Biomedical image quality improvements with Attention Mechanisms and Deep Residual Learning

Serhii Andikevych[a], Artem Shcherbatyuk[a], Yurii Poudanien[a],
Stanislav Tuzhanskyi*[a], Andrii Kozhemiako[a]
[a]Vinnytsia National Technical University, 95 Khmelnytske Shose, Vinnytsia, Ukraine

ABSTRACT

Image segmentation plays a key role in biomedical imaging, allowing different structures or regions of interest, such as organs, tissues, or blood vessels, to be clearly identified for further analysis and diagnosis. Segmentation of retinal vessels in fundus images is particularly challenging due to low contrast and complex vascular structure, which complicates the diagnosis of ophthalmic diseases. In this study, we present a state-of-the-art deep learning approach to improve biomedical image quality and segmentation accuracy using Attention U-Net++ combined with the ResNet support network. Attention mechanisms enhance the model's ability to focus on fine details of blood vessels, while deep residual learning ensures the stability of deep architecture learning. The method is evaluated on a fundus image dataset, achieving a Dice score of 0.8694 and an Intersection over Union (IoU) score of 0.7690, demonstrating competitiveness with state-of-the-art methods. The results of the visual analysis emphasize the model's ability to accurately delineate vessels and indicate areas for improvement, namely the development of methods for finding small vessels. The results demonstrate the potential of attention and residual learning mechanisms to improve biomedical image analysis, offering a reliable tool for clinical applications in the diagnosis of retinal diseases.

Keywords: image segmentation, biomedical images, image quality improvement, attention mechanisms, deep residual learning, fundus images, deep learning, Attention U-Net++, ResNet, medical imaging, retinal diagnostics

1. INTRODUCTION

Biomedical imaging is one of the most important areas of modern medicine, providing non-invasive analysis of the anatomical and functional features of the human body. It plays a key role in diagnosing, planning treatment, and monitoring the progress of various diseases[1]. One of the central tasks of biomedical imaging is image segmentation, which involves the selection of specific structures or areas of interest, such as organs, tissues, or blood vessels, for further analysis[2]. Image segmentation allows not only to visualize complex biological structures but also to quantify their characteristics, which is crucial for accurate diagnosis and prediction of pathology development[3]. In this context, segmentation of retinal vessels in fundus images is of particular importance, as retinal vessels are an important marker for the early detection of ophthalmic and systemic diseases such as diabetic retinopathy, glaucoma, hypertensive retinopathy, and even cardiovascular disorders.

Fundus images obtained with fundus cameras are a relatively affordable and widely used diagnostic method in ophthalmology. They allow visualization of the retina, optic disc, macula, and blood vessels that supply the eye[4]. However, segmentation of vessels in such images is challenging due to a number of factors. First, the low contrast between the vessels and the background makes it difficult to distinguish them clearly, especially for small capillaries. Secondly, the presence of noise caused by technical limitations of the equipment or characteristics of the eye tissue can mask important details. Third, retinal vessels have a complex topology, including branches, intersections, and variable thickness, which requires high accuracy from segmentation methods[5]. Traditional approaches, such as thresholding, gradient-based methods, or manual segmentation, are often ineffective due to the high variability of images and require significant computing resources or the involvement of expert physicians.

Recent decades have seen significant progress in machine learning, particularly deep learning, which has opened up new perspectives for automating biomedical image processing[1,6]. In particular, deep neural networks, such as U-Net, have shown high performance in segmentation tasks due to their ability to learn complex spatial dependencies in data[7]. However, the classical U-Net architecture has limitations when it comes to processing images with low contrast or small

*e-mail: slavat@vntu.edu.ua; http://vntu.edu.ua/

Photonics Applications in Astronomy, Communications, Industry, and High Energy Physics Experiments 2025, edited by Andrzej Smolarz, Ryszard S. Romaniuk, Waldemar Wójcik, Sergii V. Pavlov, Proc. of SPIE Vol. 14009
1400903 · © 2025 SPIE · 0277-786X · doi: 10.1117/12.3093542

structures such as retinal capillaries. To overcome these limitations, advanced models have been developed, such as Attention U-Net, which integrates attention mechanisms to focus on relevant areas of the image[8], and U-Net++ with improved connections between the encoder and decoder layers[9]. In addition, the use of residual networks (ResNet) as the basis for deep models can stabilize training and improve performance by reducing the effect of gradient vanishing[10].

