An optimized segmentation of optic disc and optic cup in retinal fundus images based on multimap localization and conventional U-Net

Fadia Hanifa Suwandoko
Biomedical Engineering Department
Institut Teknologi Bandung
Bandung, Indonesia
18318029@std.stei.itb.ac.id

Astri Handayani
Biomedical Engineering Department
Institut Teknologi Bandung
Bandung, Indonesia
a.handyani@itb.ac.id

Tati Latifah Erawati Rajab
Biomedical Engineering Department
Institut Teknologi Bandung
Bandung, Indonesia
tati@stei.itb.ac.id

***Abstract*— Glaucoma is a vision-threatening condition resulting from increased intraocular pressure, leading to optic nerve damage and potential vision loss. Early detection is crucial, but in Indonesia, over half of glaucoma cases are diagnosed at severe stages. One of the approaches for early detection is by using characteristics visible in a fundus image as indicators. With this approach, the segmentation of optic disc and optic cup plays an important role to extract features such as cup-to-disc ratios or rim-to-disc ratios as an indication for glaucoma. Almustofa (2021) proposed an automated glaucoma detection method using these segmentations but only tested it on limited datasets, namely Drishti-GS and REFUGE Training Set. This study presents an optimized method by incorporating the REFUGE Validation and Test sets. Optic disc segmentation attains F-Scores of 0.979 ± 0.005 for Drishti-GS and 0.942 ± 0.026 for REFUGE. Optic cup segmentation achieves F-Scores of 0.948 ± 0.020 in Drishti-GS and 0.843 ± 0.068 in REFUGE datasets. These results demonstrate the improved performance of the optimized method for glaucoma detection.**

***Keywords*— Glaucoma, optic disc, optic cup, segmentation**

I. Introduction

Glaucoma is an ophthalmic disease that affects the optic nerve and is characterized by the progressive thinning of the retinal nerve layer resulting in cell death. Glaucoma ranks as the second most common cause of blindness globally, contributing to 11% of all blindness cases [1]. In 2020, there were 4.1 million reported cases of glaucoma worldwide, with more than 25% of those cases originating from East Asia, Southeast Asia, and Oceania. In Indonesia, a staggering 51.4% of glaucoma cases are only detected at an advanced stage, when there is already a significant ocular damage or severe reduction of visual acuity [2].

The primary cause of glaucoma is increased intraocular pressure resulting from the accumulation of aqueous humor within the eye. This buildup occurs due to the obstruction of the trabecular meshwork, which is responsible for the drainage of this fluid. As a result of this obstruction, intraocular pressure increases and compresses the optic nerves located in the optic nerve head (ONH) [3]. The ONH is comprised of the optic disc and the optic cup. When intraocular pressure increases, it causes the nerve cell layer in the ONH to thin, leading to the enlargement of the optic cup. This phenomenon is known as cupping, as seen in Figure 1, and can be quantified using the cup-to-disc ratio (CDR) [4]. The CDR is calculated by determining the ratio between the size of the optic disc and the optic cup. A threshold value derived from the CDR determines the presence of normal or glaucomatous eyes. A higher CDR suggest an increased likelihood glaucoma.

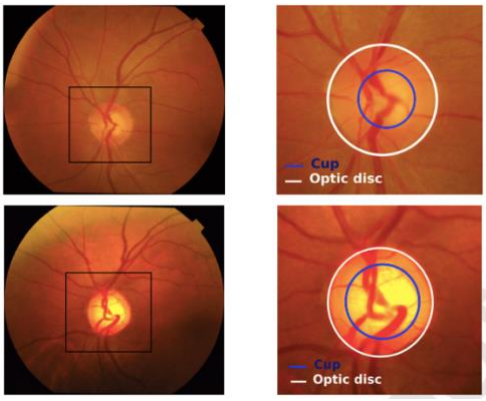

Fig. 1. Optic Nerve Head on retinal images. (1) Normal eye and (2) Glaucomatous eye.

In practice, glaucoma detection is typically performed by manual measurement of the CDR by a physician. This process is time-consuming and subjective, as the resulting segmentation and CDR are subject to the physician's judgment. Consequently, an automated method for the detection of glaucoma is necessary to address these issues [5].

II. Literature Review

Currently, convolutional neural networks (CNNs) are widely employed for semantic segmentation of the optic disc and optic cup. Among the commonly used architectures, U-Net is particularly popular for this purpose. Several previous studies have undertaken modifications to the U-Net architecture to cater to the specific requirements of optic disc and optic cup segmentation.

