

# Enhanced Segmentation of Fundus Images for Glaucoma Detection Using a New ECDR Model

**Orken Mamyrbayev**

Laboratory of Computer Engineering of Intelligent Systems, Institute of Information and Computational Technologies, Almaty, Kazakhstan  
morkenj@mail.ru

**Bakzhan Sakenov**

Laboratory of Computer Engineering of Intelligent Systems, Institute of Information and Computational Technologies, Almaty, Kazakhstan  
sbakzhan22@gmail.com

**Ardan Zhanegizov**

Satbayev University, Almaty, Kazakhstan  
ardan.zhanegiz@mail.kz (corresponding author)

**Nurzhamal Oshanova**

Abai Kazakh National Pedagogical University, National Academy of Sciences, Almaty, Kazakhstan  
n.oshanova@abaiuniversity.edu.kz (corresponding author)

**Guldina Kamalova**

Abai Kazakh National Pedagogical University, National Academy of Sciences, Almaty, Kazakhstan  
g.kamalova@abaiuniversity.edu.kz

**Sheripidin Khamrayev**

Abai Kazakh National Pedagogical University, National Academy of Sciences, Almaty, Kazakhstan  
sh.khamrayev@abaiuniversity.edu.kz

**Indira Salgozha**

Abai Kazakh National Pedagogical University, National Academy of Sciences, Almaty, Kazakhstan  
i.salgozha@abaiuniversity.edu.kz

*Received: 20 November 2025 | Revised: 17 December 2025 and 29 December 2025 | Accepted: 30 December 2025*

*Licensed under a CC-BY 4.0 license | Copyright (c) by the authors | DOI: <https://doi.org/10.48084/etasr.16392>*

## ABSTRACT

This study uses deep learning methods to detect and predict glaucoma before symptoms appear. A dataset for glaucoma analysis was collected from open data sources provided by the Institute of Eye Diseases in Almaty, Kazakhstan. This study proposes an Efficient Convolutional-Dual-Resolution (ECDR) model that employs a Convolutional Neural Network (CNN) and a Transformer-based hybrid method to segment medical images of glaucoma. The global features obtained from the Swin Transformer module are combined with local representations obtained from a CNN-based encoder. The resulting data are compared with CNN classification methods, showing that the proposed ECDR model outperforms previous solutions. This study highlights the effectiveness of the proposed ECDR model in improving the detection and prediction of glaucoma, suggesting potential implications for early diagnosis and treatment strategies.

*Keywords-ECDR; glaucoma; CNN; image segmentation; glaucoma detection*

## I. INTRODUCTION

Glaucoma is one of the most severe eye diseases, and, according to statistics from the World Health Organization, 4% of people worldwide aged 40-80 years suffer from it. One of the most important tasks is to obtain an accurate and timely diagnosis of glaucoma to prevent damage to vision in the future [1, 2]. Glaucoma disease in humans occurs with a constant or periodic increase in intraocular pressure, which damages the blood vessels and optic nerves of the eye. Glaucoma can lead to complete blindness. This disease is called a "hidden disease" because in its early stages, most patients do not show any symptoms. There are mainly two forms of glaucoma: open-angle and closed-angle. In addition, there is congenital glaucoma associated with anomalies in the development of the eye. Open-angle glaucoma occurs as a result of damage to the optic nerve. Closed-angle glaucoma, also called acute glaucoma, occurs when the drainage system of the eye completely fails.

Several factors contribute to the disease, including a drooping iris and a tiny pupil in the eye affected by the disease. When there is tension in the trabecular meshwork of the eye, the drainage channels may be forced closer to the iris. This type of damage can occur when the intraocular pressure increases too much. In many cases, nerve damage occurs when the retinal ganglion cells are lost. Increased intraocular pressure is the most important risk factor for glaucoma. However, some people may have high intraocular pressure for many years and never experience damage, while others may experience nerve damage with relatively low pressure. Since there is currently no treatment or cure for glaucoma, ophthalmologists must rely on diagnosis and treatment to keep the disease under control. Another problem with glaucoma is that different racial and ethnic groups experience its effects at different rates.



Fig. 1. Image of the eye obtained through a fundus camera.

