

Automatic Selection of the Optimal Zone for Laser Exposure According to the Fundus Images for Laser Coagulation

Nikita S. Demin^{1,2*}, Nataly Yu. Ilyasova^{1,2}, and Rustam A. Paringer^{1,2}

¹IPSI RAS – Branch of the FSRC "Crystallography and Photonics" RAS, 151 Molodogvardeyskaya str., Samara 443001, Russia

²Samara National Research University, 34 Moskovskoye Shosse, Samara 443086, Russia

*e-mail: volfgunus@gmail.com

Abstract. We analyzed a problem of extracting regions of interest (RoI) in eye-fundus images in the laser treatment for diabetic retinopathy (DR) using machine learning of deep neural networks aimed at higher-accuracy recognition of pathological and anatomical structures in the macula edema region. In this paper, we propose a method for automatic selection of the optimal zone of laser exposure based on images of the fundus for laser coagulation. Two neural networks were used to solve the problem. The first singled out anatomical objects in the fundus, and the second edema zone. The result was formed from the edema area, taking into account the location of anatomical objects on it. © 2023 Journal of Biomedical Photonics & Engineering.

Keywords: full convolutional neural networks; fundus imaging; macular edema; laser coagulation.

Paper #8988 received 13 Jun 2023; revised manuscript received 23 Oct 2023; accepted for publication 27 Oct 2023; published online 8 Dec 2023. [doi: 10.18287/JBPE23.09.040308](https://doi.org/10.18287/JBPE23.09.040308).

1 Introduction

Diabetes mellitus is a common endocrine disease affecting all human organs. On the part of the visual apparatus, this manifests itself in the form of diabetic retinopathy (DR) [1–5]. Today, there are almost 400 million patients with diabetes in the world, and by 2035 the number is expected to increase to 592 million people [6]. The most dangerous manifestation of DR is macular edema. The walls of the retinal vessels become thinner and hemorrhages occur in the retinal area, leading to partial or complete loss of vision. According to the Wisconsin Epidemiological Study of Diabetic Retinopathy (WESDR), with a duration of diabetes for more than 20 years, retinopathy is detected in 80–100% of cases, while diabetic macular edema (DME) develops in 29% of cases [7–8].

One of the methods used in medical practice for the treatment of DR is laser photocoagulation, during which certain areas of the retina are cauterized with a laser, stopping the spread of macular edema. There are no common systems for the treatment of diabetic retinopathy in the world that would provide a coagulation

plan with sufficient treatment efficiency, which is why many experienced doctors prefer manual laser guidance [9–10]. NAVILAS is a device of the German company ODOS [11, 12], which allows you to perform laser coagulation according to a pre-formed plan. The plan is formed manually, placing a hexagonal pattern in the area of influence. The limited patterns do not suit physicians [11, 13, 14], and they return to the use of old equipment with manual laser guidance, such as Valon [13].

To form a plan, it is necessary to determine which zone can be fired upon, as well as which zone is optimal, in the sense that by eliminating areas that are inappropriate for exposure, the number of laser shots and, accordingly, the costs of the substance used are reduced, and the impact on the fundus is also reduced, which should improve the quality of treatment.

In the modern world, neural network algorithms are used to solve data mining problems. With the development of technology, convolutional neural networks have found wide application in the tasks of image and video data processing. In particular, in biomedicine, neural networks allow solving problems of

semantic segmentation, search and recognition of objects in an image [15, 16], for example, determining the area of lung damage by the SARS-CoV-2 virus [17] or cancerous tumors of the human brain [18].

In this paper, we propose a method for the formation of the optimal zone of laser exposure based on the selection of the zone of macular edema and anatomical objects in the images of the fundus. This method reduces the traumatic nature of laser photocoagulation surgery.

2 Materials and Methods

2.1 Selection of the Optimal Zone for Laser Exposure

Two neural networks were used to form the laser impact zone. The first network allows us to select anatomical objects in the images of the fundus. The second network highlights the edema zone from the fundus images.

Algorithm for selecting the optimal zone of laser exposure:

1. Using the first neural network, anatomical objects are highlighted in the image.
2. Using the second neural network, the edema zone is highlighted in the image.
3. The resulting neural network segmentation maps are a probability distribution requiring post-processing. At this stage, the map is also binarized.

The edema zone is corrected taking into account the anatomical landmarks located on it (the masks of anatomical landmarks are subtracted from the mask of the edema zone), thus a preliminary mask of the laser exposure zone is obtained.

