http://dx.doi.org/10.35596/1729-7648-2025-23-1-47-53

*Original paper*

UDC 004.8; 004.8.032.26

# A METHOD FOR DETECTING DISTINCTIVE PATTERNS OF REAL PATIENTS IN GENERATED IMAGES

VASSILI A. KOVALEV

*The United Institute of Informatics Problems of the National Academy of Sciences of Belarus (Minsk, Republic of Belarus)*



**Abstract.** Generative diffusion models are a well-established method for generating high-quality images. However, there are studies that show that diffusion models are less privacy-friendly than generative models, such as generative adversarial networks and a growing family of their modifications. The discovered vulnerabilities require in-depth study of various security aspects. This is especially important for sensitive areas such as medical image analysis tasks and their practical applications. The paper describes a method for detecting image patterns presented in generated images that can potentially be identified in real CT images of patients with pulmonary tuberculosis. The method includes the following main procedures: correlation of pairs of generated and real images to pre-select pairs that involve further analysis; calculation of correlation statistics using direct and inverse Fisher transforms; performing affine image registration and calculating pairwise similarity scores; nonlinear (elastic) image registration and recalculation of similarity scores to highlight the most similar/dissimilar image areas.

**Keywords:** diffusion generative models, computed tomography, privacy preserving.

**Conflict of interests.** The author declares no conflict of interests.

**For citation.** Kovalev V. A. (2025) A Method for Detecting Distinctive Patterns of Real Patients in Generated Images. *Doklady BGUIR*. 23 (1), 47–53. http://dx.doi.org/10.35596/1729-7648-2025-23-1-47-53.

# МЕТОД ОБНАРУЖЕНИЯ ХАРАКТЕРНЫХ ПАТТЕРНОВ РЕАЛЬНЫХ ПАЦИЕНТОВ НА СГЕНЕРИРОВАННЫХ ИЗОБРАЖЕНИЯХ

В. А. КОВАЛЕВ

*Объединенный институт проблем информатики Национальной академии наук Беларуси (г. Минск, Республика Беларусь)*

**Аннотация.** Генеративные диффузионные модели являются общепризнанным методом генерации высококачественных изображений. Однако среди исследований есть примеры, подтверждающие, что диффузионные модели менее конфиденциальны, чем генеративные модели, такие как генеративные состязательные сети и растущее семейство их модификаций. Обнаруженные уязвимости требуют глубокого изучения различных аспектов безопасности. Это особенно важно для таких чувствительных областей, как задачи анализа медицинских изображений и их практическое применение. В статье рассмотрен метод обнаружения шаблонов изображений, представленных на сгенерированных изображениях, которые потенциально могут быть идентифицированы на реальных изображениях компьютерной томографии пациентов с туберкулезом лёгких. Метод включает следующие основные процедуры: корреляция пар сгенерированных и реальных изображений для предварительного выбора пар, которые предполагают дальнейший анализ; вычисление статистики корреляции с использованием прямого и обратного преобразований Фишера; выполнение аффинной регистрации изображений и расчет оценок парного сходства; нелинейная (эластичная) регистрация изображений и повторный расчет оценок сходства для выделения наиболее похожих/несхожих областей изображения.

**Ключевые слова:** диффузионные генеративные модели, компьютерная томография, сохранение конфиденциальности.

**Конфликт интересов.** Автор заявляет об отсутствии конфликта интересов.

**Для цитирования.** Ковалев, В. А. Метод обнаружения характерных паттернов реальных пациентов на сгенерированных изображениях / В. А. Ковалев // Доклады БГУИР. 2025. Т. 23, № 1. С. 47–53. http://dx.doi.org/10.35596/1729-7648-2025-23-1-47-53.

