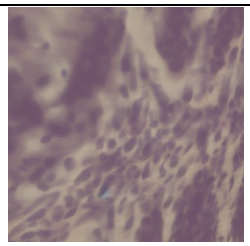
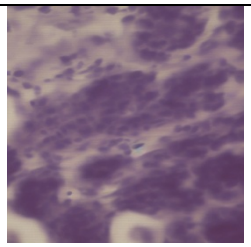
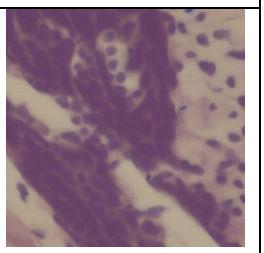
Original				
Steps	2500	5000	7500	10000
Generated				

Figure 2 – Examples of real and generated images for the class (histo_fibrocystic_cystic_mastopathy)

DOI 10.15588/1607-3274-2025-2-15

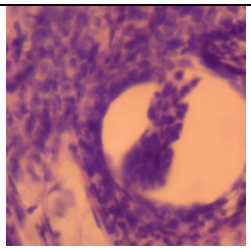
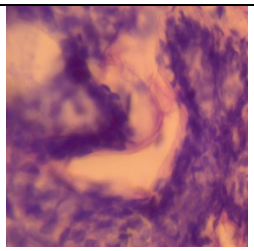
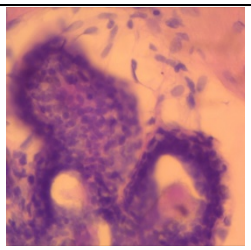
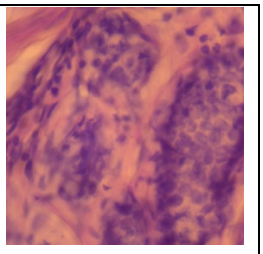
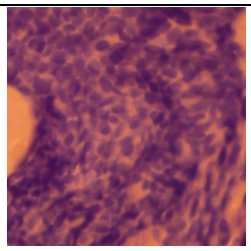
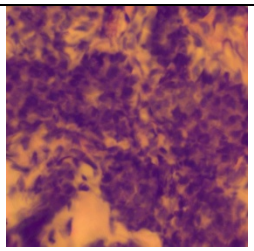
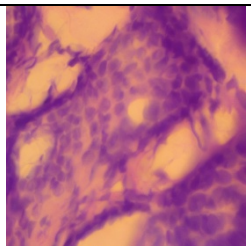
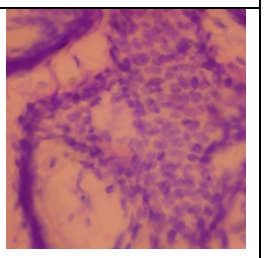
Original				
Steps	2500	5000	7500	10000
Generated				

Figure 3 – Example of real and generated images for the class (proliferative)

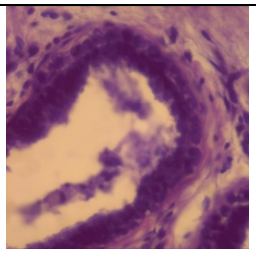
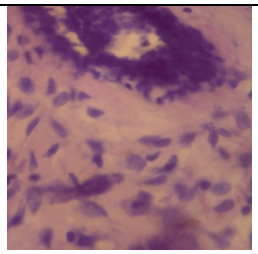
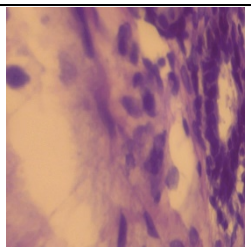
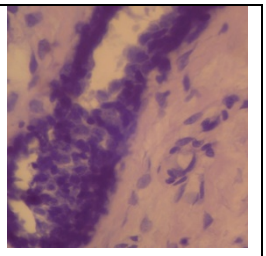
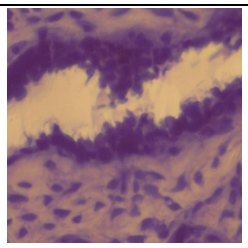
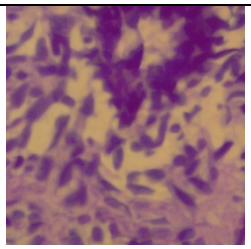
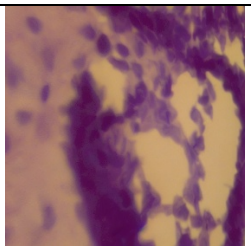
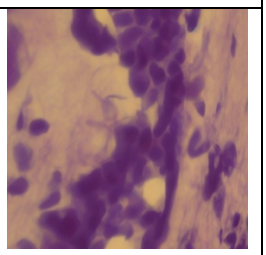
Original				
Steps	2500	5000	7500	10000
Generateed				

Figure 4 – Example of real and generated images for the class (non-proliferative)

After generation, the synthesized images were compared with the training images based on metrics calculated in Google Colab.