4. The resulting mask of the optimal zone of laser exposure is obtained from the preliminary mask by automatic processing it using the morphological functions of erosion and dilation to clarify the boundaries of macular edema with kernel size 7×7 . As well as processing with a median filter to smooth out the unevenness of the mask with kernel size 7×7 .

All neural network architectures used are fully convolutional neural networks. This type of neural networks is best suited for solving semantic segmentation problems because: the entire image context is taken into account, segmentation is performed in one pass, they are well known and widely popular.

2.2 Augmentation

The popularity of using full convolutional neural networks in semantic segmentation problems is determined by a number of factors. Thus, neural networks have a good generalizing ability and are able to take into account the entire context of the image, which allows them to be used in cases where other methods show worse results. However, neural networks also have limitations that directly depend on the characteristics of the training data set. In this regard, the use of neural networks in biomedicine faces a number of specific problems.

1. Due to patient privacy policies and the complexity of data labeling, as well as the requirements for the qualification of a specialist who performs labeling, data sets are often small [19]. The use of augmentation methods can partially compensate for the insufficient size of data for training [20]. In particular, for example, in tasks related to the analysis of biomedical images, the method of elastic data augmentation [21] is popular and effective: rotation at an arbitrary angle, reflections, and elastic deformation.

2. A frequent problem in tasks related to medicine, a pronounced problem is the imbalance of classes [22]. Solving the problem of class imbalance in the training dataset is non-trivial, but there are methods to smooth out the impact of this problem [23, 24]. For example, in the case of fundus image processing, samples often contain classes that are rarely found on most images, and, accordingly, the relative volume of these classes is extremely small. This problem of class imbalance should be taken into account when designing neural networks. Various kinds of algorithms should be applied that can level the influence of this problem [25, 26].

3. The most difficult problem is the low quality of data labeling, which is the result of the low qualification of the specialist who did the labeling [27]. This problem is extremely difficult to identify and eliminate at the training stage.

The peculiarities of our problem are the fact that the original data are unbalanced, the number of images is small, and the labelling does not exactly match the actual location of the objects.

There are a number of works devoted to solving this problem, however, they are mostly highly specialized and consider only the segmentation of images into one class, for example, the class of blood vessels [28] or exudates [29]. In paper, the problem of segmentation of fundus images into several classes is considered, which is relevant for creating a decision support technology for a doctor in the diagnosis and treatment of diabetic macular edema [30].

In accordance with the problems mentioned earlier, an important step in the application of neural networks is data preparation: smoothing out the influence of a small amount of training sample and class imbalance. In this work, the following image augmentation techniques were used: reflection, rotation by a random angle (from -30° to 30°), random shift, and elastic transformation. The use of augmentation allows us to successfully train a neural network to perform semantic segmentation, as well as to deal with network retraining when using small data sets.

2.3 Segmentation Anatomical Objects

For step 1 of the algorithm, the neural network described in Ref. [31] was used.

Based on the Unet network, the ResNetUnet, DenseNetUnet and XceptionUnet networks were built, where one of the ResNet, DenseNet, or Xception networks, respectively, was used as an encoder. The weights of the pretrained networks were used to initialize the encoder and were fixed for the duration of the training.

The ResNetUnet architecture uses the ResNet-101 network [32] pretrained on the ImageNet data set [33] as a feature encoder. The features of the ResNet architecture include its depth. It was the first architecture that allowed researchers to train neural networks with more than 20 layers. Also in this architecture, skip-connection (a modification of the architecture that allows the signal to pass through the network by skipping one or more layers) was used for the first time to prevent the fading gradient problem.

The DenseNetUnet architecture uses the DenseNet-169 network [34] as a feature encoder, which was pretrained on the ImageNet dataset. The DenseNet architecture learns well on small datasets and proposes to connect all layers using skip-connection within one building block of the network.

XceptionUnet uses the Xception-65 network [35] as a feature encoder, which was pretrained on the ImageNet dataset. The features of this architecture of the neural network include the use of a combination of pointwise (point convolution with a kernel size of 1×1) and depthwise (spatial convolution independently applied to each channel) convolutions instead of classical convolution. This replacement allows to reduce the number of trained parameters without affecting the accuracy of the network.

For the experiment, a dataset was used, consisting of 115 fundus images, which were labeled into 8 classes: optical disc (OD), macula (M), blood vessels (BV), hard exudates (HE), soft exudates (SE), new coagulates (NC), pigmented coagulates (PC), and hemorrhage (H). Images have dimensions of $1024 \text{ px} \times 1024 \text{ px}$ and 3 rgb color channels.