## Introduction

The generated medical images could substitute the real ones in different scenarios of development and use of diverse deep learning applications [1–3]. In this particular work, we considering the computed tomography (CT) images created by a denoising diffusion model. This type of generative models represents an emerging class of generative neural networks that produce images from a training distribution via an iterative denoising process [4, 5]. Compared to the previous image generation approaches, such as commonly known generative adversarial networks (GANs) together with the growing family of their modifications and variational autoencoders, diffusion models produce higher-quality samples that easier to scale and control [6].

Consequently, diffusion models have rapidly become the de-facto method for generating high-quality and high-resolution images. Nevertheless, in their work [6] authors state that diffusion models are less private than prior generative models such as GANs, and that mitigating these vulnerabilities may require new advances in privacy-preserving training. In this respect, one of the main goals of present study was to estimate the chances of disclosing private image patterns under specific conditions we are working in. In particular, a need for image generation could appear in both cases including very large and relatively small image training sets. Therefore, it was necessary to give certain priority to computational experiments, which reveal the influence of the size of image datasets used for training generative models to the probability of possible low data security.

In general, there are many kinds of image properties that can be utilized for distinguishing radiological images of a given modality. They include differences caused by specific "signatures" of images that characteristic for certain brands of radiological imaging devices, differences associated with age and gender of patients, various body lesions, specific image features coming from the artificial objects presented in the body (e. g., cardio-stimulators, objects and traces left after underwent surgery), as well as differences induced by the patients' pose and body deformations, atypical anatomy, and so on.

In this work, we limit ourselves by the methods that searching similarities associated with the patients themselves. In other words, we concentrating on the problem of detection of possible "fingerprints" (image patterns) presented in generated images that can be potentially identified in the radiological images of real lung tuberculosis patients. Similar to many others, such patterns are composed of the spatial (say, shape) and the intensity features. It is important to note that we admit that the patterns of interest could be slightly deformed in both spatial and the intensity domains due to the factors mentioned above. Nevertheless, they should also be identified and accompanied by an appropriate quantitative measure.

## Image data

*Original image data.* We used 2D axial slices of 3D CT image datasets of tuberculosis patients that satisfy to all the existing regulations, limitations, and the agreements. All the images were acquired on the same CT machine and anonymized in due course before any steps of their computerized analysis. There were no ways for disclosure, share, and other means of dissemination of personal patients' data. The main steps of CT image data preparation procedure described below.

It is obvious that different axial slices (sections) of the body are anatomically different. However, for the human eye and corresponding quantitative features they appear reasonably similar. The similarity rates depend on different factors such as the specific location of sections in the human body, the individual anatomy of each patient, the CT scanning protocol (e. g., slice thickness), and several others. However, considering the fact that in this study we are extensively using convolutional neural networks, these anatomical variations can be treated as a sort of "natural anatomical image augmentation". Clearly, such natural augmentation is far better than any artificial one produced by common software libraries. Creation of the dataset of CT image slices was done in four main steps:

– the creation of the initial dataset is based on a large collection of CT images containing up to 10,714 CT scans (approximately 0.7 TB of DICOM image data);

– converting all image files from DICOM to Niftii (also known as nii.gz). No personal data in the commonly used definitions is presented;

– excluding from the original 10,714 CT scans those that do not have information on the patient's age and gender. The resulting 8,463 CT scans included 4,662 males and 3,801 females. In total, they amounted to 288 GB, losslessly compressed;

– splitting all selected 8,463 3D images into 2D axial slices. The result was 1,002,012 2D images of 512×512 pixel sections (574,309 men and 427,703 women). They were exported to lossless PNG format with an intensity of 8 bits/pixel.

*The study image datasets.* Considering the importance of the training set size in the problem of detecting generated image patterns that inherit from the real patient images, we created 5 pairs of image datasets. The size of training sets gradually increased and consisted of 6, 60, 600, 6000, and 60 000 images respectively. Sizes of the sets of corresponding generated images were fixed to 240 items in each of 5 pairs. In each occasion the 240-sized subsets of generated images were randomly sampled from larger sets of artificial images generated based on their respective "parental" sets of the real ones. Image generation was done using diffuse generative neural models. Examples are provided in Fig. 1.