One example is the EfficientNet + Unet++ model by Kamble et al. This method replaces the U-Net encoder with EfficientNet B4 and adds skip pathways from the U-Net++ model in its decoder. This addition manages to improve model performance without increasing its computational burden [6]. Another approach is the ET-Net, which introduces an Edge guidance module (EGM) in its encoder to obtain constraints as a segmentation edge guide and a Weight Aggregation Module (WAM) to its decoder to emphasize the discovered features [7]. The POSAL method, presented by Wang et al. uses the DeepLabv3+ architecture and adversarial learning to reduce bias between training data. It provides good performance but requires significant computation [8]. Lastly, the M-Net architecture utilizes a U-Net model with multi-scale input and side output. This model receives fundus image input transformed polarly with different sizes on each layer of

its architecture and produces output at each layer with different sizes. These outputs are then weighted and summed to obtain the desired segmentation result [9].

Unlike the aforementioned studies, Almustofa [10] proposes the addition of a Region of Interest (RoI) localization module before semantic segmentation with U-Net. This method is simpler in terms of the convolutional neural network module, yet it can yield comparable results to other research findings on the Refuge and Dhristi-GS training data. However, it is important to note that the Almustofa model was solely tested on the REFUGE Train and Drishti-GS datasets. When applied to the REFUGE Validation dataset, it faced several challenges, and its performance was found to be less effective.

For localization, Almustofa employed the multiple map algorithm, which combines super pixel brightness maps and normal coefficient correlation (NCC) maps. Despite it producing sufficient results, there were still cases where localization was not captured accurately which could significantly affect the segmentation process. Additionally, due to differences in image size of the REFUGE Validation dataset compared to the dataset used during development, the ratio of the captured OD area in the RoI is not well-aligned.

In the segmentation process, Almustofa used a U-Net-based semantic segmentation technique and post-processing with ellipse fitting. This model produces binary masks of OD and OC candidate, and then fits these candidates into an ellipse on the binary mask. The performance of this model is significantly decreased on the new datasets, especially in segmentation of the optic cup. This is likely due to differences in image size and contrast between the Refuge Validation dataset and the Drishti and Refuge Training datasets. Moreover, in some images from the Refuge Validation dataset, multiple candidate OC regions were produced, which were not accounted for by Almustofa's post-processing algorithm. This led to incorrect candidate selection, resulting in an overall performance decrease of up to 16 percent.

III. METHODOLOGY

A. Dataset

The study used several datasets to train and evaluate the models developed in the study. These includes:

1. Drishti-GS dataset which contains 50 images with a size of 2047 x 1760 pixels, including 32 normal and 18 glaucomatous images [11].
2. REFUGE Training Set consisting of 400 images with a size of 2124 x 2056 pixels, including 360 normal and 40 glaucomatous images [12].
3. REFUGE Validation Set with 400 images with a size of 1634 x 1634 pixels, including 360 normal and 40 glaucomatous images [12].
4. REFUGE Test Set containing 400 images with a size of 1634 x 1634 pixels, including 360 normal and 40 glaucomatous images [12].

B. General Methodology

Previously, Almustofa [10] developed a segmentation method based on U-Net with ellipse fitting with steps as shown in Figure 2. However, the methodology faced challenges in all three steps. The localization step had issues with inaccurate localization and misalignment of the captured OD area in the region of interest (RoI). The segmentation step exhibited decreased performance on new datasets, especially in OC segmentation, due to differences in image size and contrast. The ellipse fitting post-processing did not consider multiple candidate OC regions, leading to incorrect candidate selection. These issues motivate the need for optimization in each step to enhance accuracy and generalization of the methodology, which will serve as the baseline for the current research.

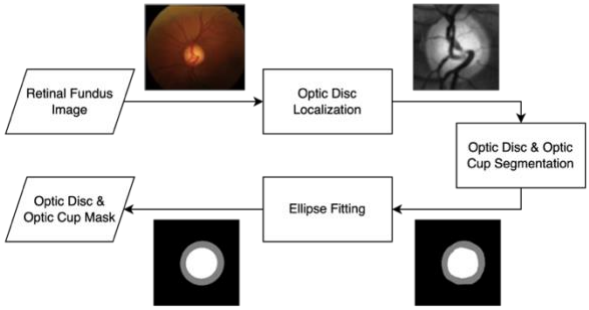

Fig. 2. The general method of optic disc and optic cup segmentation by Almustofa (2021) [10].