In this context, attention mechanisms and deep residual learning become particularly important. Attention mechanisms allow the model to identify more important structures, such as blood vessels, while ignoring noise or irrelevant background, which is especially useful in low contrast environments. On the other hand, residual learning, implemented through the ResNet architecture, ensures the stability and efficiency of training deep models, allowing for the processing of high-resolution images with complex patterns. The combination of these two approaches - attention mechanisms and residual learning - creates a powerful tool for solving biomedical image segmentation problems, particularly in ophthalmology[11].

This study proposes an innovative approach to retinal vessel segmentation based on the Attention U-Net++ architecture with ResNet as a foundation. The proposed model integrates attention mechanisms for the accurate selection of small vessels and deep residual learning for the stability of the deep network. Particular attention is paid to optimizing the model for working with high-resolution images (704×1056), which allows preserving the details of small structures. Experiments were conducted on the HRF12 fundus image dataset, where the model demonstrated high segmentation accuracy, achieving a Dice coefficient of 0.8694 and an Intersection over Union (IoU) score of 0.7690. To evaluate the results, detailed visualizations such as overlaid masks and confusion matrices were used to analyze both the model's strengths and limitations, such as the omission of small vessels. The significance of this study lies in the creation of an effective tool for automating the diagnosis of ophthalmic diseases that can be integrated into clinical practice. Improving the quality of biomedical images and the accuracy of retinal vessel segmentation opens up new opportunities for early detection of pathologies, which is critical for timely intervention and improving the quality of life of patients. In addition, the proposed approach has the potential to be scaled up to other biomedical imaging tasks, such as tumor or nerve fiber segmentation, making it a versatile solution for medical research.

2. NETWORK ALGORITHM AND ARCHITECTURE

In this study, we propose an approach to retinal vessel segmentation based on the Attention U-Net++ architecture with ResNet as its foundation. Attention U-Net++ is an advanced version of the classic U-Net model that includes attention mechanisms and improved connections between encoder and decoder layers to improve the accuracy of small structure segmentation. Attention Gates mechanisms allow the model to focus on relevant image regions, such as blood vessels, ignoring noise and irrelevant background, which is especially important for low-contrast images typical of the fundus[11,13,14].

The model's encoder is based on the ResNet residual network, which ensures the stability of training deep architectures due to residual connections. The use of ResNet allows efficient processing of high-resolution images and avoids the problem of gradient disappearance that often occurs when training deep networks. The proposed model uses ResNet-34 as a basis that provides a balance between computational complexity and performance[15]. To adapt to the limited computing resources, the number of filters in the model layers was reduced to 8, 16, 32, 64, and 128, which allowed us to optimize training on a personal computer without significantly degrading the quality of segmentation. A PC based on Intel core i5-8400 CPU, 2.8 GHz, 16 GB of RAM was used to train the model.

The Attention U-Net++ architecture consists of an encoder, a decoder, and additional nested skip connections that transfer information between different levels of the model. In each decoder block, Attention Gates are used to calculate weighting coefficients for the input features, allowing the model to pay more attention to the vessels and suppress the influence of background pixels. The output of the model is a binary mask, where the value of 1 corresponds to vessel pixels and 0 to the background. The architecture of the developed model is shown in Figure 1.

The model is trained using a combined loss function that combines Binary Cross-Entropy (BCE) and Dice Loss. BCE provides sensitivity to local differences between the predicted and true masks, while Dice Loss optimizes global similarity, which is important for segmentation tasks with unbalanced classes where vessel pixels occupy a much smaller area compared to the background. The combined loss function is defined as[16]:

$$L = L_{BCE} + (1 - Dice\ Coefficient),$$

where L_{BCE} is the binary cross-entropy, and the Dice Coefficient is calculated as:

$$Dice\ Coefficient\ = \frac{2 * \sum(y_{true} * y_{pred})\ + \varepsilon}{\sum y_{true}\ + \sum y_{pred} + \varepsilon},$$

where y_{true} is the ground truth mask, y_{pred} is the predicted mask, and $\varepsilon\ = 10^{-6}$ is the small constant to avoid division by zero.