Currently, in medicine, there is a deepening of knowledge about the development of pathological changes in the human body, so modern information devices and methods are needed to process biomedical information in images. To obtain an image of the eye, special equipment is used, such as a fundus camera. With the advancement of deep learning, medical image processing has become very effective. Researchers aim to identify and predict glaucoma using deep learning technologies before it becomes a significant problem. Many studies, such as [3, 4], use open data, Convolutional Neural Networks (CNNs), and segmentation methods, but there are still limitations that make the detection of glaucoma less reliable.

This study aimed to increase the accuracy of the diagnosis of glaucoma diagnosis using fundus images. The proposed Efficient Convolutional-Dual-Resolution (ECDR) model can significantly improve the ability to detect glaucoma. ECDR adds a layer of analysis through its hierarchical attention structure, allowing the system to more accurately interpret complex medical images, improving diagnosis. The proposed method involves the following steps:

- Preprocessing: Eye images are preprocessed to improve quality and highlight important features.
- CNNs analyze images to identify signs associated with glaucoma, such as changes in the optical disc or nerve fibers.
- ECDR is used to further analyze features at different levels of abstraction, which helps to improve classification and prediction accuracy.
- Classification and diagnosis: The model classifies images for the presence or absence of signs of glaucoma and helps in diagnosis.

This approach may lead to the development of more sensitive and accurate diagnostic tools capable of detecting glaucoma early and preventing vision loss. This study presents the integration of ECDR with a CNN to improve the performance of structured glaucoma prediction.

## II. BACKGROUND

Medical image processing and data extraction play a crucial role in pattern recognition, offering valuable insights into the anatomical regions necessary for detailed analysis. This information is vital for healthcare professionals to characterize injuries, monitor disease progression, and determine appropriate treatment strategies. As the use of medical image analysis continues to expand, the demand for highly accurate and reliable disease detection methods becomes increasingly critical. CNNs have emerged as a powerful tool for various medical image segmentation tasks due to their ability to effectively extract image features. The development of hybrid models, such as encoder-decoder networks, Fully Convolutional Networks (FCNs), and U-Net architectures, has significantly advanced the field of medical image processing [5].

For instance, in [6], an automated method was proposed for detecting the Retinal Nerve Fiber Layer (RNFL), which is crucial in the diagnosis of glaucomatous optic neuropathy. This approach utilized texture features derived from a coincidence matrix, combined with a backpropagation neural network as a classifier. In [7], an ACNN was trained and tested on a large glaucoma database, demonstrating impressive abilities to generalize classifications, recognize potentially dangerous conditions, and predict areas requiring attention. This model achieved 96.2% accuracy with an Area Under the Curve (AUC) of 0.983. However, the model faced challenges in accurately identifying small problem areas, and the return on investment was not thoroughly evaluated. In [8], integrating pretrained transfer learning models with a U-Net architecture and employing a DenseNet-201 deep convolutional neural

network offered high accuracy rates of 98.82% in training and 96.90% in testing, demonstrating improved performance over existing classification methods.

In [9], the focus was on distinguishing between the optic disc and fundus images using neural networks trained on data from various cameras. Separation accuracies of 93% for the optic cup and 98% for differentiating the cup from the disc were achieved by employing architectures such as Xception, VGG-16, VGG-19, and ResNet50. In [10], the effectiveness of pretrained CNNs, specifically OverFeat and VGG-S, was investigated for automated glaucoma detection in fundus images. This study explored various preprocessing techniques, such as vessel inpainting and Contrast-Limited Adaptive Histogram Equalization (CLAHE), to enhance feature discrimination, ultimately demonstrating the potential of this approach with significant positive results on the Drishti-GS1 dataset.

Currently, there are numerous algorithms for segmentation and detection of glaucoma, all of which require high-quality data. These algorithms often involve learning feature descriptors based on Support Vector Machines (SVM) and their combinations. In particular, studies utilizing DeepLabV3+ [11] and MobileNet [12] have reported object identification and segmentation accuracies of 99.7% and 99.53%, respectively. Deep learning techniques have further enhanced models such as G-Net [13] and U-Net [14] for glaucoma diagnosis.