All training image datasets were perfectly balanced by gender, i. e., consisted of 50 % of female and 50 % of male subjects. Also, all the patients were aged of 26 complete years of life except for the ones included in the dataset containing 60 000 images which composed of people aged 24–39 years. The five versions of generative neural networks were trained on each of five training sets separately.

In order to provide the necessary variability of images used for training the diffuse generative models, we conditionally sub-divided image slices into the following three anatomical categories which are conditionally referred to as "classes" (Fig. 1):

c1: The upper part of liver;

c2: The heart class, which was represented by a middle heart section plus some limited amount of adjacent axial slices along with *Z* neighborhood;

c3: The shoulders which include the upper part of lungs and their close neighboring sections above and below them along with *Z* axis.

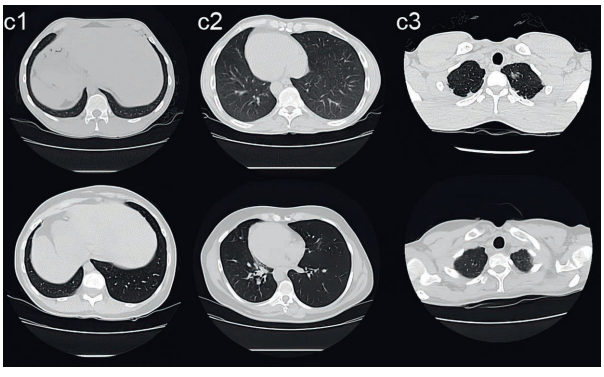



**Fig. 1.** Examples of original (top row) and generated (bottom row) images

It should be noted that it was impossible to consider other distinct anatomical sections such as the ones situated at the neck and kidney levels. This is because all the patients were suffering from lung tuberculosis and therefore the patients were scanned only within the regions of lungs plus few additional safety slices.

### Evaluating the difference of real and generated images using statistical metrics

Statistical metrics are used at the preliminary stage of image comparison based on correlation measures computed over the whole images included into the real-vs-generated pairs. The main goal of this stage is to select candidates for the further, more thorough analysis.

*Details of computing raw correlation coefficients.* We start with pixel-wise Pearson correlation of the images. It is easy to see that this case even such simple image miss-match as relative shift of generated images compared to the original ones may lead to a dramatic drop of correlation coefficient value. Same consequences may have place with mutual rotation. Simply speaking, even comparison of the shifted/rotated versions of a generated image with absolutely identical real image may not indicate the complete image duplication by the 1.0 correlation coefficient. Thus, prior to the comparison of images of pair, we apply the rigid-body affine transform (rotation, translation, scaling) to the generated

image which brings them to the best possible correspondence. Besides, this operation does not perform any local, non-linear deformations of image patterns.

Once generated and real images are overlapped and compared with each other, the image pair with maximal correlation is passed to the further analysis. It is important to note that each generated image is compared with the real ones but not vice versa. This is because the only set of real images represents the whole variety of possible image features being analyzed in the given experimental setup. However, any set (portion) of generated images represent only a small fraction of the practically unlimited number of generated samples.

The maximal values of correlation of original and generated images as a function of the size of five training sets are depicted in Fig. 2. Note that to ease perception and interpretation of correlation results, the correlation coefficient values of all 240 generated images are presented not in their original random order but sorted in ascending order of correlation coefficient to ease the visual perception.