The data was stored in tables containing information on the number of iterations, metric values, and normalized indicators.

Based on the conducted experiments, the following were constructed:

1. Graphs of metric value changes during training (M_{FID}^{N}, M_{IS}^{N}, M_C as a function of iterations).

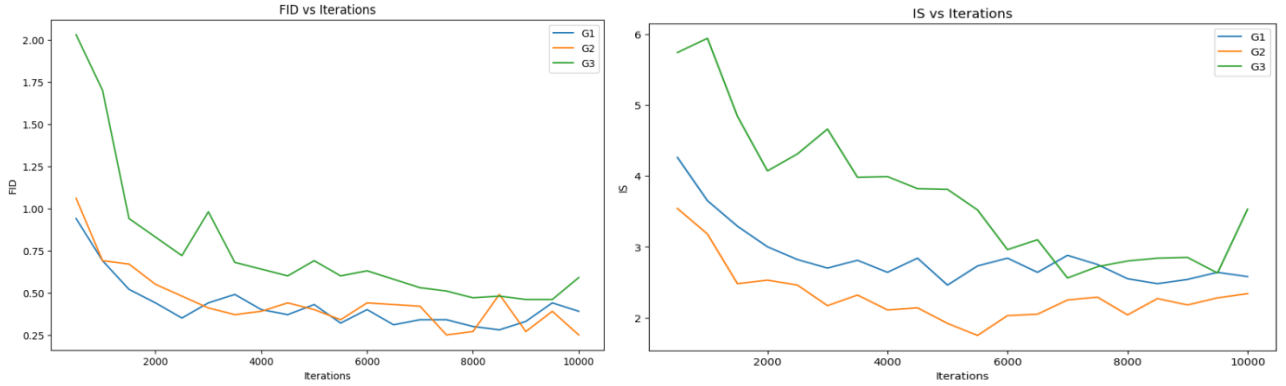
2. 3D graphs showing the relationship between M_C, M_{FID}^{N}, and M_{IS}^{N} for different values of the parameter α.

The graphs were plotted in the Google Colab environment. Figures 5a and 5b present the graphs of the IS and FID metric dependencies on the number of iterations for three classes: G1, G2 and G3.

Figures 6a and 6b present the dependencies of M_{FID}^{N} and M_{IS}^{N} on the number of iterations for the three classes: G1, G2 and G3.

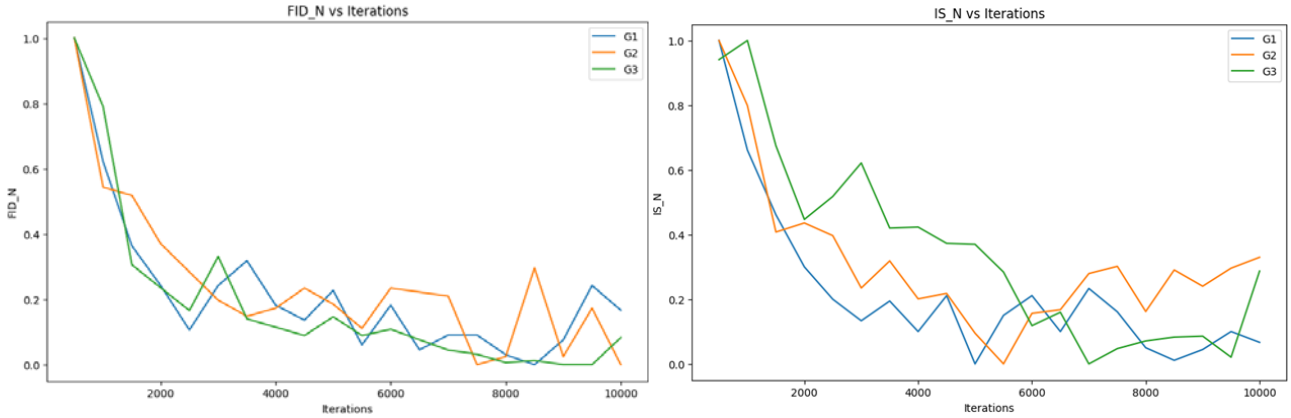
Figure 7 presents the graph of the dependence of M_C on the number of iterations for class G1. Similarly, graphs of the dependence of M_C on the number of iterations for classes G2 and G3 are provided.

DOI 10.15588/1607-3274-2025-2-15



a b

Figure 5 – a – Dependence of IS on the number of iterations; b – Dependence of FID on the number of iterations



a b

Figure 6 – a – Dependence of M_{FID}^{N} on the number of iterations; b – Dependence of M_{IS}^{N} on the number of iterations

Figure 7 – Dependence of M_C on the number of iterations

Figure 8 presents a 3D graph of the dependence of M_C on the number of iterations for class G3. Similar dependencies are shown for other classes.