In this study, we address these challenges by proposing an optimization methodology for optic disc analysis in retinal fundus images. Firstly, we optimize the localization step by introducing a Region of Interest (RoI) size optimization that handles variations in image size within and between datasets. Secondly, to improve segmentation, we enhance the supervised learning method by incorporating an additional dataset, enhancing model robustness. Lastly, we address limitations in ellipse fitting by meticulously inspecting each obtained contour to select the appropriate optic cup candidate based on its relationship with the optic disc.

C. Optic Disc Localization

The preprocessing step proposed by Almustofa, prior to performing optic disc (OD) and optic cup (OC) segmentation, is the localization of the OD. This localization process serves to prepare the image before the segmentation process. The localization module in general utilizes an unsupervised algorithm-based localization method. This localization process involves several maps, including the Normalized Correlation Coefficient (NCC) map and the super pixel brightness map. From the input RGB retina image, three maps are generated: the red-channel NCC map, the green-channel NCC map, and the super pixel map. These three maps are weighted and summed, and the coordinate with the highest pixel value is taken as the center point of the RoI. The RoI is then extracted from the green-channel retina image that has undergone contrast-limited adaptive histogram equalization (CLAHE), with a cropping size of 550 x 550 pixel centered on the coordinate [10].

In this study, we propose an RoI size optimization using ratios relative to the image size to tackle variations in image size present within the Refuge dataset as well as between Drishti-GS and Refuge datasets. Due to the difference in image size between datasets, the initial fixed size of 550 × 550 pixels could no longer capture the optic disc (OD) in a manner that maintained uniform proportions, which is essential for accurate segmentation, especially in the presence of class imbalance. To address this issue, we proposed the use of ratios

relative to the image size, rather than a fixed pixel size. We then compare the segmentation performance under various RoI dimensions to evaluate the effectiveness of our proposed approach.

D. Segmentation

The semantic segmentation of OD and OC utilizes a supervised learning method based on U-Net with a workflow depicted in [13]. The evaluation metric used for assessing the performance of the segmentation model is F-Score. F-Score is particularly useful in situations where there is an imbalance between the classes, which is often the case in semantic segmentation tasks. The F-Score is calculated using equation (1).

$$F - Score = \frac{2 \times TP}{2 \times TP + FP + FN} \quad (1)$$

Where:

- True Positive (TP) represents correctly predicted positive samples.
- False Positive (FP) represents incorrectly predicted positive samples.
- False Negative (FN) represents incorrectly predicted negative samples.

Currently, the training data for this research consists of 450 images from the REFUGE Train and Drishti datasets, and the validation data consists of 400 images. These two datasets exhibit significantly different characteristics. In terms of image resolution, the Refuge validation data properly represent the Refuge test set but are much smaller than the resolution of the Refuge Training set. To fairly represents these variations, we choose to train the segmentation with a training to validation data ratio of 3:1.

E. Ellipse Fitting

After the segmentation process, post-processing is performed using ellipses fitting. The ellipses fitting process in [10] has not accounted for the possibility of multiple candidates. Thus, while searching for an OC candidate among detected ellipses, only the ellipse with the largest size is selected. Consequently, there is a possibility of OC entirely not detected in the fundus image.

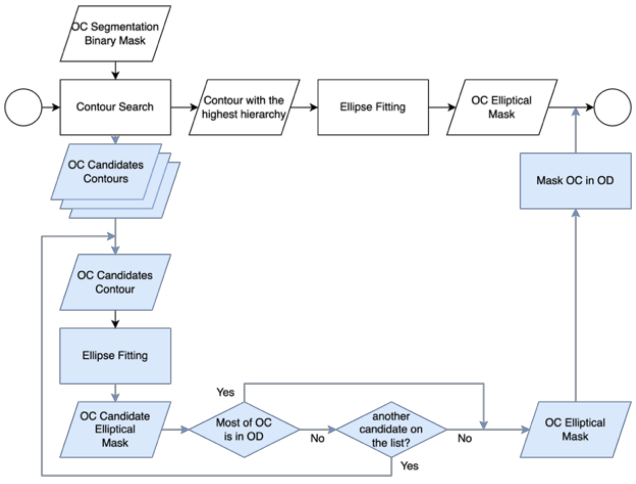

Fig. 3. The optimized ellips fitting workflow with the addition of multiple candidate handling (in blue).