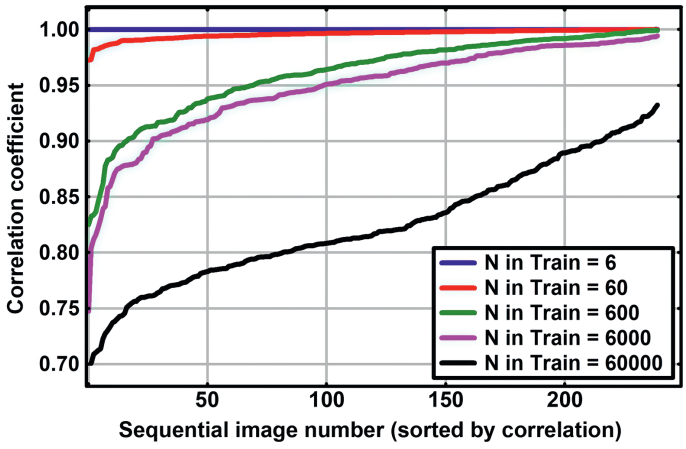



**Fig. 2.** Dependence of max pair-wise correlations on training set size for 240 generated images taken at random

It is easy to note from Fig. 2 that generative model trained on tiny image set consisting of six images reproduces practically exact copies of real images. However, the variability of generated images and their deviation from the images placed in the training set is growing up with the increase of training set size. This is evident from the systematic shift of plots downwards. In particular, for the training set consisted of 60 000 images the Minimum value of Maximal correlation drops down to 0.678.

Statistics of Pearson correlation coefficient. The regularity which was noticed above by way of visual analysis of correlation curves can be simply expressed and easily captured by comparison of mean correlation values and other basic statistics. However, corresponding calculations may not be done directly. The reason is that arithmetic manipulations with Pearson's correlation coefficients (e. g., calculation of the mean and percentiles) are not mathematically correct [7]. This is because when the correlation coefficient is close to 1.0, what is often the case in this study, its distribution is highly skewed. Such a property makes it difficult to estimate the confidence intervals and apply significance tests to the correlation coefficients for original and generated image datasets. The Fisher's transformation solves this problem by yielding a variable whose distribution is approximately normal, with a variance which is stable over different values of correlations. Mathematically, the Fisher's $z$-transform is an inverse hyperbolic tangent (artanh). For converting the correlation coefficients into $z$-scores and back we employed the direct and reverse versions of $z$-transform as implemented in the DescTools package of R [8] wherever necessary. The resultant statistical characteristics of Fisher's $z$-scores and Persons' correlation coefficients with respect to the train set size are given in Tab. 1.

**Table 1.** Changes of statistical significance values of $z$-score and correlation $r$ with respect to the train set size

| Train set size | Mean $z$/Mean $r$ | Min $z$/Min $r$ | Max $z$/Max $r$ | STD $z$/STD $r$ |
|---|---|---|---|---|
| 6 | 5.091/0.9999 | 4.720/0.9998 | 5.358/0.9999 | 0.1344/0.1336 |
| 60 | 3.316/0.9977 | 2.135/0.9724 | 4.482/0.9997 | 0.5186/0.4766 |
| 600 | 2.211/0.9763 | 1.172/0.8248 | 3.810/0.9990 | 0.5762/0.5199 |
| 6000 | 1.968/0.9617 | 0.967/0.7473 | 2.937/0.9944 | 0.4231/0.3995 |
| 60000 | 1.196/0.8323 | 0.825/0.6776 | 1.674/0.9321 | 0.1830/0.1810 |

The graphical illustration of Fisher's *z*-scores of the statistical significance of differences of real and generated images as a function of train set size is depicted in Fig. 3.