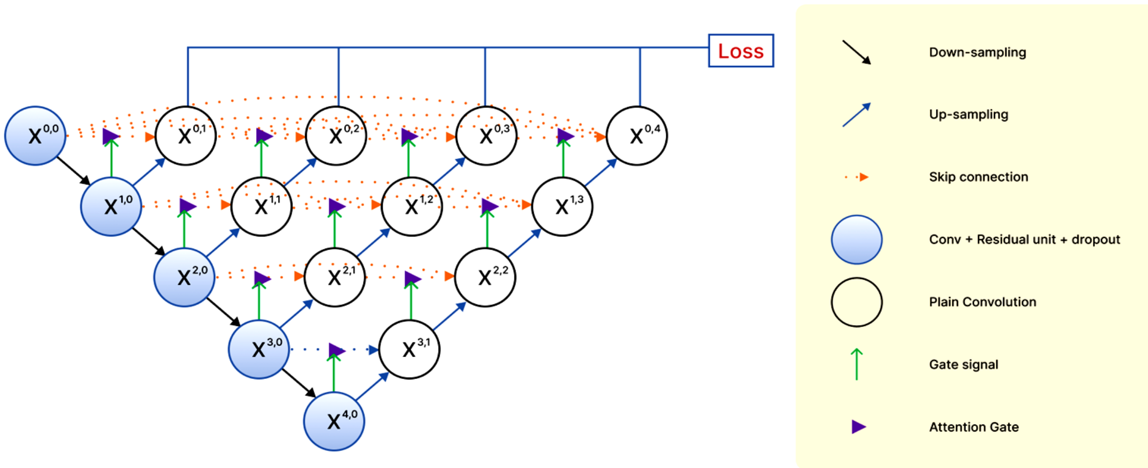


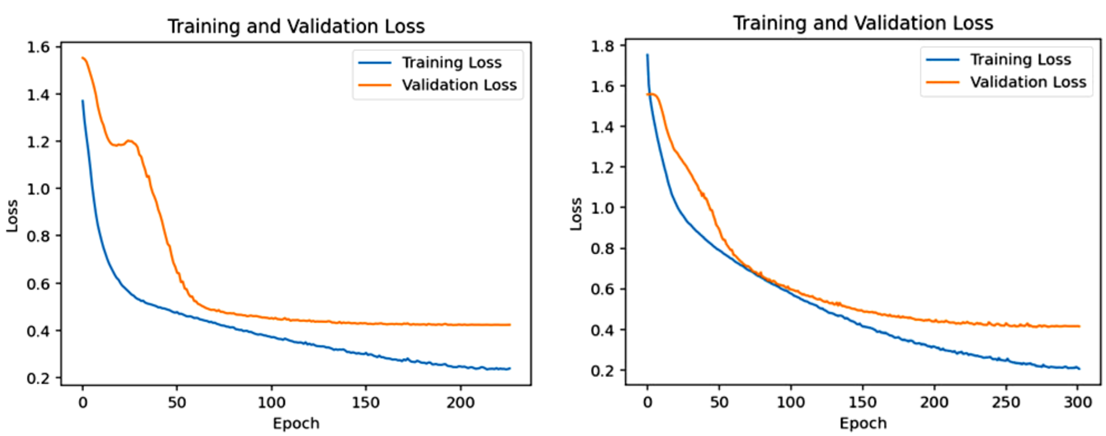
Figure 1. Architecture of the Developed Model

For evaluating the model's performance, two main metrics were used: Dice Coefficient and Intersection over Union (IoU). The formula for IoU is as follows:

$$IoU = \frac{\sum(y_{true} * y_{pred})\ + \ \varepsilon}{\sum y_{true} + \sum y_{pred} - \sum((y_{true} * y_{pred})\ + \ \varepsilon},$$

where $\varepsilon\ =\ 10^{-6}$ is added for computational stability (to avoid division by zero)

Training was performed for 300 epochs using the Adam optimizer and an initial learning rate of 0.001. To prevent overfitting, an early stopping mechanism was used, which stopped training if the loss value on the validation set did not decrease for 30 epochs. In addition, a learning rate scheduling strategy was applied: the learning rate was halved every time the loss value on the validation set did not improve for 10 epochs. This allowed the model to adapt to complex data and achieve better convergence (Fig. 2b). The final model was saved in Keras format after training[17].



a) b)

Figure 2. Dynamics of Model Training Metrics Over 300 Epochs: a) Model Training Losses without ResNet; b) Model Training Losses with ResNet

3. DATASET

Experiments with medical images were conducted on the publicly available HRF (High-Resolution Fundus) dataset, which is designed to evaluate retinal vessel segmentation methods. The HRF dataset contains 45 high-quality fundus images divided into three categories: 15 images of healthy eyes, 15 images of diabetic retinopathy, and 15 images of glaucoma. Each image has a resolution of 3504×2336 pixels and is accompanied by an expert binary mask created by specialists, where vessels are marked in white (1) and the background is black (0)12.