The efficacy of deep learning methods in classifying and segmenting diseases in fundus images has been well-documented, particularly as the volume of images continues to grow. However, training these models requires clear labels for successful generalization, which can be challenging due to the high costs associated with expert annotations. In [15], multi-task deep learning models leveraged similarities between related fundus tasks and measurements used in the diagnosis of glaucoma. This model simultaneously learns various segmentation and classification tasks, resulting in a glaucoma detection accuracy of  $93.56 \pm 1.48$ .

In [16], machine learning models demonstrated high predictive capacity and interpretability for glaucoma diagnosis based on the thickness of the retinal nerve fiber layer and visual field assessments. This approach synthesized features from the original data and selected the most suitable ones for classification. The Random Forest model exhibited the best performance, achieving a classification accuracy of 98%, with sensitivity at 98.3%, specificity at 97.5%, and an AUC of 97.9%. These models showed remarkable accuracy in distinguishing between glaucoma patients and healthy individuals. The methodologies developed for glaucoma diagnosis effectively combine results from two distinct types of measurements—temporal and spatial data—leading to reliable diagnoses of eye conditions. This technology has the potential to enhance the identification and treatment of patients suffering from the disease.

### III. DATA PREPARATION

This study used two publicly available datasets: REFUGE [17] and QazGo [18]. The REFUGE dataset comprises 1,200 annotated color fundus images, organized into three subsets of 400 images, each for training, validation, and testing purposes. The training set specifically includes 40 glaucomatous and 360 normal images, emphasizing the importance of early detection of glaucoma, as lost vision cannot be restored and regular eye check-ups are essential. This dataset facilitates independent analysis of fundus images for both eyes. In contrast, the QazGo dataset, collected from an eye center in Almaty, Qazaqstan, consists of 791 fundus images that reflect a variety of conditions and demographics, enhancing model generalizability. Each image in QazGo is annotated with a mask for the optical disk, and although initial annotations included both cup and disc structures, the approach was streamlined to ensure uniformity by overlapping cup annotations with the disc. All images in this dataset are standardized to a resolution of 640×640 pixels.

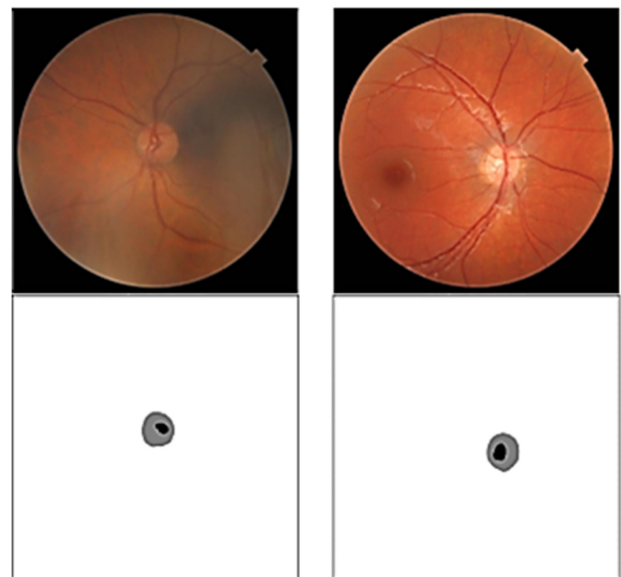


Fig. 2. Examples of fundus images from the QazGo dataset.

### IV. METHODS

ECDR is a hierarchical encoder-decoder architecture designed for high-accuracy, fast glaucoma detection from retinal fundus images. Figure 3 shows its architecture. The model combines a lightweight EfficientNet-V2 backbone with Convolutional Block Attention Modules (CBAM) in the encoder and a novel Dense Upsampling Convolution Kernel (DUCK) decoder that employs multi-scale residual blocks. This design maximizes feature representation while keeping computational demands modest, making it suitable for both large-scale clinical datasets and resource-constrained environments.