With this optimization, each obtained contour will be inspected. Using the knowledge that the OC is always located inside the OD, each OC contour will be examined to determine whether most of its area is located within the OD. Contours that meet this specification will then be selected as the OC candidate, resulting in the detection of the appropriate contour. The optimized ellipse fitting workflow is illustrated in Figure 3.

IV. Results and Discussion

A. Optimized Localization

The proposed methodology involved experimenting with ratios calculated based on the size of the optic disc and the room for error in the gap between the brightest spot in the image and the actual center of the OD. This approach allowed us to consider the inherent variations in optic disc sizes and their positions within the images, enabling a more flexible and adaptive segmentation process.

We evaluated various ratios between 20% and 40% to determine the most appropriate one. The results of the experiment revealed that using a ratio of 30% was the best candidate for achieving optimal segmentation outcomes. By selecting the 30% ratio, we ensured a good balance between capturing sufficient information about the optic disc and leaving enough room for error in the gap between the brightest spot and the true center of the OD. This balance played a crucial role in achieving accurate and robust semantic segmentation results across different datasets, despite variations in image resolutions and other characteristics.

TABLE I. Segmentation Result with New RoI

Model No.	RoI of Model Data		OD F-Score	OC F-Score
	Training	Validation		
Baseline	550×550	550×550	0.913 ± 0.049	0.748 ± 0.154
1	550×550	30%	0.912 ± 0.051	0.720 ± 0.166
2	**30%**	**30%**	**0.917 ± 0.046**	**0.784 ± 0.122**

Initially, the RoI size was tested using the base as to test the new Region of Interest (RoI) size, the modified RoI size was evaluated using the baseline model trained with 550 × 550 data. However, in comparison to the baseline, the segmentation performance experienced a significant reduction. In response to the observed decrease in segmentation performance, we took proactive measures to address this issue. The model was retrained using the training data that incorporated the latest cropping specification based on the 30% image size ratio. This revision enabled the model to better accommodate the new RoI size and exploit the advantages of the ratio-based approach, leading to a more precise and accurate segmentation of the optic disc (OD) and optic cup (OC) regions.

Notably, the retraining process demonstrated a substantial improvement in OD and OC segmentation performance compared to the baseline model. The adoption of the 30% of the original image size as the superior image size alignment method proved to be an effective strategy, yielding superior results in terms of segmentation accuracy and generalization capability.

B. Optimized Segmentation

To begin optimization of the segmentation process for the additional datasets, a training and validation ratio first needed to be established. Due to limited number of data in the Drishti-

GS dataset and its resemblance in resolution with those of the Refuge Training set, we use all images of Drishti-GS dataset as training data and maintain the overall training:validation data ratio of approximately 3:1. With such proportion, our 4-fold stratified cross validations gives the average F-Score of 0.945 ± 0.004 and 0.873 ± 0.011 for OD and OC, respectively.

C. Optimized Ellipse Fitting

To further improve the segmentation result of the optimized model, a post-processing algorithm using ellipse fitting will be utilized. The optimization of the ellipse fitting method involves checking each of the detected contour as candidates for the optic cup. The effect of this is demonstrated in an example in Figure 4.

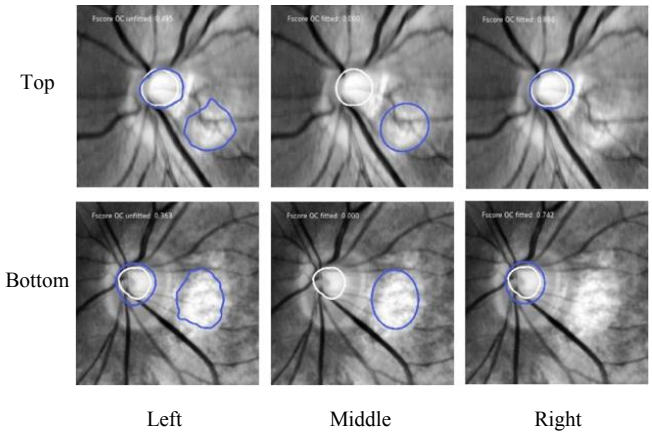

Fig. 4. The distribution of the segmentation results with the best model and optimized ellipse-fitting algorithm is shown in Figure 5, where OC F-Score (right) is shown to contain more variations compared to OD (left), both within and between each data origin. Problems in OC segmentation include subtle boundaries with surrounding OD and intricate blood vessel pattern that sometimes reside within OC.