DOI 10.15588/1607-3274-2025-2-15

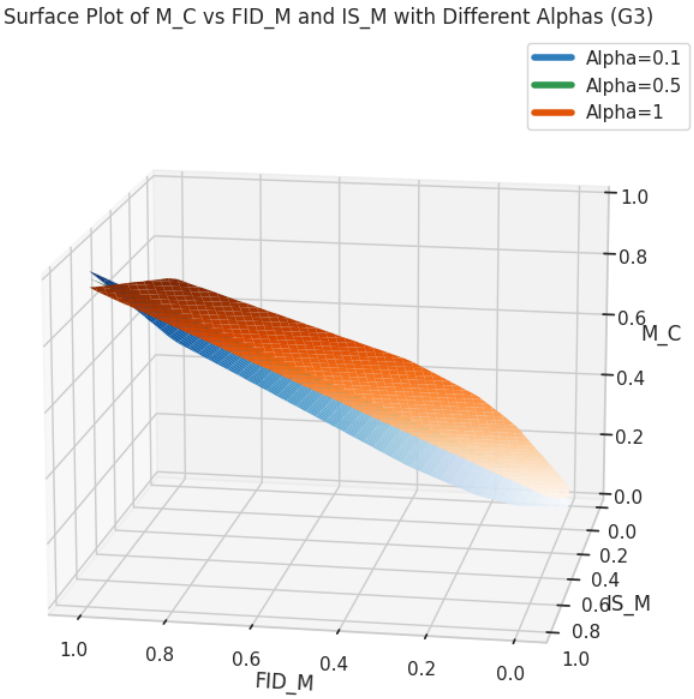


Figure 8 – Dependence of M_C on the number of iterations (class G3)

5 DISCUSSION

Modern classifiers are built on deep neural networks. The required classification accuracy is achieved through a large dataset of input images and prolonged neural network training. However, biomedical image datasets are limited due to objective constraints in obtaining biomedical images. Therefore, an urgent task is the generation of synthetic biomedical images based on real ones. The study [34] demonstrates that a promising approach for biomedical image generation is the use of diffusion models and modern generative tools such as Stable Diffusion. In this work, diffusion models are utilized for generating HIs.

The second crucial task is the evaluation of the generated images. Typically, artificial image evaluation relies on metrics and their modifications. The M_{FID} metric computes the similarity between real and generated images, while the M_{IS} metric measures the diversity of images in the generated dataset. Ideally, the M_{FID} metric should indicate minimal distance between real and generated images, whereas the M_{IS} metric should reflect maximum diversity in the synthesized dataset. This necessitates the combination of these metrics, leading to the development of the combined metric M_C, which allows for adjusting the influence of M_{IS} and M_{FID} on the quality assessment of generated images.

The initial biomedical image dataset consisted of three classes: G1 –37 images, G2 – 45 images, and G3 – 23 images. Using affine transformations, the dataset was expanded to 165 images in each class. Based on Stable Diffusion, 4000 images were generated for each class.

The total training time was 43 minutes per class. The synthesis of 4000 images took approximately 3 hours and 20 minutes for each class. The numerical parameters of the conducted experiments are as follows:

1. For class G1: $M_{FID_{\min}} = 0.28$, $M_{IS_{\min}} = 2.46$; $M_{FID_{\max}} = 0.94$, $M_{IS_{\max}} = 4.26$.
2. For class G2: $M_{FID_{\min}} = 0.25$, $M_{IS_{\min}} = 1.75$; $M_{FID_{\max}} = 1.06$, $M_{IS_{\max}} = 3.54$.
3. For class G3: $M_{FID_{\min}} = 0.46$, $M_{IS_{\min}} = 2.56$; $M_{FID_{\max}} = 2.03$, $M_{IS_{\max}} = 5.94$.

As shown in Figures 5a, 5b, 6a, 6b, and 7, approximately 2500 iterations are sufficient to achieve the required quality of generated images. This trend is observed for the M_{IS}, M_{FID}, and M_C metrics.

The developed combined metric will be integrated into the image quality assessment module based on metric evaluation. The metric module will be one of the components of the CAD software system.

CONCLUSIONS

This paper addresses the problem of evaluating the quality of synthesized images. To this end, a combined metric M_C was developed, based on the M_{IS} and M_{FID} metrics. Additionally, an algorithm for biomedical image synthesis using diffusion models was proposed. The synthesized biomedical images of three breast pathology classes were evaluated using the M_{IS}, M_{FID}, and M_C metrics.

The training and generation time for HIs using diffusion models was reduced by an order of magnitude compared to GANs. High-quality synthetic images were obtained for each histopathological class (G1, G2 and G3),

DOI 10.15588/1607-3274-2025-2-15

demonstrating good performance in terms of the M_{IS}, M_{FID}, and M_C metrics.

The **scientific novelty** of this study lies in the development of a combined metric for evaluating the quality of synthesized images.