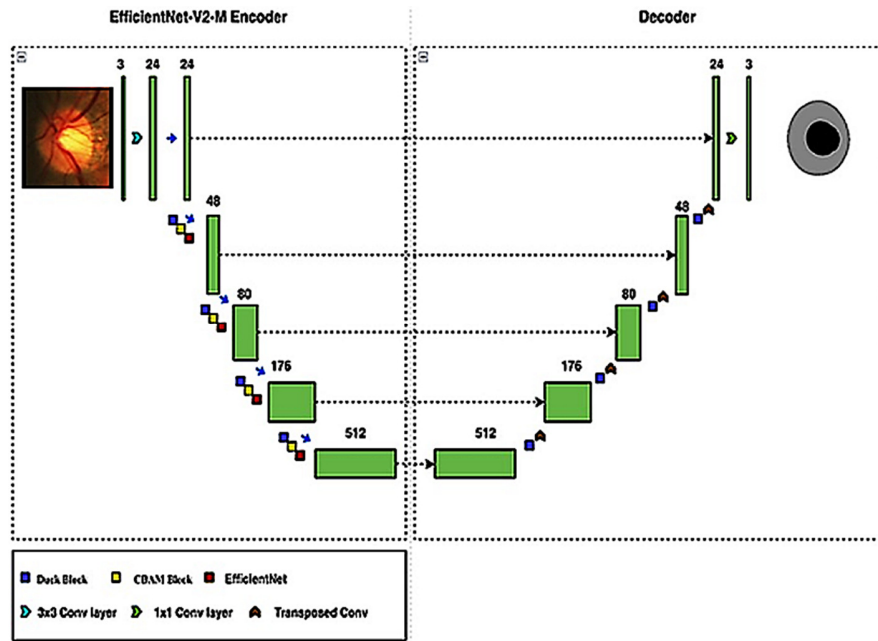


Fig. 3. ECDR architecture.

### A. Encoder

The encoder adopts EfficientNet-V2-M as its backbone [19], leveraging fused MBConv blocks and progressive image-size scaling to achieve a favorable trade-off between accuracy and training speed. EfficientNet-V2-M processes the input RGB fundus image ( $3 \times H \times W$ ) through an initial  $1 \times 1$  convolutional projection, followed by a cascade of EfficientNet blocks that progressively reduce spatial resolution while expanding channel depth ( $24 \rightarrow 48 \rightarrow 80 \rightarrow 176 \rightarrow 512$ ). The EfficientNetV2 compound scaling techniques optimize the architecture for both accuracy and efficiency. EfficientNetV2, initially introduced for classification, seamlessly adapts to segmentation tasks. Using it as an encoder accelerates training, reduces the model's footprint, and preserves—or even enhances—segmentation performance across diverse datasets.

To enhance discriminative power, each EfficientNet block is augmented with a CBAM attention module. CBAM applies sequential channel attention, which aggregates spatial information via global average- and max-pooling to generate a channel-wise weighting, and spatial attention, which compresses channel information using a  $7 \times 7$  convolution to produce a spatial mask. The combined attention maps are multiplied with the feature maps, enabling the encoder to emphasize clinically relevant structures such as the optic disc and cup while suppressing background noise.

### B. Decoder

The decoder mirrors the encoder's hierarchical structure but reverses the resolution changes using transposed convolutions for up-sampling. At each decoding stage, the following operations are performed:

1. Transposed convolution (kernel = 2, stride = 2) to double the spatial dimensions.

2. DUCK block, a dense up-sampling module that replaces conventional interpolation. Each DUCK block consists of three parallel residual pathways that simulate effective kernel sizes of  $5 \times 5$ ,  $9 \times 9$ , and  $13 \times 13$  by stacking one, two, and three residual units, respectively. Dilated convolutions are employed in the  $9 \times 9$  and  $13 \times 13$  pathways (Midscope and Widescope) to enlarge receptive fields without increasing parameter count.
3.  $3 \times 3$  convolution to refine the up-sampled features.

Skip connections connect the encoder and decoder stages of identical resolution, concatenating high-resolution encoder features with the up-sampled decoder maps. This fusion preserves fine-grained anatomical details essential for precise segmentation of the optic nerve head.