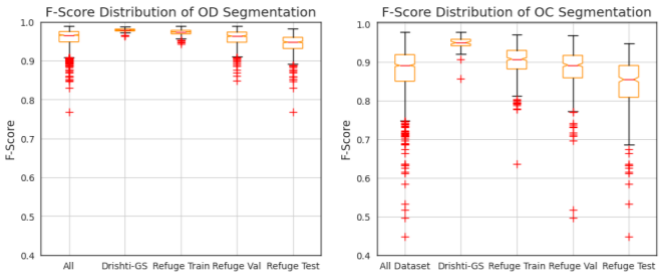

Fig. 5. Distribution of F-Score of optic disc (left) and optic cup (right) segmentation results.

The images affected by the ellipse fitting algorithm are those with OC segmentation results forming more than one area. In the previous algorithm, the contour with the largest area was taken, resulting in the detection of an area other than OC in some images, as shown in Figure 4. The optimization of the ellipse fitting successfully detected the correct OC contour compared to the previous method, thereby enabling the model to detect the location of the OC and avoid images with an OC F-Score close to zero. The distribution of the segmentation results with the best model and optimized ellipse-fitting algorithm is shown in Figure 5.

Table II presents a comparison of our results with those of other studies. All results for the Refuge dataset in Table II are exclusively from the test set, unless stated otherwise. In comparison to the method proposed in [10], our approach yields improvements of 0.04 and 0.13 in the segmentation F-Scores for the optic disc (OD) and optic cup (OC), respectively, on the Refuge validation dataset.

TABLE II. COMPARISON OF SEGMENTATION RESULTS BETWEEN DIFFERENT MODELS

Reference Model	Method	Drishti-GS		REFUGE	
		OD	OC	OD	OC
EfficienNet + U-Net++ [6]	Modified U-Net ++, EfficientNet B4 encoder	0.978	0.938	0.957	0.876
ET-Net [7]	ResNet-50, EGM, WAM	0.975	0.931	0.953	0.891
pOSAL [8]	DeepLabv3+ and adversarial learning	0.974	0.901	0.946	0.875
M-Net [9]	U-net with Multi Scale Input Layer and Side Output, Polar Transform	0.966	0.886	0.936	0.865
Multimap Localization + U-Net[10]	Multimap Localization + U-Net + Ellipse Fitting	0.957[a]	0.929[a]	0.968[a] 0.913[b]	0.913[a] 0.748[b]
Our study	Multimap Localization + U-Net + Ellipse Fitting	0.979[a]	0.948[a]	0.973[a] 0.957[b] 0.942[c]	0.903[a] 0.882[b] 0.843[c]

[a] Calculated with Training Set
[b] calculated with validation Set
[c] calculated with test set

Moreover, despite the simplicity of our method, we achieve differences of less than 0.02 and 0.05 in the F-Scores of OD and OC, respectively, on the Refuge test dataset when compared to other studies. For the Drishti-GS dataset, our F-Scores surpass those reported in other studies. However, it is essential to note that, in our study, the Drishti-GS dataset was solely used as a training set. Other studies did not provide a detailed explanation of their exact utilization of the Drishti-GS dataset.

D. Limitations

While our optimization methodology for optic disc analysis demonstrates significant advancements, it also has certain limitations that warrant consideration. Firstly, the RoI size optimization, aimed at addressing variations in image size across datasets, may not fully account for extreme variations in image resolutions, which could potentially affect segmentation performance on challenging images. To ensure broader applicability, future research could explore more robust approaches to handle extreme variations in image sizes and resolutions.

Secondly, despite enhancing the segmentation method by incorporating an additional dataset and adjusting the training to validation data ratio, the model's performance could still be influenced by dataset-specific biases or limited diversity in the training data. Further increasing the size and diversity of the training dataset, including images with rare pathological cases, may help in improving the model's generalization capabilities and robustness.

Additionally, the optimization methodology has been evaluated on specific datasets, and its performance may vary when applied to other datasets or clinical settings. Therefore, it is crucial to validate and fine-tune the methodology on

diverse and independent datasets to assess its generalizability and applicability in real-world scenarios.

V. Conclusion

In this study, we developed an optimization method for an optic disc and optic cup segmentation algorithm used for glaucoma detection in a previous study. We concluded that the best image preprocessing strategy is to align the RoI at the proposed center of OD with the crop size of 30% times of the original image. This method has been found to improve the segmentation F-Score of OD and OC to 0.917 ± 0.046 and 0.784 ± 0.122 for Refuge Validation compared to the previous study.