The **practical significance** of this research includes the development of a HI synthesis algorithm, generation of HIs for three classes, and the execution of computational experiments to assess the quality of the synthesized images.

Future research directions involve exploring new biomedical image synthesis algorithms based on diffusion models and fuzzy metrics for quality assessment of synthesized images.

ACKNOWLEDGEMENTS

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

REFERENCES

1. Cancer Facts for Women. American Cancer Society. [Electronic resource]. Access mode: https://www.cancer.org/cancer/risk-prevention/understanding-cancer-risk/cancer-facts/cancer-facts-for-women.html
2. Cancer in Ukraine 2021–2022: Incidence, mortality, prevalence and other relevant statistics. Bulletin of the National Cancer Registry of Ukraine № 24, 2021–2022. [Electronic resource]. Access mode: http://www.ncru.inf.ua/publications/BULL_24/PDF_E/bull_eng_24.pdf
3. Berezsky O., Liashchynskyi P., Pitsun O., Izonin I. Synthesis of Convolutional Neural Network architectures for biomedical image classification, *Biomedical Signal Processing and Control,* 2024, №95, P. 106325. DOI: 10.1016/j.bspc.2024.106325
4. Nie D., Cao X., Gao Y., Wang L., Shen D. Estimating CT image from MRI data using 3d fully convolutional networks *In: Deep Learning and Data Labeling for Medical Applications*. Springer International Publishing, 2016, pp. 170–178. DOI: 10.1007/978-3-319-46976-8_18
5. Nie D., Trullo R., Lian J., Petitjean C., Ruan S., Wang Q., Shen D. Medical image synthesis with context-aware generative adversarial networks, *In: Medical Image Computing and Computer Assisted Intervention – MICCAI 2017*. Springer International Publishing, 2017, pp. 417–425. DOI: 10.1007/978-3-319-66179-7_48
6. Khader F., Müller-Franzes G., Arasteh S. T., Han T., Haarburger C., Schulze Hagen M., Schad P., Engelhardt S., Baeßler B., Foersch S., Stegmaier J., Kuhl C., Nebelung S., Kather J. N., Truhn D. Denoising diffusion probabilistic models for 3d medical image generation, *Sci. Rep. 2023 May 5,* Vol. 13, № 1, P. 7303. DOI: 10.1038/s41598-023-34341-2
7. Xu Q., Huang G., Yuan Y., Guo C., Sun Y., Wu F., Weinberger K. An empirical study on evaluation metrics of generative adversarial networks, *ArXiv1806.07755 Cs Stat*, 2018. DOI: 10.48550/arXiv.1806.07755
8. Borji A. Pros and cons of GAN evaluation measures, *Computer Vision and Image Understanding*, 2019, Vol. 179, pp. 41–65. DOI: 10.1016/j.cviu.2018.10.009
9. Rendon N., Giraldo J. H., Bouwmans T., Rodríguez-Buritica S., E. Ramirez, Isaza C. Uncertainty clustering internal validity assessment using Fréchet distance for unsupervised learning, *Engineering Applications of Artificial Intelligence*, 2023, Vol. 124, P. 106635. DOI: 10.1016/j.engappai.2023.106635
10. Buzuti L. F., Thomaz C. E. Fréchet AutoEncoder Distance: A new approach for evaluation of Generative Adversarial Networks, *Computer Vision and Image Understanding,* 2023, Vol. 235, P. 103768. DOI: 10.1016/j.cviu.2023.103768
11. Heusel M., Ramsauer H., Unterthiner T., Nessler B., Hochreiter S. GANs Trained by a Two Time-Scale Update Rule Converge to a Local Nash Equilibrium [Electronic resource]. Access mode: https://arxiv.org/abs/1801.01401. DOI:10.48550/arXiv.1801.01401
12. Bińkowski M., Sutherland Danica J., Arbel Michael, Gretton Arthur Demystifying MMD GANs [Electronic resource], Access mode: https://arxiv.org/abs/1801.01401. DOI:10.48550/arXiv.1801.01401
13. Chong M. J., Forsyth D. Effectively Unbiased FID and Inception Score and where to find them [Electronic resource]. Access mode: https://arxiv.org/abs/1911.07023v3. DOI: 10.48550/arXiv.1911.07023
14. Parmar G. Zhang R., Zhu J.-Y. On Aliased Resizing and Surprising Subtleties in GAN Evaluation [Electronic resource]. Access mode: https://arxiv.org/abs/2104.11222v3. DOI: 10.48550/arXiv.2104.11222
15. Jayasumana S., Ramalingam S., Veit A., Glasner D., Chakrabarti A., Kumar S. Rethinking FID: Towards a Better Evaluation Metric for Image Generation [Electronic resource]. Access mode: https://arxiv.org/abs/2401.09603v2. DOI: 10.48550/arXiv.2401.09603
16. Betzalel E., Penso C., Fetaya E. Evaluation Metrics for Generative Models: An Empirical Study, *Machine Learning and Knowledge Extraction,* 2024, Vol. 6, № 3, pp. 1531–1544. DOI: 10.3390/make6030073
17. Benny Y., Galanti T., Benaim S., Wolf L. Wolf Evaluation Metrics for Conditional Image Generation, *International Journal of Computer Vision,* 2021, Vol. 129, № 5, pp. 1712–1731. DOI: 10.1007/s11263-020-01424-w
18. Arjovsky M., Chintala S., Bottou L. Wasserstein GAN [Electronic resource]. Access mode: https://arxiv.org/abs/1701.07875. DOI: 10.48550/ARXIV.1701.07875
19. Dayarathna S., Islam K. T., Uribe S., Yang G., Hayat M., Chen Z. Deep learning based synthesis of MRI, CT and PET: Review and analysis, *Medical Image Analysis,* 2024, Vol. 92, P. 103046. DOI: 10.1016/j.media.2023.103046