For processing, the images were resized to 704×1056 pixels with preserved proportions, which allowed the model to work with detailed vessel structures while maintaining acceptable computational complexity. Image pixel values were normalized to the range [0, 1] by dividing by 255. The masks were binarized: pixels with a value above 0.5 were considered vessels (1), and below - background (0). The dataset was divided into training (90%) and validation (10%) sets to evaluate the model during training.

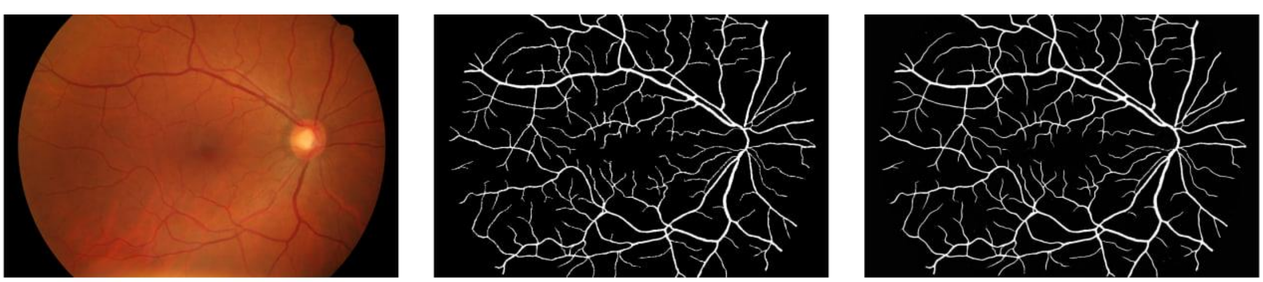
4. RESULTS OF THE DEVELOPED MODEL

The dynamics of the model training process are shown in Fig. 2, which illustrates the change in the value of the combined loss function for two architecture variants: with ResNet (Fig. 2 b) and without ResNet (Fig. 2 a). Figure 2 a) shows how the training and validation losses gradually decrease: from an initial value of approximately 1.8 at the start of training to a level of 0.2-0.3 after 150 epochs for the model without ResNet. Similarly, Figure 2(b) shows a similar trend for the model with ResNet, where the losses stabilize at 0.2-0.3 after 150 epochs. This decline indicates an effective optimization of the combined loss function (BCE + Dice Loss) in both cases. The stabilization of the metrics after 150 epochs in both architectures confirms the effectiveness of the applied regularization strategies, in particular, early stopping and reducing the learning rate. Thanks to this, the models achieved optimal performance in the early stages while maintaining stability and accuracy throughout the process.

Table 3 shows the performance of the models (U-Net and SegNet) from previous studies[18–22] and the performance of the proposed Attention U-Net model with and without ResNet. The results showed that Attention U-Net achieved a Dice of 0.8334 and an IoU of 0.7144, which indicates a significant improvement due to the attention mechanisms. Attention U-Net performed best with ResNet, with a Dice of 0.8694 and an IoU of 0.7690, which is 54.03% higher than U-Net and 22.57% higher than SegNet in terms of Dice. These data emphasize the advantage of integrating ResNet and attention mechanisms to improve the accuracy of retinal vessel segmentation.

Table1. Results of Experimental Measurements of Model Performance

Model	Dice	IOU
U-Net	0.5644	0.3931
SegNet	0.7093	0.5496
Attention U-Net	0.8334	0.7144
Attention U-Net with ResNet	0.8694	0.7690



a) b) c)

Figure 3. Test Image Processing Result: a) Typical fundus view; b) Real mask; c) Predicted mask

The result of the test image processing is shown in Figure 3. The original image (a) shows a typical fundus view, while the real mask (b) and the predicted mask (c) show the segmentation of the vessels. Figure 4 shows an overlay of the original and predicted masks, and an overlay of this image on the original to facilitate visual analysis, with green coloring illustrating areas where the model failed to detect vessels.

