RESEARCH ARTICLE

Glaucoma Identification Using Convolutional Neural Networks Ensemble for Optic Disc and Cup Segmentation

SANDRA VIRBUKAITĖ [1], (Student Member, IEEE), **JOLITA BERNATAVIČIENĖ**[1], (Member, IEEE), **AND DAIVA IMBRASIENĖ**[2]

[1]Institute of Data Science and Digital Technologies, Vilnius University, 03101 Vilnius, Lithuania
[2]Institute of Physiology and Pharmacology, Lithuanian University of Health Sciences, 44307 Kaunas, Lithuania

Corresponding author: Sandra Virbukaitė (sandra.virbukaite@mif.vu.lt)

ABSTRACT Glaucoma is one of the diseases that can cause incurable blindness. The first noticeable symptoms appear only when the disease has progressed but an early diagnosis of the disease prevents the severe consequences of disease progression. In this paper, we developed a Convolutional Neural Networks (CNNs) based ensemble for joint optic disc (OD) and optic cup (OC) segmentation using modified Attention U-Net architecture with pre-trained ResNet34, ResNet50, MobileNet, Inceptionv3, DenseNet121 as backbones. The ensemble was trained on a mixed dataset consisting of REFUGE, Drishti-GS, and RIM-ONE r3 (RIM-ONE) datasets of eye fundus images and tested on images of each dataset separately. The most accurate joint OD and OC segmentation is achieved using an ensemble consisting of modified Attention U-Net with pre-trained ResNet34, Inceptionv3, and DenseNet121 as backbones and majority voting for final prediction. The highest Dice of 0.961 for OD and 0.894 for OC is achieved on the REFUGE test dataset, 0.974 for OD and 0.916 for OC on the Drishti-GS test dataset, and 0.978 for OD and 0.902 for OC on the RIM-ONE test dataset. The highest Intersection over Union of 0.925 for OD and 0.808 for OC is achieved on the REFUGE test dataset, 0.950 for OD and 0.845 for OC on the Drishti-GS test dataset, and 0.957 for OD and 0.822 for OC on the RIM-ONE test dataset. Using the segmentation results, the cup-to-disc ratio (CDR) has been calculated to classify eye fundus images into mild-stage, moderate-stage, severe-stage glaucoma, and non-glaucoma cases.

INDEX TERMS Ensemble networks, glaucoma, machine learning (ML), segmentation, voting methods.

I. INTRODUCTION

Vision plays a major role in every aspect of human life. Unfortunately, various eye diseases can cause vision impairment and even blindness. Glaucoma is one of those eye diseases that damages the eye's optic nerve and it is ranked as the second-leading cause of blindness and the fourth-leading cause of moderate and severe vision impairment [1]. According to the World Health Organization (WHO), the number of people with glaucoma worldwide will increase to 111.8 million in 2040 [2]. The diagnosis

The associate editor coordinating the review of this manuscript and approving it for publication was Sangsoon Lim.

of glaucoma is based on three main indicators: increased intraocular pressure, changes in the visual field, and the size of the optic nerve disc excavation. Visual field assessment is time-consuming and measuring intraocular pressure can be also not enough to diagnose glaucoma, as visual impairment may be present in the absence of an increase in intraocular pressure [3]. The mathematical assessment of optic nerve disc excavation is of great practical benefit to ophthalmologists, as fundus photographs are taken for many patients. Measuring the excavation would allow a more accurate evaluation of atrophic changes in the optic nerve in the early stage of optic nerve damage. The diagnosis of optic nerve atrophy is important not only for those with glaucoma but also in cases

of multiple sclerosis, brain or optic nerve tumors, and other pathologies leading to optic nerve atrophy. Early diagnosis of optic nerve atrophy and timely treatment would help manage the condition in its early stages and preserve better visual functions. Therefore, early detection of glaucoma is crucial.

Clinically, glaucoma examination is based on visual field evaluation, intraocular pressure examination, and cup-to-disc ratio (CDR) measurement [4], [5]. CDR is one of the most significant techniques for automated glaucoma detection. CDR is calculated as the ratio of the vertical OC diameter (VCD) to the vertical OD diameter (VDD). Based on the CDR value different stages of glaucoma are distinguished [6]:

- Early or mild glaucoma: CDR in (0.3–0.4]
- Moderate glaucoma: CDR in (0.4–0.7)
- Severe glaucoma: CDR above 0.7.

A healthy eye has a CDR of 0.3.

An accurate measurement of CDR relies on the precise segmentation of the optic disc (OD) and optic cup (OC), which in most cases is obtained manually by ophthalmologists. Since manual contouring of the OD and OC borders is a complex and time-consuming task, various methods, especially deep neural networks, have been proposed for automatic segmentation of OD and OC.