DOI 10.15588/1607-3274-2025-2-15

20. Luz D. S., Lima T. J. B., Silva R. R. V., Magalhães D. M. V., Araujo F. H. D. Automatic detection metastasis in breast histopathological images based on ensemble learning and color adjustment, *Biomedical Signal Processing and Control,* 2022, Vol. 75, P. 103564. DOI: 10.1016/j.bspc.2022.103564
21. Jiménez-Gaona Y., Carrión-Figueroa D., Lakshminarayanan V., Rodríguez-Álvarez M. José GAN-based data augmentation to improve breast ultrasound and mammography mass classification, *Biomedical Signal Processing and Control,* 2024, Vol. 94, P. 106255. DOI: 10.1016/j.bspc.2024.106255
22. Ukwuoma C. C., Cai D., Eziefuna E. O., Oluwasanmi A., Abdi Sabirin F., Muoka G. W., Thomas D., Sarpong K. Enhancing histopathological medical image classification for Early cancer diagnosis using deep learning and explainable AI – LIME & SHAP, *Biomedical Signal Processing and Control*, 2025, Vol. 100, P. 107014. DOI: 10.1016/j.bspc.2024.107014
23. Shi S., Li H., Zhang Y., Wang X. Semantic information-guided attentional GAN-based ultrasound image synthesis method, *Biomedical Signal Processing and Control,* 2025, Vol. 102, P. 107273. DOI: 10.1016/j.bspc.2024.107273
24. Zeng X., Lu B., Zhang J. Medical image synthesis algorithm based on vision graph neural network with manifold matching, *Biomedical Signal Processing and Control*, 2025, Vol. 103, P. 107381. DOI: 10.1016/j.bspc.2024.107381
25. Hu Y., Zhang S., Li W., Sun J., Xu L. X. Unsupervised medical image synthesis based on multi-branch attention structure, *Biomedical Signal Processing and Control*, 2025, Vol. 104, P. 107495. DOI: 10.1016/j.bspc.2025.107495
26. Jose L., Liu S., Russo C., Nadort A., Di Ieva Antonio Generative Adversarial Networks in Digital Pathology and Histopathological Image Processing: A Review, *Journal of Pathology Informatics,* 2021, Vol. 12, № 1, P. 43. DOI: 10.4103/jpi.jpi_103_20
27. Giri S. K., Dash S. Synthesis of clinical images for oral cancer detection and prediction using deep learning, *Mining Biomedical Text, Images and Visual Features for Information Retrieval,* 2025, pp. 339–356. DOI: 10.1016/b978-0-443-15452-2.00017-0
28. Berezsky O. M., Pitsun O. Y. Evaluation methods of image segmentation quality, *Radio Electronics, Computer Science, Control,* 2018, №1, pp. 119–128. DOI: 10.15588/1607-3274-2018-1-14
29. Berezsky O. M., Liashchynskyi P. B. Method of generative-adversarial networks searching architectures for biomedical images synthesis, *Radio Electronics, Computer Science, Control*, 2024, № 1, pp. 104–117. DOI https://doi.org/10.15588/1607-3274-2024-1-10
30. Berezsky O. M., Liashchynskyi P. B., Pitsun O. Y., Melnyk G. M. Deep network-based method and software for small sample biomedical image generation and classification, *Radio Electronics, Computer Science, Control,* 2023, №4,P. 76. DOI: 10.15588/1607-3274-2023-4-8
31. Rantalainen M., Hartman J. ACROBAT – a multi-stain breast cancer histological whole-slide-image data set from routine diagnostics for computational pathology. Karolinska Institutet. [Electronic resource]. Access mode: https://doi.org/10.48723/w728-p041. DOI:10.48723/w728-p041
32. Wan Siti Halimatul Munirah Wan Ahmad, Mohammad Faizal Ahmad Fauzi, Md Jahid Hasan, Zaka Ur Rehman, Jenny Tung Hiong Lee, See Yee Khor, Lai Meng Looi, Fazly Salleh Abas, Afzan Adam, Elaine Wan Ling Chan, Sei-ichiro Kamata UMMC ER-IHC Breast Histopathology Whole Slide Image and Allred Score IEEE Dataport, 2023. [Electronic resource]. Access mode: https://dx.doi.org/10.21227/9gbq-gz50. DOI: 10.21227/9gbq-gz50
33. Berezsky O., Datsko T., Melnyk G. Cytological and histological images of breast cancer [Electronic resource]. Access mode: https://zenodo.org/records/7890874. DOI: 10.5281/zenodo.7890873
34. Berezsky O., Liashchynskyi P., Melnyk G., Dombrovskyi M., Berezkyi M. Synthesis of biomedical images based on generative intelligence tools, *Informatics & Data-Driven Medicine (IDDM 2024): 7th International Conference, Birmingham, UK, November 14–16, 2024, CEUR Workshop Proceedings*, 2024, Vol. 3892, pp. 349–362. DOI: CEUR_WS.org/Vol-3892/paper23.pdf
35. Deza M. M., Deza E. Encyclopedia of Distances. Springer-Verlag Berlin Heidelberg, 2013, 583 p.
36. GitHub – liashchynskyi_rudi_ Lightweight image converter and dataset augmentor, 2024. [Electronic resource]. Access mode: https://github.com/liashchynskyi/rudi