Neural networks are built and trained using the TensorFlow library. The following parameters were used in training:

- Input size: $1024 \text{ px} \times 1024 \text{ px} \times 3$.
- Number of epochs: 120.
- Loss function: FocalLoss [36].
- Optimizer: Adam [37].
- Learning rate: 0.003.

The specified number of epochs is the maximum value of possible training epochs. Each neural network had its own optimal epoch, at which training stopped if the results did not improve within 20 epochs. During the research, it was found that the XceptionUnet network shows the highest accuracy among those studied.

2.4 Segmentation Edema Zone

To search for a suitable neural network architecture for the second point of the laser exposure area selection algorithm, in this work, such neural networks as Unet, Unet++, Manet, Linknet, Feature Pyramid Network (FPN), Pyramid Scene Parsing Network (PSPNet), Pyramid Attention Network (PAN), and Deeplabv3 were used.

All neural networks used in this work are based on the resnet-34 network pre-trained on images from the ImageNet dataset. Using a pre-trained neural network allows us to train your own networks faster, because

such networks are already capable of extracting a huge number of features from images.

Unet [38] is one of the first fully convolutional neural network architectures successfully applied to solve the problem of semantic segmentation of biomedical images. The Unet architecture can be represented schematically in the form of the letter U. This architecture consists of two parts, an encoder that acts as a classifier or feature extractor and a decoder that maps the image segmentation from the features extracted by the encoder.

Unet++ is an improved version of the Unet architecture, which was also developed for use in tasks of semantic segmentation of biomedical images [39]. The main change in this architecture is the replacement of the links between the encoder and decoder with small intermediate networks. Unlike simply concatenating the outputs of the encoder layers to the outputs of the decoder layers, the use of such intermediate networks improves the internal representation of the features of the neural network in such a way that it allows the decoder to more accurately build a segmentation map.

The architecture of the MA-net neural network was developed to solve the problem of semantic segmentation of images obtained as a result of scanning the organs of the liver and brain [40]. A feature of this neural network is the use of the attention mechanism. This mechanism models the human ability to focus on an object or area of interest. Thus, this mechanism allows the neural network to pay attention to certain areas, features, channels, depending on the area under consideration. Accounting for such information allows us to train more accurate neural networks. According to the study, the addition of the attention mechanism improves the accuracy of semantic segmentation in comparison with the neural network of the Unet and Unet++ architectures.

The Linknet architecture is similar to the Unet architecture with the exception of some changes [41]. So, in the original Unet, the decoder and encoder are interconnected by simple links. These links extend the outputs of the encoder with the outputs of the decoder. In the Linknet architecture, it is proposed to use the residual connection proposed by the authors of the Resnet architecture. Thus, the outputs of the decoder are added to the outputs of the encoder. According to the study conducted by the authors, this modification improves the accuracy of the final model.

The FPN architecture is also similar to the Unet architecture [42]. Its feature is how the outputs of each decoder block are used. In Unet, the final segmentation map is the output from the last layer of the neural network. Whereas in FPN the final segmentation map is made by weighted summation of the output of each decoder block. This feature to a greater extent allows us to better take into account the general context of the image at different scales, as well as deal with overfitting of the neural network.

The architecture of PSPNet is fundamentally different from those considered earlier [43]. The typical U-shaped architecture does not apply here. So PSPNet consists of a convolutional neural network (CNN) which also acts as

a feature encoder. Further, the features from the last layer of the CNN fall into a special module, in which convolutions of different sizes are applied to the features. The convolution data performs feature localization. Next, each convolution output is upsampled to the size of the original image using an upsampling layer. After that, all layers are concatenated together with the encoder feature map. For this tensor, the final convolutional layer is applied, which forms a segmentation map.

The PAN architecture [44] similarly to MAnet uses the attention mechanism to improve the accuracy of the neural network. While I use Position-wise Attention Block (PAB) and Multi-scale Fusion Attention Block (MFAB) in MAnet, Feature Pyramid Attention (FPA) is used in PAN architecture. This block is designed to improve the perception of the attention block itself by building a structure inside the FPA similar to that implemented in the FPN network architecture. Thus, the FPA block allows you to highlight features from different scales, thereby making up for the lack of classical mechanisms for attention to pixel information.

The Deeplabv3 [45] architecture also consists of an encoder and a decoder. However, unlike all previous architectures, the main idea of this architecture is to replace conventional convolutional layers with sparse convolution layers (atrous convolution). These layers are designed to help train the neural network to better understand the context of the image.

For the experiments, we used a set of 50 fundus images (with dimensions of $1024 \text{ px} \times 1024 \text{ px}$ and 3 rgb color channels) labeled by an ophthalmologist. This dataset differs from the dataset used in the segmentation of anatomical objects. Fig. 1 shows an example of an image of the fundus and Fig. 2 shows its markings.