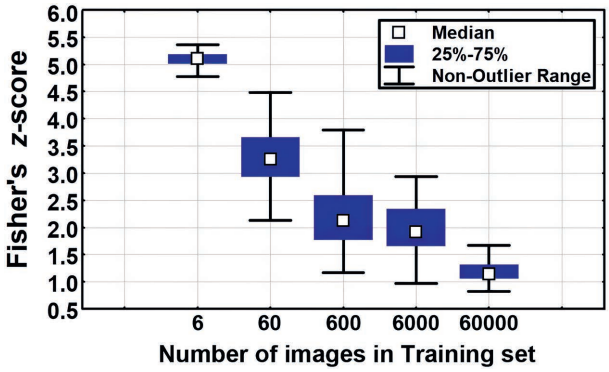



**Fig. 3.** *z*-scores of the significance of differences of real and generated images

**Assessing the difference of real and generated images using Fréshet Inception Distance score**

The Fréshet Inception Distance (FID) [9] can be viewed as a further development of the Inception metric which is based on the Inception V3 neural network model. The FID metric essentially is an algorithm that combines two different, partly controversial, integral measure of the quality of image generation. These components are:

– a metric of the "realism" of generated images;

– quantitative measure of the diversity of generated images.

The above term "realism" can be understood to mean that the generated images generally look like the real ones to a human expert. It is difficult to distinguish them visually and, with certain limitations, they can replace the real patient images even in different computation scenarios. Note that according to the definition, FID score may not be computed for very small image datasets.

Results of computing FID score of four largest test image datasets examined in this study are presented in Tab. 2. Note that the smaller FID values, the better the results of generation are. Also, the smallest training set contained only six images is excluded from consideration here because this case generated images are appearing pretty much as the exact copies of training images and the calculated FID score values are unstable.

**Table 2.** FID scores for relatively large datasets examined in this study

| Train set size | 60 | 600 | 6000 | 60 000 |
|---|---|---|---|---|
| Fréshet Inception Distance | 70.08 | 17.02 | 14.45 | 10.71 |

It is easy to see that the quality of generated images, especially in relation to their variability (Fig. 2, Tab. 1) is growing up while increasing the training sets. This is well agreed with the other results reported above.

In addition to the comparison of FID scores for gradually increasing four training sets, we performed an experiment to testify that FID score is sensitive enough to recognize whether a set of generated images was obtained by the given set of real parental ones. This test was accomplished with the help of seven image sets of randomly sampled real images and seven sets of corresponding generated ones. Each of seven real image sets consisted of 600 images of patients aged 26 years including exactly 300 females and 300 males. Results are summarized in Fig. 4.

| | Gen-1 | Gen-2 | Gen-3 | Gen-4 | Gen-5 | Gen-6 | Gen-7 |
|---|---|---|---|---|---|---|---|
| Real-1 | 17.1 | 31.3 | 39.7 | 35.5 | 34.4 | 33.9 | 34.8 |
| Real-2 | 34.8 | 16.6 | 39.3 | 33.2 | 32.3 | 33.5 | 31.9 |
| Real-3 | 39.9 | 35.4 | 17.9 | 37.9 | 37.6 | 37.8 | 37.4 |
| Real-4 | 39.2 | 33.4 | 41.9 | 16.9 | 35.6 | 36.5 | 35.9 |
| Real-5 | 35.3 | 31.2 | 39.9 | 33.3 | 17.5 | 34.2 | 32.5 |
| Real-6 | 36.9 | 33.3 | 40.9 | 36.1 | 36.1 | 16.6 | 34.3 |
| Real-7 | 39.2 | 32.4 | 42.2 | 35.7 | 34.8 | 36.1 | 17.2 |

**Fig. 4.** Fréshet Inception Distance scores calculated for all pairs composed of seven real and seven generated image sets

Fig. 4. clearly demonstrates that each set of generated images is much closer to their respective parental training sets. Indeed, the elements of resultant matrix of FID scores situated on the leading diagonal are substantially smaller than the others. The variability within the elements of the leading diagonal themselves can be explained by the natural anatomical diversity of the real images from which the training datasets were composed.