Received 29.01.2025.
Accepted 25.04.2025.

DOI 10.15588/1607-3274-2025-2-15

УДК 004.932

КОМБІНОВАНА МЕТРИКА ДЛЯ ОЦІНКИ ЯКОСТІ СИНТЕЗОВАНИХ БІОМЕДИЧНИХ ЗОБРАЖЕНЬ

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

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

Домбровський М. О. – магістр з комп'ютерної інженерії Західноукраїнського національного університету, Тернопіль, Україна.

Лящинський П. Б. – д-р філософії з комп'ютерних наук, викладач кафедри комп'ютерної інженерії Західноукраїнського національного університету, Тернопіль, Україна.

Мельник Г. М. – канд. техн. наук, доцент кафедри комп'ютерної інженерії Західноукраїнського національного університету, Тернопіль, Україна.

АНОТАЦІЯ

Актуальність. У статті досліджено проблему розробки нової метрики для оцінки якості синтезованих зображень. Актуальність проблеми пояснюється необхідністю оцінки якості штучних зображень. Крім цього у роботі показано перспективність синтезу біомедичних зображень на основі дифузійних моделей. Результати дослідження можуть бути використані для синтезу біомедичних зображень та кількісної оцінки якості синтезованих зображень.

Мета роботи – розробка комбінованої метрики та алгоритму синтезу біомедичних зображень для оцінки якості синтезованих зображень.

Метод. У статті розроблено комбіновану метрику M_C для оцінки якості синтезованих зображень. Ця метрика базується на основі двох метрик M_{IS}, M_{FID}. Також розроблено алгоритм синтезу гістопатологічних зображень на основі дифузійних моделей.

Результати. Для дослідження метрик M_{IS}, M_{FID} і M_C використано гістопатологічні зображення, які знаходяться на платформі Zenodo. Цей dataset містить три класи G1, G2, G3 гістопатологічних зображень патологічних станів молочної залози. На основі розробленого алгоритму синтезу зображень отримано три класи штучних гістопатологічних зображень.

На основі метрик M_{IS}, M_{FID} і M_C отримано оцінки якості синтезованих гістопатологічних зображень: Розроблена метрика вийде в основу програмного модуля для оцінки якості зображень на основі метрик. Цей програмний модуль буде інтегрований у CAD.

Висновки. Розроблена комбінована метрика для оцінки якості синтезованих зображень і запропонований алгоритм синтезу біомедичних зображень. Програмна реалізація комбінованої метрики і алгоритму синтезу зображень інтегровані у модуль оцінки якості зображень.

КЛЮЧОВІ СЛОВА: метрика, метрики IS, FID, гістопатологічні зображення, глибокі нейронні мережі, дифузійні моделі, Stable Diffusion.