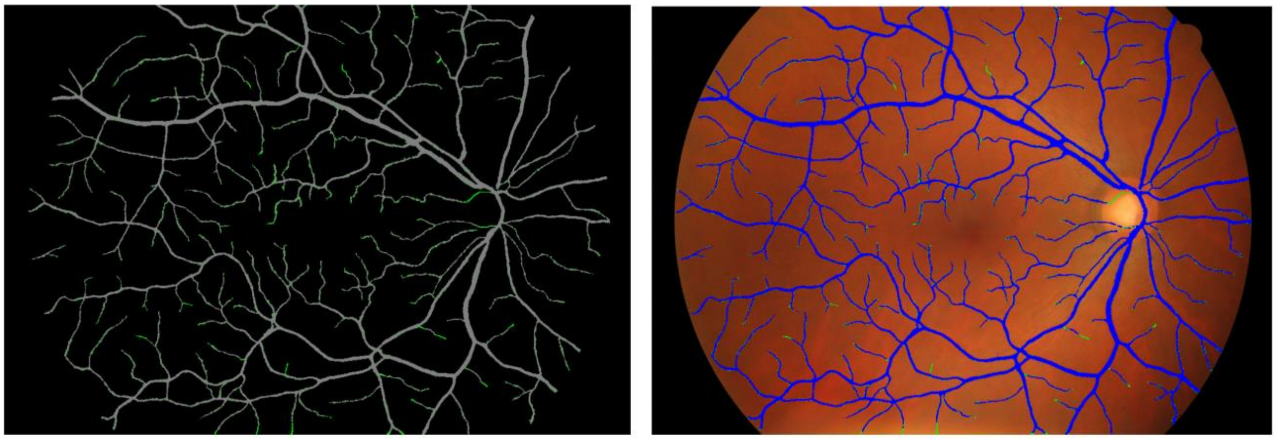


Figure 4. Overlaid images

The visual analysis shows that the model effectively identifies the main vessels, as can be seen from the overlay of predicted and real vessels. However, the analysis of false negative pixels shows that the model failed to detect some small vessels and capillaries, especially in areas with low contrast or complex vascular network structure. This indicates the potential for further improvement of the model. One possible way to improve is to increase the size of the training dataset, which would allow the model to better generalize across a variety of images. Additionally, applying data augmentation, such as varying contrast, brightness, or adding noise, can help the model better recognize small vessels. It is worth noting that the real mask created by experts may not always be perfect. Experts may have labeled some vessels based on their own experience or assumptions, especially in areas where vessels are not clearly visible due to poor image quality. This may partially explain the discrepancies between the predicted and true masks, and future studies should take into account the possible subjectivity of expert labeling when evaluating the model.

In general, the results demonstrate the high efficiency of the proposed model, but the identified shortcomings indicate opportunities for its further improvement, which can enhance the accuracy of retinal vessel segmentation.

5. CONCLUSIONS

The proposed Attention U-Net++ model with the ResNet framework demonstrates significant potential for retinal vessel segmentation in fundus images, achieving a Dice coefficient of 0.8694 and an IoU of 0.7690 on the test image. These results exceed those of previous studies: U-Net (Dice 0.5644) by about 54.03% and SegNet (Dice 0.7093) by about 22.57%, which emphasizes the effectiveness of integrating attention mechanisms and network architecture with residual connections. Visual analysis demonstrates accurate segmentation of major vessels, although detection of small capillaries remains problematic, indicating room for improvement through dataset expansion, augmentation, and algorithm optimization. Further research will be aimed at improving the model's performance to overcome these limitations, contributing to more accurate non-invasive diagnostics in ophthalmology.

REFERENCES

[1] Romanyuk, O., Smolarz, A., et al., "Method of anti-aliasing with the use of the new pixel model," Proc. SPIE 9816, 981617 (2015).

[2] Komenchuk, O.V., Mokin, O.B. "Analysis of Methods for Preprocessing of Panoramic Dental X-Rays for Image Segmentation Tasks," Visnyk of Vinnytsia Politechnical Institute, (5), 41–49 (2023).

[3] Romanyuk, A.N., et al., "Fast ray casting of function-based surfaces," Przeglad Elektrotechniczny, 93(5), pp. 83–86 (2017).