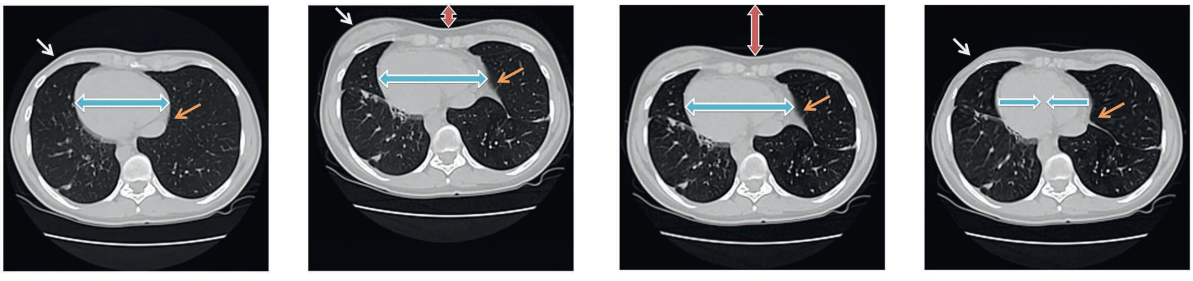
### Localizing and measuring the differences of real and generated images using nonlinear image deformations

The nonlinear image registration based on elastic image deformations is aimed at automatic establishing geometrical correspondences between the content of two images [10]. Typically, the nonlinear registration applied after the affine registration which put the image pair in a rough correspondence by applying linear transformations such as image shift, rotation, and scaling. Mathematically, the nonlinear registration is an ill-posed problem. Nevertheless, in practice applying the nonlinear image deformations allows to improve mutual image correspondence dramatically.

From computational point of view, the nonlinear registration is implemented as iterative optimization task whose cost function forces to perform deformations which ultimately resulted in the best image match. This particularly means that nonlinear registration is computationally-expensive. This is partly overcome by the parallelization using recent multi-core processors or GPU-based parallel implementation. In this work, both affine and nonlinear registrations were performed using RNiftyReg library of R language environment [8]. The cost function of the iterative optimization process includes suitable similarity metric to measure the quality of image match at every iteration step. Typically, the similarity measured by correlation of current and target images, or by the value of mutual information, or by an ad hoc synthetic function, or somehow else.

The mutual information (MI) is the most commonly used measure of pixel-wise similarity of two compared images that has solid theoretical basis [11]. Indeed, MI is a basic concept from information theory. It intimately linked to that of entropy of a random variable, a fundamental notion in information theory that quantifies the expected “amount of information” held in a random variable. In the context of image registration, it used to measure the amount of information that one image contains about the other [12]. The MI registration criterion postulates that MI is maximal when the images are correctly aligned. Thus, the value of MI at the final iteration step could be used as a good and robust metric of the similarity of real and generated images. The robustness here is due to the fact that the random vectors in this particular case are image pixels, the number of which is measured in hundreds of thousands or even millions.

Fig. 5 illustrates how the generated image transformed in two steps to its deformed version that match best the real target image. The key differences are highlighted by arrows. The values of MI score are given underneath.



*a* *b* *c* *d*

**Fig. 5.** Example of real (*a*) and the most similar (*b*) generated images along (MI-1 = 10.5) with the results of fitting generated image to the real one using linear affine (*c*) (MI-2 = 11.7) and nonlinear (*d*) (MI-3 = 12.5) transforms; the key differences are highlighted by arrows

### Results and discussions

In this paper, we have introduced a method for detection of image patterns presented in generated CT image slices that can be identified in the real images of lung tuberculosis patients. The method includes the following basic procedures: correlation of pairs of generated and real images for selection of pairs suggestive for the further analysis; computing correlation statistics using the direct and inverse Fisher’s transforms; performing affine image registration to put image pairs in rough corres-

pondence and calculating pair-wise similarity scores; non-linear (elastic) image registration, re-calculation of similarity scores and highlight the most similar/dissimilar image regions. Finally, we compute the FID distance of original and generated image datasets for assessing the overall quality of generation. Computational experiments were performed on 5 image training sets consisted of 6, 60, 600, 6000, and 60 000 CT image slices. As a result, it was found that images generated on small training sets (about 100 images and less) are nearly duplicates of the real images. The fraction of "distinct" artificial images grows with the increase of training set size. For instance, the inter-group FID distance is equal to 17.02, 14.45, and 10.71 for $N = 600$, $N = 6000$, and $N = 60\,000$ respectively.