ЛІТЕРАТУРА

1. Cancer Facts for Women. American Cancer Society. [Electronic resource]. – Access mode: https://www.cancer.org/cancer/risk-prevention/understanding-cancer-risk/cancer-facts/cancer-facts-for-women.html
2. Cancer in Ukraine 2021–2022: Incidence, mortality, prevalence and other relevant statistics. Bulletin of the National Cancer Registry of Ukraine № 24, 2021–2022. [Electronic resource]. – Access mode: http://www.ncru.inf.ua/publications/BULL_24/PDF_E/bull_eng_24.pdf
3. Synthesis of Convolutional Neural Network architectures for biomedical image classification / [O. Berezsky, P. Liashchynskyi, O. Pitsun, I. Izonin] // Biomedical Signal Processing and Control. – 2024. – №95. – P. 106325. DOI: 10.1016/j.bspc.2024.106325
4. Estimating CT image from MRI data using 3d fully convolutional networks / [D. Nie, X. Cao, Y. Gao et al.] // In: Deep Learning and Data Labeling for Medical Applications. Springer International Publishing, 2016. – P. 170–178. DOI: 10.1007/978-3-319-46976-8_18
5. Medical image synthesis with context-aware generative adversarial networks / [D. Nie, R. Trullo, J. Lian et al.] // In: Medical Image Computing and Computer Assisted Intervention – MICCAI 2017. Springer International Publishing, 2017. – P. 417–425. DOI: 10.1007/978-3-319-66179-7_48
6. Denoising diffusion probabilistic models for 3d medical image generation / [F. Khader, G. Müller-Franzes, S. T. Arasteh et al.] // Sci. Rep. 2023 May 5. – Vol. 13, №1. – P. 7303. DOI: 10.1038/s41598-023-34341-2
7. An empirical study on evaluation metrics of generative adversarial networks / [Q. Xu, G. Huang, Y. Yuan et al.] // ArXiv1806.07755 Cs Stat, 2018. DOI: 10.48550/arXiv.1806.07755
8. Borji A. Pros and cons of GAN evaluation measures / A. Borji // Computer Vision and Image Understanding. – 2019. – Vol. 179. – P. 41–65. DOI: 10.1016/j.cviu.2018.10.009
9. Uncertainty clustering internal validity assessment using Fréchet distance for unsupervised learning / [N. Rendon, J. H. Giraldo, T. Bouwmans et al.] // Engineering Applications of Artificial Intelligence. – 2023. – Vol. 124. – P. 106635. DOI: 10.1016/j.engappai.2023.106635
10. Buzuti L. F. Fréchet AutoEncoder Distance: A new approach for evaluation of Generative Adversarial Networks / L. F. Buzuti, C. E. Thomaz // Computer Vision and Image Understanding. – 2023. – Vol. 235 – P. 103768. DOI: 10.1016/j.cviu.2023.103768
11. GANs Trained by a Two Time-Scale Update Rule Converge to a Local Nash Equilibrium [Electronic resource] / [M. Heusel, H. Ramsauer, T. Unterthiner et al.] – Access mode: https://arxiv.org/abs/1801.01401. DOI:10.48550/arXiv.1801.01401