So, using augmentation methods, it allowed us to expand our data set to 5000+ images. The initial data set was divided into training and test sets in a ratio of 4 to 1. Thus, 40 fundus images represented the training set and 10 images represented the test set.

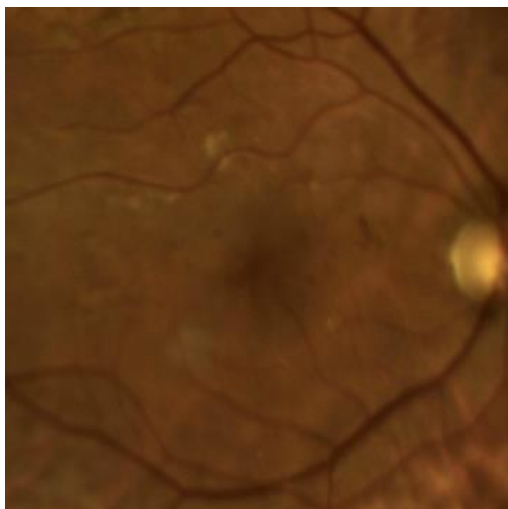


Fig. 1 Source image ($1024 \text{ px} \times 1024 \text{ px}$).

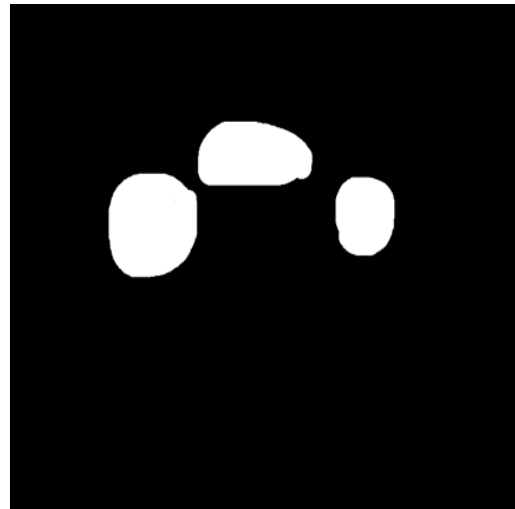


Fig. 2 Binary segmentation mask for macular edema ($1024 \text{ px} \times 1024 \text{ px}$).

All neural networks were trained with the same parameters:

- Input size: $1024 \text{ px} \times 1024 \text{ px} \times 3$.
- Number of epochs: 150.
- Loss function: Cross Entropy.
- Optimizer: Adam.
- Learning rate: 0.001.

3 Result and Discussion

During the experiments, four neural networks were selected that have the highest accuracy according to the f1 metric. Table 1 presents the results of this experiment.

Table 1 Training results for various neural network architectures.

Architecture	f1
Unet	0.584
Unet++	0.562
MAnet	0.508
Linknet	0.575
FPN	0.438
PSPNet	0.399
PAN	0.512
DeepLabV3	0.478

The following networks became the best by the metric: Unet, Unet++, Linknet, PAN. Further research was carried out with them. In the course of the research, it was found that if, in addition to the image of the fundus, a map of objects (optical disk, macula, vessels, etc.) is submitted to the input of the neural network, this will improve the accuracy of the model when highlighting the edema area. The objects were selected using a neural network from the first step of the algorithm for selecting

the area of the laser impact zone. The results of the corresponding experiment are presented in Table 2.

As can be seen from the results in the Table 2, for the three neural networks, the accuracy of edema area selection is improved. Further research will be carried out with the training of neural networks on images of the fundus combined with a map of objects. For fundus images, some researchers use a special kind of preprocessing, in which the original fundus image is added (weighted sum) to the same image processed using a Gaussian filter with a large kernel. Fig. 3 shows source image and Fig. 4 shows an example of applying this preprocessing to one of the images from the original dataset. Table 3 presents the results of an experiment that checks the effect of this preprocessing on neural network training in terms of the f1 metric.

The results of the conducted research showed that the use of such preprocessing does not have a clear positive effect on the accuracy of the neural network.

In order to obtain reliable results of the experiment, k-fold cross-validation was carried out. The data set was randomly divided into 5 non-overlapping groups so that 40 images were included in the training set, and 10 of them were included in the test set.

Table 4 shows the average result of the experiment in terms of precision, recall, f1 metrics. To evaluate each group, the best epoch in terms of the f1 metric was taken.

Table 2 Results of the experiment to check the effect of adding map of objects (f1 metric).