## Conclusion

1. Based on several quantitative evaluations, it is found that images generated on small training sets (around 100 images or less) are almost duplicates of real patient images. The proportion of visually "distinguishable" synthetic images increases with the training set size. The Fréshet Inception Distance distance between groups decreases from 17.02 for $N = 600$ to 10.71 for $N = 60{,}000$, respectively.

2. Further research is needed to explore precise patterns for extracting and matching images that appear visually similar but cannot be compared using feature-based or pixel-based algorithms.

## References

1. Koshino K., Werner R. A., Pomper M. G., Bundschuh R. A. Toriumi F., Higuchi T., et al. (2021) Narrative Review of Generative Adversarial Networks in Medical and Molecular Imaging. *Annals of Translational Medicine*. 9 (9), 821–835.
2. Chambon P., Bluethgen C., Langlotz C. P., Chaudhari A. (2022) Adapting Pretrained Vision-Language Foundational Models to Medical Imaging Domains. *arXiv paper arXiv: 2210.04133*. 1–17.
3. Kozlovski S., Kovalev V. (2019) Generation of Artificial Biomedical Image Datasets for Training Deep Learning Models. *Pattern Recognition and Information Processing, 14th International Conference, Minsk, May 21–23, Belarusian State University of Informatics and Radioelectronics*. 278–281.
4. Song Y., Ermon S. (2019) Generative Modeling by Estimating Gradients of the Data Distribution. *33rd Conference on Neural Information Processing Systems, Vancouver, Canada, Dec. 8–14*. 115–119.
5. Ho J., Jain A., Abbeel P. (2020) Denoising Diffusion Probabilistic Models. *Advances in Neural Information Processing Systems*. 33, 6840–6851.
6. Carlini N., Hayes J., Nasr M., Jagielski M., Sehwag V., Tramer F., et al. (2023) Extracting Training Data from Diffusion Models. *Proceedings of the 32nd USENIX Security Symposium, Anaheim, CA, USA, Aug. 9–11*. 5253–5270.
7. Fisher R. A. (1915) The Frequency Distribution of the Values of the Correlation Coefficient in Samples of an Indefinitely Large Population. *Biometrika*. 10 (4), 507–521.
8. R Core Team (2022) *R: A Language and Environment for Statistical Computing.* Vienna, Foundation for Statistical Computing.
9. Heusel M., Ramsauer H., Unterthiner T., Nessler B., Hochreiter S. (2017) GANs Trained by a Two Time-Scale Update Rule Converge to a Local Nash Equilibrium. *NIPS'17: Proceedings of the 31st International Conference on Neural Information Processing Systems*. 6629–6640.
10. Ruthotto L., Modersitzki J. (2015) Non-Linear Image Registration. *Handbook of Mathematical Methods in Imaging*. Springer, New York.
11. Cover T. M., Thomas J. A. (1991) *Elements of Information Theory.* New York, John Wiley & Sons.
12. Maes F., Loeckx D., Vandermeulen D., Suetens P. (2015) Image Registration Using Mutual Information. *Handbook of Biomedical Imaging*. Springer. 295–308.

**Information about the author**

**Kovalev V. A.,** Cand. of Sci., Leading Researcher,
The United Institute of Informatics Problems
of the National Academy of Sciences of Belarus

**Address for correspondence**

220013, Belarus, Minsk, Surganova St., 6
The United Institute of Informatics Problems
of the National Academy of Sciences of Belarus
Tel.: +375 29 199-97-70
E-mail: vassili.kovalev@gmail.com
Kovalev Vassili Alekseevich