DOI 10.15588/1607-3274-2025-2-15

12. Demystifying MMD GANs [Electronic resource] / [M. Bińkowski, Danica J. Sutherland, Michael Arbel, Arthur Gretton] – Access mode: https://arxiv.org/abs/1801.01401. DOI:10.48550/arXiv.1801.01401
13. Chong M. J. Effectively Unbiased FID and Inception Score and where to find them [Electronic resource] / M. J. Chong, D. Forsyth. – Access mode: https://arxiv.org/abs/1911.07023v3. DOI: 10.48550/arXiv.1911.07023
14. Parmar G. On Aliased Resizing and Surprising Subtleties in GAN Evaluation [Electronic resource] / G. Parmar, R. Zhang, J.-Y. Zhu. – Access mode: https://arxiv.org/abs/2104.11222v3. DOI: 10.48550/arXiv.2104.11222
15. Rethinking FID: Towards a Better Evaluation Metric for Image Generation [Electronic resource] / [S. Jayasumana, S. Ramalingam, A. Veit et al.]. – Access mode: https://arxiv.org/abs/2401.09603v2. DOI: 10.48550/arXiv.2401.09603
16. Betzalel E. Evaluation Metrics for Generative Models: An Empirical Study / E. Betzalel, C. Penso, E. Fetaya // Machine Learning and Knowledge Extraction. – 2024. – Vol. 6, № 3. – P. 1531–1544. DOI: 10.3390/make6030073
17. Wolf Evaluation Metrics for Conditional Image Generation / [Y. Benny, T. Galanti, S. Benaim, L. Wolf] // International Journal of Computer Vision. – 2021. – Vol. 129, № 5. – P. 1712–1731. DOI: 10.1007/s11263-020-01424-w
18. Arjovsky M. Wasserstein GAN [Electronic resource] / M. Arjovsky, S. Chintala, L. Bottou. – Access mode: https://arxiv.org/abs/1701.07875. DOI: 10.48550/ARXIV.1701.07875
19. Deep learning based synthesis of MRI, CT and PET: Review and analysis / [S. Dayarathna, K. T. Islam, S. Uribe et al.] // Medical Image Analysis. – 2024. – Vol. 92. – P. 103046. DOI: 10.1016/j.media.2023.103046
20. Automatic detection metastasis in breast histopathological images based on ensemble learning and color adjustment / [D. S. Luz, T. J. B. Lima, R. R.V. Silva et al.] // Biomedical Signal Processing and Control. – 2022. – Vol. 75. – P. 103564. DOI: 10.1016/j.bspc.2022.103564
21. GAN-based data augmentation to improve breast ultrasound and mammography mass classification / [Y. Jiménez-Gaona, D. Carrión-Figueroa, V. Lakshminarayanan, M. José Rodríguez-Álvarez] // Biomedical Signal Processing and Control. – 2024. – Vol. 94. – P. 106255. DOI: 10.1016/j.bspc.2024.106255
22. Enhancing histopathological medical image classification for Early cancer diagnosis using deep learning and explainable AI – LIME & SHAP / [C. C. Ukwuoma, D. Cai, E. O. Eziefuna et al.] // Biomedical Signal Processing and Control. – 2025. – Vol. 100. – P. 107014. DOI: 10.1016/j.bspc.2024.107014
23. Semantic information-guided attentional GAN-based ultrasound image synthesis method / [S. Shi, H. Li, Y. Zhang, X. Wang] // Biomedical Signal Processing and Control. – 2025. – Vol. 102. – P. 107273. DOI: 10.1016/j.bspc.2024.107273
24. Zeng X. Medical image synthesis algorithm based on vision graph neural network with manifold matching / X. Zeng, B. Lu, J. Zhang // Biomedical Signal Processing and Control. – 2025. – Vol. 103. – P. 107381. DOI: 10.1016/j.bspc.2024.107381
25. Unsupervised medical image synthesis based on multi-branch attention structure / [Y. Hu, S. Zhang, W. Li et al.] // Biomedical Signal Processing and Control. –2025. – Vol. 104 – P. 107495. DOI: 10.1016/j.bspc.2025.107495
26. Generative Adversarial Networks in Digital Pathology and Histopathological Image Processing: A Review / [L. Jose, S. Liu, C. Russo et al.] // Journal of Pathology Informatics. – 2021. – Vol. 12, № 1. – P. 43. DOI: 10.4103/jpi.jpi_103_20
27. Giri S. K. Synthesis of clinical images for oral cancer detection and prediction using deep learning / S. K. Giri, S. Dash // Mining Biomedical Text, Images and Visual Features for Information Retrieval. – 2025. – P. 339–356. DOI: 10.1016/b978-0-443-15452-2.00017-0
28. Berezsky O. M. Evaluation methods of image segmentation quality / O. M. Berezsky, O. Y. Pitsun // Radio Electronics, Computer Science, Control. – 2018. – №1. – P. 119–128. DOI: 10.15588/1607-3274-2018-1-14
29. Berezsky O. M. Method of generative-adversarial networks searching architectures for biomedical images synthesis / O. M. Berezsky, P. B. Liashchynskyi // Radio Electronics, Computer Science, Control. – 2024. – № 1. – P. 104–117. DOI https://doi.org/10.15588/1607-3274-2024-1-10
30. Deep network-based method and software for small sample biomedical image generation and classification / [O. M. Berezsky, P. B. Liashchynskyi, O. Y. Pitsun, G. M. Melnyk] // Radio Electronics, Computer Science, Control. – 2023. – №4. – P. 76. DOI: 10.15588/1607-3274-2023-4-8
31. Rantalainen M. ACROBAT – a multi-stain breast cancer histological whole-slide-image data set from routine diagnostics for computational pathology. Karolinska Institutet. [Electronic resource] / M. Rantalainen, J. Hartman. – Access mode: https://doi.org/10.48723/w728-p041. DOI:10.48723/w728-p041
32. UMMC ER-IHC Breast Histopathology Whole Slide Image and Allred Score IEEE Dataport, 2023. [Electronic resource] / [Wan Siti Halimatul Munirah Wan Ahmad, Mohammad Faizal Ahmad Fauzi, Md Jahid Hasan et al.] – Access mode: https://dx.doi.org/10.21227/9gbq-gz50. DOI: 10.21227/9gbq-gz50
33. Berezsky O. Cytological and histological images of breast cancer [Electronic resource] / O. Berezsky, T. Datsko, G. Melnyk – Access mode: https://zenodo.org/records/7890874. DOI: 10.5281/zenodo.7890873
34. Synthesis of biomedical images based on generative intelligence tools / [O. Berezsky, P. Liashchynskyi, G. Melnyk et al.] // Informatics & Data-Driven Medicine (IDDM 2024): 7th International Conference, Birmingham, UK, November 14–16, 2024, CEUR Workshop Proceedings, 2024. – Vol. 3892. – P. 349–362. DOI: CEUR_WS.org/Vol-3892/paper23.pdf
35. Deza M. M. Encyclopedia of Distances / M. M. Deza, E. Deza. – Springer-Verlag Berlin Heidelberg, 2013. – 583 p.
36. GitHub – liashchynskyi_rudi_ Lightweight image converter and dataset augmentor, 2024. [Electronic resource]. – Access mode: https://github.com/liashchynskyi/rudi

DOI 10.15588/1607-3274-2025-2-15