Architecture	Source image	Combined images
Unet	0.584	0.620
Unet++	0.562	0.632
Linknet	0.575	0.608
PAN	0.512	0.446

Table 3 Results of an experiment to test the effect of fundus image preprocessing (f1 metric).

Architecture	Source image	Preprocessed image
Unet	0.620	0.609
Unet++	0.632	0.626
Linknet	0.608	0.594
PAN	0.446	0.468

Table 4 Results of cross-validation (f1 metric).

Architecture	Linknet	PAN	Unet	UnetPlus Plus
precision	0.728	0.674	0.648	0.716
recall	0.565	0.533	0.470	0.592
f1	0.634	0.594	0.544	0.647

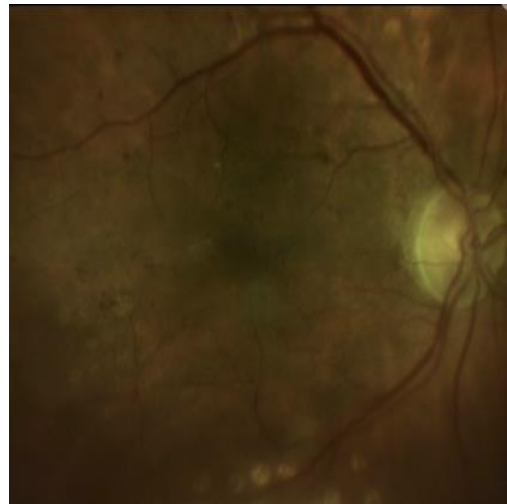


Fig. 3 Source image (1024 px × 1024 px).

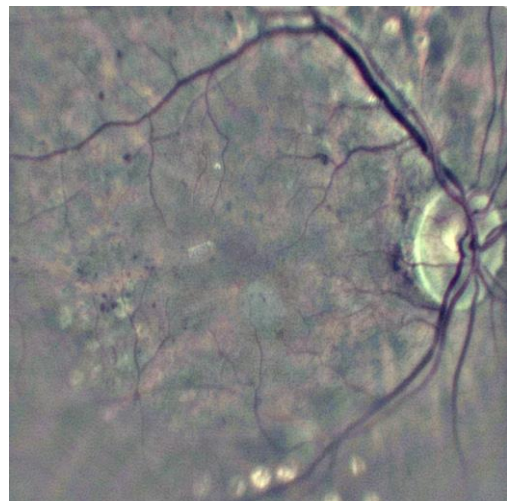


Fig. 4 Image after preprocessing (1024 px × 1024 px).

Thus, according to the results obtained, the most accurate model turned out to be Unet++. Fig. 5 shows an example of the results of semantic segmentation of images.

As mentioned earlier, when working with biomedical data, there is a problem of inaccurate labeling. In this work, the labeling problem is due to two factors. The marking of anatomical objects in the fundus is not absolutely accurate at the pixel level. For example, the edge of the vessels is inaccurately segmented, there is a small group of pixels that belong to the edge, but not marked by the expert. Also, the labeling of the edema area is not accurate due to the problem that the edema itself is not always visible in a planar image. For this reason, the ophthalmologist was guided by OCT data when marking, so the resulting marking is not always correct in the area where the edema is not clearly expressed.

With the help of augmentation techniques, this problem was partially solved using the generalizing ability of the neural network. Also, due to the specificity of this task, a direct comparison with the results of other studies cannot be correctly carried out.