## V. RESULTS

### A. Evaluation Metrics

The Dice coefficient (DC) and Intersection over Union (IoU) were used to evaluate the segmentation results. The Dice Coefficient measures the overlap between two samples.

$$DC = \frac{2 * TP}{2 * TP + FP + FN} \quad (1)$$

where TP is True Positives, FP is True Negatives, and FN is False Negatives. IoU is another metric to evaluate the accuracy of an object detector:

$$IoU = \frac{TP}{TP + FP + FN} \quad (2)$$

### B. Implementation

The PyTorch platform was used on a desktop computer with Linux, an Intel® Xeon® E5-2620 v, and an NVIDIA A100 x2 graphics card, to provide sufficient computing power to train the network. An 80:20 training-testing split was

adopted, and the model was trained for 100 epochs. An input image size of  $512 \times 512$ , batch size of 3, and a learning rate of  $1 \times 10^{-5}$  were the default parameters for training. The model was optimized using the SGD optimizer with a momentum of 0.9. Automatic learning rate decay was based on the validation set score. To provide an objective comparison, the proposed ECDR was combined with CNN methods, as well as with models formulated based on their combination.

### C. Experimental Results

Figure 4 illustrates the training progress of the ECDR model. The validation DC indicates the model's performance in accurately segmenting the target objects during validation. The training loss values are shown, reflecting the model's learning progress over time. This dual representation allows for a comprehensive understanding of the model's training dynamics, highlighting the relationship between training loss and validation performance. Table I presents the evaluation results of the ECDR in terms of  $DC_{disc}$ ,  $DC_{cup}$ , and  $IoU$ .

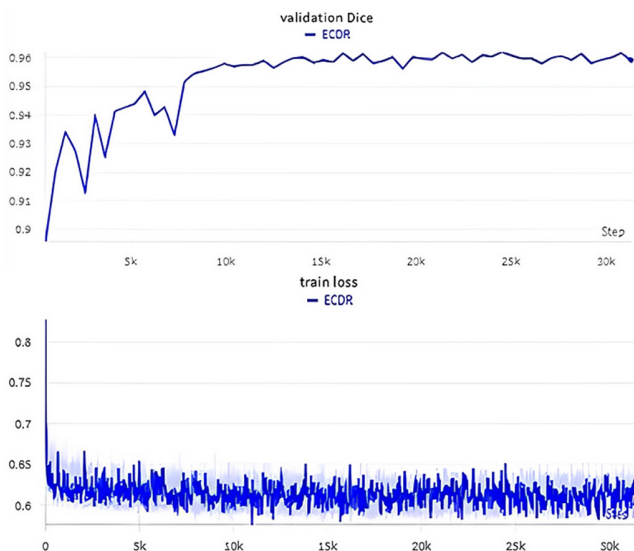


Fig. 4. ECDR model's training progress.

TABLE I. EVALUATION PERFORMANCE RESULTS

Method	$DC_{disc}$	$DC_{cup}$	$IoU$
[20]	0.9488	0.8643	0.9311
[21]	0.9532	0.8600	0.9042
[17]	0.9202	0.8826	-
Proposed	0.9660	0.8890	0.8890

## VI. DISCUSSION

The proposed ECDR model is a hybrid method developed for the segmentation of medical images to detect glaucoma. By combining global characteristics obtained from the EfficientNet-V2-M module with local representations from a CNN-based encoder, the model effectively captures both the overall context and fine details of the images. The incorporation of the DUCK-Net and CBAM module further enhances the model's ability to unify these representations, allowing for improved attention to relevant features while minimizing the impact of

complex backgrounds. The ECDR model demonstrated superior performance compared to existing methods, particularly in  $DC_{disc}$ , where it achieved a score of 0.9660, outperforming all other methods, including previous benchmarks such as [20] (0.9488) and [21] (0.9532). In the  $DC_{cup}$  metric, the proposed method scored 0.8890, which is competitive and underscores the model's overall effectiveness.

The results highlight the capacity of the ECDR model to balance the preservation of low-level object details with the modeling of long-term interactions, making it a promising tool for clinical applications in glaucoma detection. Notable improvements over other methods, such as [20, 21], further contextualize the achievements of the ECDR model, demonstrating its superior performance in  $DC_{disc}$  and strong competitiveness in  $DC_{cup}$  compared to existing techniques. These findings not only validate the effectiveness of the proposed method but also emphasize its potential for early diagnosis and intervention in glaucoma, ultimately aiming to improve patient outcomes and reduce the risk of vision loss. Future research could explore avenues such as transfer learning [22], advanced data augmentation techniques, and the integration of multimodal data to improve robustness and diagnostic accuracy, contributing to more effective tools in ophthalmic imaging.