[4] Pavlov, S.V., Vassilenko, V.B., Saldan, I.R., Vovkotrub, D.V., Poplavskaya A.A., and Kuzin O.O., "Methods of processing biomedical image of retinal macular region of the eye," Proc. SPIE 9961, 99610X (2016).
[5] Kukharchuk, V.V., Holodiuk, V.V., et al., "Information conversion in measuring channels with optoelectronic sensors," Sensors, 22(1), 271 (2022).
[6] Komenchuk, O. V., and Mokin, O. B., "Information technology for accelerated annotation of medical images in segmentation tasks based on deep learning models," Visnyk of VPI [Internet]. (4): 95-103 (2024).
[7] Ronneberger, O., Fischer, P., and Brox, T., "U-Net: Convolutional Networks for Biomedical Image Segmentation," Version 1. arXiv. (2015).https://doi.org/10.48550/ARXIV.1505.04597
[8] Oktay, O., Schlemper, J., Folgoc, L. L., Lee, M., Heinrich, M., Misawa, K., et al., "Attention u-net: Learning where to look for the pancreas," arXiv preprint arXiv:1804.03999. (2018).
[9] Micallef, N., Seychell, D., & Bajada, C. J., "Exploring the u-net++ model for automatic brain tumor segmentation," IEEE Access, 9, 125523-125539 (2021).
[10] He, K., Zhang, X., Ren, S., & Sun, J., "Deep Residual Learning for Image Recognition," (Version 1). arXiv. https://doi.org/10.48550/ARXIV.1512.03385 (2015).
[11] Li, Z.; Zhang, H.; Li, Z.; Ren, Z., "Residual-Attention UNet++: A Nested Residual-Attention U-Net for Medical Image Segmentation," Appl. Sci. 12, 7149. (2022).
[12] Dataset Ninja, Visualization Tools for High Resolution Fundus Dataset. Dataset Ninja; 2025. Accessed February 10, https://datasetninja.com/high-resolution-fundus (2025).
[13] Fraz, M. M., Remagnino, P., Hoppe, A., Uyyanonvara, B., Rudnicka, A. R., Owen, C. G., & Barman, S. A., "An ensemble classification-based approach applied to retinal blood vessel segmentation," IEEE Transactions 1 on Biomedical Engineering, 59(9), 2538-2548. (2012).
[14] Ronneberger, O., Fischer, P., & Brox, T., "U-Net: Convolutional networks for biomedical image segmentation," In International Conference on Medical image computing and computer-assisted intervention, Springer, Cham., 234-241 (2015).
[15] Almoosawi, N. M., & Khudeyer, R. S., "ResNet-34/DR: a residual convolutional neural network for the diagnosis of diabetic retinopathy," Informatica, 45(7) (2021).
[16] Su, J., Liu, Z., Zhang, J., Sheng, V. S., Song, Y., Zhu, Y., & Liu, Y., "DV-Net: Accurate liver vessel segmentation via dense connection model with D-BCE loss function," Knowledge-Based Systems, 232, 107471 (2021).
[17] Tensor Flow Developers. (n.d.). Keras: The Python deep learning API. TensorFlow. Retrieved February 11, from https://www.tensorflow.org/guide/keras (2025).
[18] Dzierżak, R., and Michalska, M., "Analysis of the effectiveness of selected segmentation methods of anatomical brain structures," Informatyka, Automatyka, Pomiary w Gospodarce i Ochronie Środowiska, 8(2), 58–61, (2018). https://doi.org/10.5604/01.3001.0012.0713
[19] Andrikevych, S., & Tuzhansky, S., "Retinal vessel segmentation on fundus images using SegNet", VNTKP VNTU. Faculty of Information Electronic Systems. Available at: https://conferences.vntu.edu.ua/index.php/all-frtzp/all-frtzp-2025/paper/view/23933 (2025).
[20] Andrikevych, S., & Tuzhanskyi, S. "Application of U-Net for retinal vessel segmentation in fundus images", In VNTKP VNTU. International Scientific and Technical Conference on Optoelectronic Information Technologies "Photonics − ODS 2025", (2025). Retrieved from https://conferences.vntu.edu.ua/index.php/ods/ods_2025/paper/view/23294
[21] Wójcik, W., Pavlov, S., Kalimoldayev, M., "Information Technology in Medical Diagnostics II," London: Taylor & Francis Group, CRC Press, Balkema book, p. 336 (2019).
[22] Dubolazov, O.V., Ushenko, A.G., Ushenko, Y.A., Sakhnovskiy, M.Y., et al., "Laser müller matrix diagnostics of changes in the optical anisotropy of biological tissues", Information Technology in Medical Diagnostics II, 195-203 (2019).