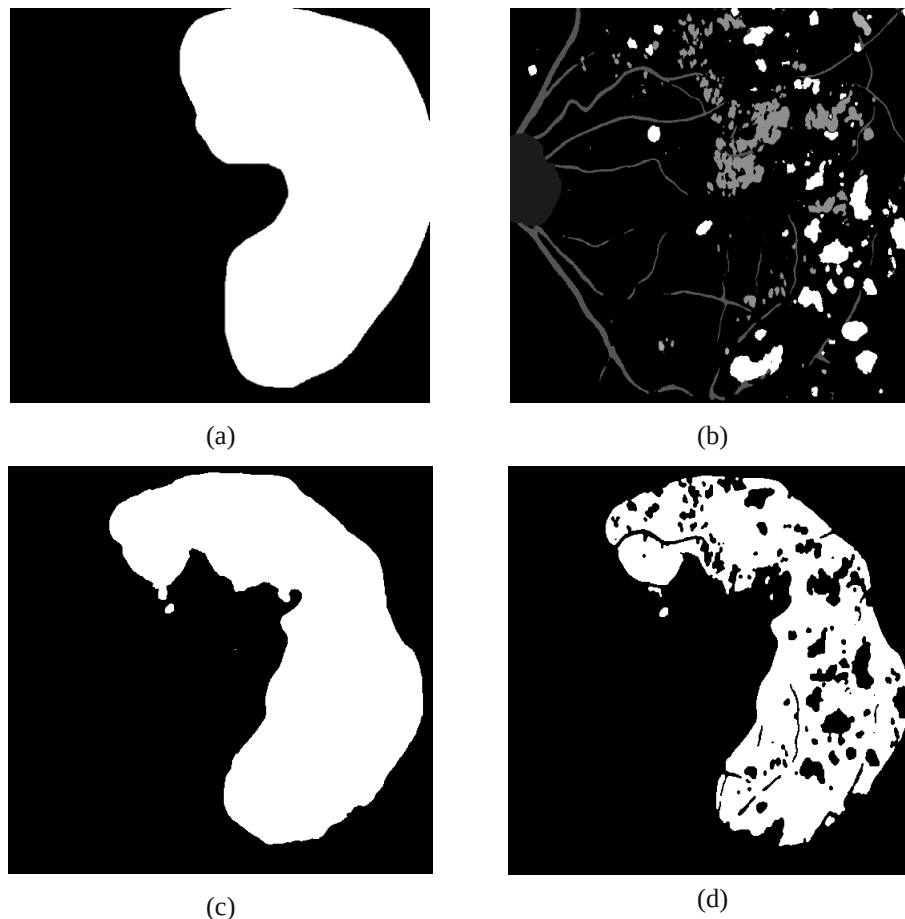


Fig. 5 Results of predictions of the area of macular edema by the neural network (1024 px × 1024px): (a) expert labels, (b) prediction of the neural network of the object map, (c) results of predictions of the neural network of the edema area, and (d) processed result.

Both the formulation of the problem itself and the initial data differ. This work is part of a semi-automatic system for creating a coagulation plan. The results obtained in the course of the work were subjected to expert analysis by an ophthalmologist and were considered sufficient for use in the system for constructing a coagulation plan. The hypothesis was confirmed that despite the weak severity of edema in the images of the fundus, neural networks are able to highlight it.

Further research will be aimed at improving the accuracy of segmentation of the fundus regions.

4 Conclusion

In this work, a method was developed and investigated for automatic selection of the optimal laser exposure zone from the fundus images for laser coagulation, based on the use of two neural networks. An important step

presented in this paper is the selection of the area of macular edema.

As a result of the research, it was shown that the use of the Unet++ model to highlight the edema area gives the highest accuracy relative to other architectures considered in the framework of the work. improve the accuracy of differential diagnosis and is devoid of subjectivity.

5 Acknowledgment

This work was performed within the State Assignment of FSRC “Crystallography and Photonics” RAS.

Disclosures

The authors declare no conflict of interest.

References

1. D. N. Louis, A. Perry, P. Wesseling, D. J. Brat, I. A. Cree, D. Figarella-Branger, C. Hawkins, H. K. Ng, S. M. Pfister, G. Reifenberger, R. Soffietti, A. von Deimling, and D. W. Ellison, “[The 2021 WHO Classification of Tumors of the Central Nervous System: a summary](#),” *Neuro-Oncology* 23(8), 1231–1251 (2021).
2. Y. Zheng, M. He, and N. Congdon, “[The worldwide epidemic of diabetic retinopathy](#),” *Indian Journal of Ophthalmology* 60(5), 428–431 (2012).