Furthermore, we have concluded that the best optimization strategy for OD and OC segmentation is to improve the ellipse fitting algorithm by checking the contour and using a 3:1 ratio for training data division. This strategy has been found to achieve an OD segmentation F-Score of 0.979 ± 0.005 in the training set of Drishti-GS and 0.942 ± 0.026 in REFUGE dataset. The optic cup segmentation has been found to achieve an F-Score of 0.948 ± 0.020 for Drishti-GS training set and an F-Score of 0.843 ± 0.055 for the REFUGE dataset.

References

[1] J. D. Steinmetz *et al.*, "Causes of blindness and vision impairment in 2020 and trends over 30 years, and prevalence of avoidable blindness in relation to VISION 2020: the Right to Sight: an analysis for the Global Burden of Disease Study," *Lancet Glob Health*, vol. 9, no. 2, pp. e144–e160, Feb. 2021, doi: 10.1016/S2214-109X(20)30489-7.

[2] Infodatin, "Situasi Glaukoma di Indonesia," 2019. Accessed: Dec. 08, 2021. [Online]. Available: https://www.pusdatin.kemkes.go.id/resources/download/pusdatin/infodatin/infoDatin_glaukoma_2019.pdf

[3] P. Harasymowycz *et al.*, "Medical Management of Glaucoma in the 21st Century from a Canadian Perspective," *J Ophthalmol*, vol. 2016, pp. 1–22, 2016, doi: 10.1155/2016/6509809.

[4] R. N. Weinreb *et al.*, "Primary open-angle glaucoma," *Nature Reviews Disease Primers*, vol. 2. 2016. doi: 10.1038/nrdp.2016.67.

[5] J. Carrillo, L. Bautista, J. Villamizar, J. Rueda, M. Sanchez, and D. rueda, "Glaucoma Detection Using Fundus Images of The Eye," in *2019 XXII Symposium on Image, Signal Processing and Artificial Vision (STSIVA)*, IEEE, Apr. 2019, pp. 1–4. doi: 10.1109/STSIVA.2019.8730250.

[6] R. Kamble, P. Samanta, and N. Singhal, "Optic Disc, Cup and Fovea Detection from Retinal Images Using U-Net++ with EfficientNet Encoder," 2020, pp. 93–103. doi: 10.1007/978-3-030-63419-3_10.

[7] Z. Zhang, H. Fu, H. Dai, J. Shen, Y. Pang, and L. Shao, "ET-Net: A Generic Edge-aTtention Guidance Network for Medical Image Segmentation," Jul. 2019.

[8] S. Wang, L. Yu, X. Yang, C.-W. Fu, and P.-A. Heng, "Patch-based Output Space Adversarial Learning for Joint Optic Disc and Cup Segmentation," Feb. 2019, doi: 10.1109/TMI.2019.2899910.

[9] H. Fu, J. Cheng, Y. Xu, D. W. K. Wong, J. Liu, and X. Cao, "Joint Optic Disc and Cup Segmentation Based on Multi-Label Deep Network and Polar Transformation," *IEEE Trans Med Imaging*, vol. 37, no. 7, pp. 1597–1605, Jul. 2018, doi: 10.1109/TMI.2018.2791488.

[10] A. N. Almustofa, A. Handayani, and T. L. R. Mengko, "Optic Disc and Optic Cup Segmentation on Retinal Image Based on Multimap Localization and U-Net Convolutional Neural Network," *Journal of Image and Graphics*, vol. 10, no. 3, 2022, doi: 10.18178/joig.10.3.109-115.

[11] J. Sivaswamy, S. R. Krishnadas, G. Datt Joshi, M. Jain, and A. U. Syed Tabish, "Drishti-GS: Retinal image dataset for optic nerve head(ONH) segmentation," in *2014 IEEE 11th International Symposium on Biomedical Imaging (ISBI)*, IEEE, Apr. 2014, pp. 53–56. doi: 10.1109/ISBI.2014.6867807.

[12] J. I. Orlando *et al.*, "REFUGE Challenge: A Unified Framework for Evaluating Automated Methods for Glaucoma Assessment from Fundus Photographs," Oct. 2019, doi: 10.1016/j.media.2019.101570.

[13] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation," May 2015.