Although automatic segmentation methods have shown great potential in glaucoma detection in general, the identification of mild-stage glaucoma remains a challenging task [7]. The automated segmentation algorithms cannot distinguish the boundaries of the optic cup due to heavy overlap and weak contrast between the optic cup and neuroretinal rim regions. In addition, most previous studies based on deep learning techniques have focused on identifying the image of the eye fundus in glaucomatous and nonglaucomatous cases [8], [9], [10], [11] or segmentation of the optic disc and optic cup [12], [13], [14] without going into the stage of vision impairment. Another notable point is that most recent works still use eye fundus images from the same dataset to train and test deep neural networks [15], [16], [17].

To overcome the limitations above, this paper presents the applied approach as follows:

- A mixed training data strategy by combining eye fundus images of different datasets Drishti-GS, REFUGE, and RIM-ONE is used to increase image diversity.
- The joint OD and OC segmentation approach is used to optimize computational power.
- An ensemble of five deep neural networks is applied to improve OD and OC segmentation for further CDR calculation.
- According to the calculated CDR the eye fundus images are classified into nonglaucoma cases and different stages of glaucoma, namely mild glaucoma, moderate glaucoma and severe glaucoma.

The rest of this paper is divided into the following sections: Section II presents the related work of the applied CNNs ensemble; Section III describes the methodology used in our experiment; Section IV describes the experiment; Section V

provides the experimental results; Section VI concludes our paper.

II. RELATED WORK

During the past few years, ensemble learning-based CNNs architectures have been carried out in automatic glaucoma assessment systems development based on fundus image analysis.

Kim et al. [8] proposed two ensemble models based on three Fully Convolutional Networks (FCN) with a modified U-Net structure to segment OD and OC where a different region of interest (ROI) was used as input for each FCN. In each ensemble model, the final results were estimated by merging the results of three FCNs using an averaging operator. The raw ROIs were used as input for the OD segmentation model and masked ROIs were used as inputs for the OC segmentation model. The RIGA and the REFUGE datasets were used for the training and evaluation of the proposed method. The proposed method achieved an IoU of 0.930 and a Dice of 0.964 in OD segmentation and an IoU of 0.810 and a Dice of 0.892 in OC segmentation accordingly.

Civit-Masot et al. [9] used an ensemble approach to predict glaucoma on a combined dataset consisting of RIM-ONE and Drishti-GS datasets in such a way that OD and OC were segmented using a generalized U-Net to calculate the CDR, and Random sample consensus (RANSAC) was used to find out if the predicted shapes are similar to ellipse. The transfer learning on MobileNet V2, pre-trained with weights from the ImageNet 1K challenge, was used for the prediction of glaucoma. The results were blended to provide a likelihood score for glaucoma. The applied approach achieved the Dice of 0.920 and 0.840 for OD and OC segmentation, respectively, on the RIM-ONE dataset, and the Dice of 0.930 and 0.890 for OD and OC segmentation, respectively, on the Drishti-GS dataset.

Zilly et al. [18] presented a general framework for retinal image segmentation using CNN architectures based on ensemble learning. A deep CNN network was trained on numerous patches from the same Drishti-GS dataset. An entropy sampling technique was used for information point reduction and allowed to reduce the computational complexity. The proposed approach achieved a Dice of 0.970 for OD segmentation and a Dice of 0.870 for OC segmentation in the Drishti-GS dataset.

Ali et al. [12] presented an OD segmentation system based on an ensemble of ten deep learning-based semantic segmentation models such as U-Net, Gated Skip Connections (GSCs), DoubleU-Net, DeepLabV3+, CGNet, ERFNet, SegNet, ESNet, LinkNet, and SQNet. For the aggregation step, the Ordered Weighted Average operator has been used. The aggregation has been applied to each pixel of the input image. A threshold of 0.5 was applied to the result and the class with the maximum activation has been taken as a label. The best results for segmenting the OD in fundus images collected from the Hospital Sant Joan de Reus have

been obtained with the ensemble of models GSCs, DoubleU-Net, and DeepLabV3+ achieving an IoU of 0.954, Dice of 0.951, a precision higher than 0.960, and a recall higher than 0.930.

Rehman et al. [10] proposed a two-stage glaucoma classification scheme based on four pre-trained deep convolutional neural networks (deep CNNs) AlexNet, NasNet-Large, InceptionResNetV2 and InceptionV3. The usage of transfer learning where the weights in pre-trained networks are used as the starting point for the training process helped to reduce the training time. These four Deep CNNs were tested using extracted Optic Nerve Head (ONH) from three publicly available datasets ACRIMA, ORIGA-Light, and RIM-ONE, and private datasets collected from local hospitals. The classification accuracy has been improved by combining these four Deep CNNs into one classifier where the final decision has been made by the five voting techniques, i.e., Proportional Voting (PV), Majority Voting (MV), Averaging (AV), Accuracy/Score based Weighted Averaging (ASWA), and Accuracy based Weighted Voting (AWV). In the case of the ACRIMA dataset, the accuracy of 0.995 of AlexNet and NasNet-Large has been improved to 0.996 by the ensemble with AWV. For the ORIGA-Light, the accuracy has increased from 0.879 to 0.883 with ASWA.