3. I. V. Vorobieva, D. A. Merkusheva, "Diabetic retinopathy in patients with type 2 diabetes mellitus. Epidemiology, a modern view of pathogenesis," *Ophthalmology* 9(4), 18–21 (2012).
4. I. I. Dedov, M. V. Shestakova, and O. K. Vikulova, "State Register of Diabetes Mellitus in the Russian Federation: Status of 2014 and Development Prospects," *Diabetes Mellitus* 18(3), 5–23 (2015).
5. I. I. Dedov, M. Shestakova, and G. R. Galstyan, "Prevalence of type 2 diabetes mellitus in the adult population of Russia (NATION study)," *Diabetes Mellitus* 19(2), 104–112 (2016).
6. X. Zhang, J. B. Saaddine, C. F. Chou, M. F. Cotch, Y. J. Cheng, L. S. Geiss, E. W. Gregg, A. L. Albright, B. E. Klein, and R. Klein, "Prevalence of diabetic retinopathy in the United States 2005–2008," *JAMA* 304, 649–656 (2010).
7. L. Guariguata, D. R. Whiting, I. Hambleton, J. Beagley, U. Linnenkamp, and J. E. Shaw, "Global estimates of diabetes prevalence for 2013 and projections for 2035," *Diabetes Research and Clinical Practice* 103(2), 137–49 (2014).
8. A. N. Amirov, E. A. Abdulaeva, and E. L. Minkhuzina, "Diabetic macular edema: Epidemiology, pathogenesis, diagnosis, clinical presentation, and treatment," *Kazan Medical Journal* 96(1), 70–74 (2015).
9. R. Klein, B. E. K. Klein, S. E. Moss, M. D. Davis, and D. L. DeMets, "The Wisconsin Epidemiologic Study of Diabetic Retinopathy IV. Diabetic Macular Edema," *Ophthalmology* 91(12), 1464–1474 (1984).
10. A. P. Goidin, O. L. Fabrikantov, and E. V. Sukhorukova, "The effectiveness of classical and pattern laser coagulation in diabetic retinopathy," *Bulletin of Russian Universities. Mathematics* 19(4), 1105–1107 (2014). [in Russian]
11. T. Moutray, J. R. Evans, N. Lois, D. J. Armstrong, T. Peto, and A. Azuara-Blanco, "Different lasers and techniques for proliferative diabetic retinopathy," *Cochrane Database of Systematic Reviews* (3), John Wiley & Sons (2018).
12. E. A. Zamytsky, A. V. Zolotarev, E. V. Karlova, and P. A. Zamytsky, "Analysis of the coagulates intensity in laser treatment of diabetic macular edema in a Navilas robotic laser system," *Saratov Journal of Medical Scientific Research* 13(2), 375–378 (2017). [in Russian]
13. J. J. Jung, R. Gallego-Pinazo, A. Lleó-Pérez, J. I. Huz, and I. A. Barbazetto, "Navilas laser system focal laser treatment for diabetic macular edema - one-year results of a case series," *The Open Ophthalmology Journal* 7, 48–53 (2013).
14. P. B. Velichko, "Comprehensive treatment of diabetic macular edema," *Bulletin of Russian Universities. Mathematics* 19(4), 1097–1101 (2014).
15. V. Vinokurov, Y. Khristoforova, O. Myakinin, I. Bratchenko, A. Moryatov, A. Machikhin, and V. Zakharov, "Neural network classifier of hyperspectral images of skin pathologies," *Computer Optics* 45(6), 879–886 (2021).
16. Yu. Kh. Ganeeva, E. V. Myasnikov, "Identifying persons from Iris images using neural networks for image segmentation and feature extraction," *Computer Optics* 46(2), 308–316 (2022).
17. I. D. Apostolopoulos, T. A. Mpesiana, "COVID-19: Automatic detection from X-ray images utilizing transfer learning with Convolutional Neural Networks," *Physical and Engineering Sciences in Medicine* 43(2), 635–640 (2020).
18. S. A. Abdelaziz Ismael, A. Mohammed, and H. Hefny, "An enhanced deep learning approach for brain cancer MRI images classification using residual networks," *Artificial Intelligence in Medicine* 102, 101779 (2020).
19. A. M. Arellano, W. Dai, S. Wang, X. Jiang, and L. Ohno-Machado, "Privacy policy and technology in biomedical data science," *Annual Review of Biomedical Data Science* 1, 115–129 (2018).
20. C. Shorten, T. M. Khoshgoftaar, "A survey on image data augmentation for Deep Learning," *Journal of Big Data* 6(1), 1–48 (2019).
21. E. Castro, J. S. Cardoso, and J. C. Pereira, "Elastic deformations for data augmentation in breast cancer mass detection," in *IEEE EMBS International Conference on Biomedical & Health Informatics (BHI)*, 230–234 (2018).
22. H. Ishwaran, R. O'Brien, "Commentary: The problem of class imbalance in biomedical data," *The Journal of Thoracic and Cardiovascular Surgery* 161(6), 1940–1941 (2021).
23. F. Charte, A. J. Rivera, M. J. del Jesus, and F. Herrera, "MLSMOTE: Approaching imbalanced multilabel learning through synthetic instance generation," *Knowledge-Based Systems* 89, 385–397 (2015).
24. R. M. Pereira, Y. M. G. Costa, and C.N. Silla Jr., "MLTL: A multi-label approach for the Tomek link undersampling algorithm," *Neurocomputing* 383, 95–105 (2020).
25. A. Mukhin, I. Kilbas, R. Paringer, and N. Ilyasova, "Application of the gradient descent for data balancing in diagnostic image analysis problems," in *International Conference on Information Technology and Nanotechnology (ITNT)*, Samara, Russia (2020).
26. A. V. Mukhin, I. A. Kilbas, R. A. Paringer, N. Y. Ilyasova, and A. V. Kupriyanov, "A method for balancing a multi-labeled biomedical dataset," *Integrated Computer-Aided Engineering* 29(2), 209–225 (2022).