## VII. CONCLUSIONS

This study presented a novel hybrid model, ECDR, designed for the segmentation and detection of glaucoma in fundus images. The knowledge gap addressed by this research involves the limitations of existing CNN-based methods in accurately capturing both global and local features within medical images, particularly on complex backgrounds. The proposed approach integrates EfficientNetV2-M as an encoder and a DUCK-based decoder architecture, enhanced by CBAM attention modules, to effectively capture and unify these essential features critical for accurate glaucoma diagnosis. The basic steps included the collection and preprocessing of datasets, namely REFUGE and images from the Qazaqstan Eye Research Institute, followed by training the ECDR model under optimized conditions. The experimental results demonstrated that ECDR achieves a  $DC_{disc}$  score of 0.9660, outperforming existing methods. Additionally, ECDR maintained competitive results in  $DC_{cup}$ , scoring 0.8890, underscoring the model's overall effectiveness. The novelty of this work lies in the model's combination of hierarchical attention and multiscale convolution blocks, which significantly enhanced segmentation performance while maintaining computational efficiency. These advances not only contribute to the existing body of knowledge in ophthalmic imaging but also offer practical solutions for early screening and monitoring of glaucoma.

These findings validate the potential of ECDR as a reliable tool for detecting glaucoma, highlighting its ability to extract meaningful anatomical structures, even under challenging imaging conditions. Future work should focus on expanding the size of the dataset, testing in real-world clinical settings, and further exploring the generalizability of the model to other ophthalmological conditions, aiming to improve patient outcomes and reduce the risk of vision loss.

## ACKNOWLEDGMENT

This research was funded by the internal grant program of Abai Kazakh National Pedagogical University, according to Order No. 05-04/250 dated April 3, 2025, and supported by the University's internal budget.