27. D. Hao, L. Zhang, J. Sumkin, A. Mohamed, and S. Wu, “[Inaccurate labels in weakly-supervised deep learning: Automatic identification and correction and their impact on classification performance](#),” *IEEE Journal of Biomedical and Health Informatics* 24(9), 2701–2710 (2020).
28. C. Tian, T. Fang, Y. Fan, and W. Wu, “[Multi-path convolutional neural network in fundus segmentation of blood vessels](#),” *Biocybernetics and Biomedical Engineering* 40(2), 583–595 (2020).
29. J. Kaur, D. Mittal, “[A generalized method for the segmentation of exudates from pathological retinal fundus images](#),” *Biocybernetics and Biomedical Engineering* 38(1), 27–53 (2018).
30. N. Bhagat, R. A. Grigorian, A. Tutela, and M. A. Zarbin, “[Diabetic macular edema: Pathogenesis and treatment](#),” *Survey of Ophthalmology* 54(1), 1–32 (2009).
31. R. A. Paringer, A. V. Mukhin, N. Yu. Ilyasova, and N. S. Demin, “[Neural networks application for semantic segmentation of fundus](#),” *Computer Optics* 46(4), 596–602 (2022).
32. K. He, X. Zhang, S. Ren, and J. Sun, “[Deep Residual Learning for Image Recognition](#),” *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 770–778 (2016).
33. A. Krizhevsky, I. Sutskever, and G. Hinton, “[Imagenet classification with deep convolutional neural networks](#),” *Advances in Neural Information Processing Systems* 25, 1097–1105 (2012).
34. F. Iandola, M. Moskewicz, S. Karayev, R. Girshick, T. Darrell, and K. Keutzer, “[DenseNet: Implementing efficient convnet descriptor pyramids](#),” *arXiv preprint arXiv:1404.1869* (2014).
35. F. Chollet, “[Xception: Deep learning with depthwise separable convolutions](#),” *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 1251–1258 (2017).
36. T. Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollar, “[Focal loss for dense object detection](#),” *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* 6 2980–2988 (2017).
37. D. P. Kingma, J. Ba, “[Adam: A Method for Stochastic Optimization](#),” *arXiv preprint arXiv:1412.6980* (2014).
38. O. Ronneberger, P. Fischer, and T. Brox, “[U-net: Convolutional networks for biomedical image segmentation](#),” in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, N. Navab, J. Hornegger, W. M. Wells, and A. F. Frangi (Eds.), Springer International Publishing, 9351, 234–241 (2015).
39. Z. Zhou, M. M. Rahman Siddiquee, N. Tajbakhsh, and J. Liang, “[UNet++: A nested U-Net Architecture for Medical Image segmentation](#),” in *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*, D. Stoyanov, Z. Taylor, G. Carneiro, T. Syeda-Mahmood, A. Martel, L. Maier-Hein, J. M. R. S. Tavares, A. Bradley, J. P. Papa, V. Belagiannis, J. C. Nascimento, Z. Lu, S. Conjeti, M. Moradi, H. Greenspan, and A. Madabhushi (Eds.), Springer International Publishing, 11045, 3–11 (2018).
40. T. Fan, G. Wang, Y. Li, and H. Wang, “[Ma-net: A multi-scale attention network for liver and tumor segmentation](#),” *IEEE Access* 8, 179656–179665 (2020).
41. A. Chaurasia, E. Culurciello, “[LinkNet: Exploiting encoder representations for efficient semantic segmentation](#),” in *IEEE Visual Communications and Image Processing (VCIP)*, Saint Petersburg, FL, USA (2017).
42. T.Y. Lin, P. Dollar, R. Girshick, K. He, B. Hariharan, and S. Belongie, “[Feature Pyramid Networks for Object Detection](#),” *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2117–2125 (2017).
43. H. Zhao, J. Shi, X. Qi, X. Wang, and J. Jia, “[Pyramid Scene Parsing Network](#),” *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2881–2890 (2017).
44. H. Li, P. Xiong, J. An, and L. Wang, “[Pyramid Attention Network for Semantic Segmentation](#),” *arXiv preprint arXiv:1805.10180* (2018).
45. L. C. Chen, G. Papandreou, F. Schroff, and H. Adam, “[Rethinking Atrous Convolution for Semantic Image Segmentation](#),” *arXiv preprint arXiv:1706.05587* (2017).