## REFERENCES

- [1] F. Li *et al.*, "Joint optic disk and cup segmentation for glaucoma screening using a region-based deep learning network," *Eye*, vol. 37, no. 6, pp. 1080–1087, Apr. 2023, <https://doi.org/10.1038/s41433-022-02055-w>.
- [2] A. Murugan, B. Ashok, M. Dhanush, and S. Elankathir, "Automatic Classification and Earlier Detection of Diabetic Retinopathy Using Deep Learning," in *2023 9th International Conference on Advanced Computing and Communication Systems (ICACCS)*, Coimbatore, India, Mar. 2023, pp. 1455–1459, <https://doi.org/10.1109/ICACCS57279.2023.10113025>.
- [3] X. Chen, Y. Xu, D. W. Kee Wong, T. Y. Wong, and J. Liu, "Glaucoma detection based on deep convolutional neural network," in *2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, Milan, Aug. 2015, pp. 715–718, <https://doi.org/10.1109/EMBC.2015.7318462>.
- [4] Q. Abbas, "Glaucoma-Deep: Detection of Glaucoma Eye Disease on Retinal Fundus Images using Deep Learning," *International Journal of Advanced Computer Science and Applications*, vol. 8, no. 6, 2017, <https://doi.org/10.14569/IJACSA.2017.080606>.
- [5] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Boston, MA, USA, June 2015, pp. 3431–3440, <https://doi.org/10.1109/CVPR.2015.7298965>.
- [6] A. Septiarini, A. Harjoko, R. Pulungan, and R. Ekantini, "Automated Detection of Retinal Nerve Fiber Layer by Texture-Based Analysis for Glaucoma Evaluation," *Healthcare Informatics Research*, vol. 24, no. 4, pp. 335–345, Oct. 2018, <https://doi.org/10.4258/hir.2018.24.4.335>.
- [7] S. Gheisari *et al.*, "A combined convolutional and recurrent neural network for enhanced glaucoma detection," *Scientific Reports*, vol. 11, no. 1, Jan. 2021, Art. no. 1945, <https://doi.org/10.1038/s41598-021-81554-4>.
- [8] R. Kashyap *et al.*, "Glaucoma Detection and Classification Using Improved U-Net Deep Learning Model," *Healthcare*, vol. 10, no. 12, Dec. 2022, <https://doi.org/10.3390/healthcare10122497>.
- [9] H. S. Alghamdi, H. L. Tang, S. A. Waheeb, and T. Peto, "Automatic Optic Disc Abnormality Detection in Fundus Images: A Deep Learning Approach," *Proceedings of the Ophthalmic Medical Image Analysis International Workshop*, vol. 3, no. 2016, Oct. 2016, <https://doi.org/10.17077/omia.1042>.
- [10] J. I. Orlando, E. Prokofyeva, M. del Fresno, and M. B. Blaschko, "Convolutional neural network transfer for automated glaucoma identification," in *12th International Symposium on Medical Information Processing and Analysis*, Jan. 2017, Tandil, Argentina, <https://doi.org/10.1117/12.2255740>.
- [11] R. Kashyap and A. D. Pierson, "Big Data Challenges and Solutions in the Medical Industries," in *Handbook of Research on Pattern Engineering System Development for Big Data Analytics*, IGI Global Scientific Publishing, 2018, pp. 1–24.
- [12] R. Kashyap, "Big Data Analytics Challenges and Solutions," in *Big Data Analytics for Intelligent Healthcare Management*, Elsevier, 2019, pp. 19–41.
- [13] M. Juneja *et al.*, "Automated detection of Glaucoma using deep learning convolution network (G-net)," *Multimedia Tools and Applications*, vol. 79, no. 21, pp. 15531–15553, June 2020, <https://doi.org/10.1007/s11042-019-7460-4>.
- [14] P. H. Prastyo, A. S. Sumi, and A. Nuraini, "Optic Cup Segmentation using U-Net Architecture on Retinal Fundus Image," *JITCE (Journal of Information Technology and Computer Engineering)*, vol. 4, no. 02, pp. 105–109, Sept. 2020, <https://doi.org/10.25077/jitce.4.02.105-109.2020>.
- [15] L. Pascal, O. J. Perdomo, X. Bost, B. Huet, S. Otálora, and M. A. Zuluaga, "Multi-task deep learning for glaucoma detection from color fundus images," *Scientific Reports*, vol. 12, no. 1, July 2022, Art. no. 12361, <https://doi.org/10.1038/s41598-022-16262-8>.
- [16] S. J. Kim, K. J. Cho, and S. Oh, "Development of machine learning models for diagnosis of glaucoma," *PLOS ONE*, vol. 12, no. 5, 2017, Art. no. e0177726, <https://doi.org/10.1371/journal.pone.0177726>.
- [17] J. I. Orlando *et al.*, "REFUGEChallenge: A unified framework for evaluating automated methods for glaucoma assessment from fundus photographs," *Medical Image Analysis*, vol. 59, Jan. 2020, Art. no. 101570, <https://doi.org/10.1016/j.media.2019.101570>.
- [18] "IICT QazGo\_Glaucoma." Hugging Face, [Online]. [https://huggingface.co/datasets/iict/QazGo\\_Glaucoma](https://huggingface.co/datasets/iict/QazGo_Glaucoma).
- [19] O. Mamyrbayev, S. Pavlov, O. Karas, Y. Saldan, K. Momynzhanova, and S. Zhumagulova, "Increasing the reliability of diagnosis of diabetic retinopathy based on machine learning," *Eastern-European Journal of Enterprise Technologies*, vol. 2, no. 9 (128), pp. 17–26, Apr. 2024, <https://doi.org/10.15587/1729-4061.2024.297849>.
- [20] A. Sevastopolsky, "Optic disc and cup segmentation methods for glaucoma detection with modification of U-Net convolutional neural network," *Pattern Recognition and Image Analysis*, vol. 27, no. 3, pp. 618–624, July 2017, <https://doi.org/10.1134/S1054661817030269>.
- [21] B. Al-Bander *et al.*, "Dense Fully Convolutional Segmentation of the Optic Disc and Cup in Colour Fundus for Glaucoma Diagnosis," *Symmetry*, vol. 10, no. 4, Mar. 2018, <https://doi.org/10.3390/sym10040087>.
- [22] S. S. Mahmood, S. Chaabouni, and A. Fakhfakh, "Improving Automated Detection of Cataract Disease through Transfer Learning using ResNet50," *Engineering, Technology & Applied Science Research*, vol. 14, no. 5, pp. 17541–17547, Oct. 2024, <https://doi.org/10.48084/etasr.8530>.