Kurilová et al. [11] applied an ensemble method composed of deep learning models VGG-16, MobileNet, and ResNet-50 using hard voting and average voting to classify OD using fundus images from the REFUGE dataset. The models were monitored using binary accuracy, precision, recall, Area Under the Curve (AUC), true positives, true negatives, false positives, and false negatives. The best accuracy of 0.980 and AUC of 0.880 were achieved using the average voting method.

The detail studio prepared by Elangovan and Nath [19] presented a deep ensemble model based on stacking ensemble technique implementing such Deep CNNs as Xception, Inceptionv3, Densenet-201, Mobilenet-v2, Efficientnet-b0, VGG-16, VGG-19, Googlenet, Alexnet, Resnet-18, Resnet-50, Resnet-101, Squeezenet, and using support vector machine (SVM) for the final classification of glaucoma and normal images. With the suggested approach the classification accuracy of 0.996, 0.995, 0.934, 0.913, 0.796 in LAG-R, ACRIMA-R, Drishti-GS1-R, RIM-ONE2-R, ORIGA-R has been achieved.

It is observed from the literature above that lately proposed methods employ CNNs ensemble for various glaucoma detection tasks. In [8] with the help of the proposed CNN ensemble method, unexpected segmentation errors have been removed and a post-processing to fine-tune the results was no longer needed. In [9] an ensemble assisted in the development of a diagnostic aid tool for glaucoma detection by analyzing complete images of the eye fundus. The advantages of the proposed CNN ensemble in [18] are that the computational effort has been reduced while providing better performance than the simple uniform sampling approach.

III. METHODOLOGY

This section presents the research methodology, applied Convolutional Neural Networks, voting methods, and proposed ensemble.

CNNs and their variants have shown great performance in various glaucoma identification tasks, namely classification and segmentation [8], [20]. In this paper, for the development of an ensemble, the Attention U-Net architecture [21] has been chosen as a base for decoding layers due to the performance results in our previous work [22]. To improve the OD segmentation and especially OC segmentation and to save computational resources at the same time, pre-trained CNN models namely ResNet-34 and ResNet-50 [23], MobileNet [24], Inception-v3 [25], and DenseNet [26] have been chosen as encoding layers due their characteristics. ResNet with its deep layers and residual connections excels in learning complex images. MobilNet requires fewer parameters while maintaining reasonable performance. Inception-v3 is trained to learn the visual representations and capture both local and global features. DenseNet allows each layer to use the features of all previous layers, so the gradient flow is optimized during training and the network can acquire knowledge more effectively. The efficiency of these models was demonstrated in [19] through solving classification tasks. The final prediction in our proposed approach is given by the five most common [10], [27] voting techniques such as Majority voting (MV), Averaging voting (AV), Weighted average voting (WAV), Unanimous voting method (UV), and Max voting. The following subsections describe each CNN and voting method in more detail.

A. CONVOLUTIONAL NEURAL NETWORK

Attention U-Net [21] is an extended U-Net [28] with incorporated attention mechanisms that help the network to learn more fine-grained features by focusing on the important regions of the input image. The network consists of a contraction path at the left and an expansion path at the right as the original U-Net. A contraction path with a series of convolution layers and max pooling is intended for the extraction of local features. The expansion path with a series of up-sampling and convolution layers is appointed for global features. Also like in the original U-Net, there is the concatenation of feature maps using a skip connection but the Attention U-Net differs by having an attention gate at each level of the skip connection. With the help of the attention gate, the feature activations in irrelevant regions are repressed which improves the accuracy and sensitivity of the model for dense label predictions.

Resnet [23] is a deep-learning model where the weight layers learn residual functions by the layer inputs, instead of learning unreferenced functions. Rather than relying on each of the multiple stacked layers to directly correspond to the desired underlying mapping, residual networks allow these layers to apply a residual mapping. The residual blocks are

stacked on top of each other to form a network. ResNet-34 and ResNet-50 have thirty-four and fifty layers accordingly using these blocks.

MobileNet [24] is based on a simplified architecture that uses depthwise separable convolutions to create lightweight deep neural networks. The Depth-wise separable convolution consists of the depth-wise convolution layer which is used to filter the input channels and the point-wise convolution layer which is used to merge them to create a new feature. With the usage of the depth-wise convolutions, a single filter is applied to each input channel. Since the depthwise convolution filters the input channel only rather than combining them to produce new features, the pointwise convolution layer is added to calculate a linear combination of the output of depthwise convolution using a 1×1 convolution.

Inception-v3 [25] is a modification of the previous structure of a deep neural network Inception-v1 [29] and Inception-v2 [30]. The main idea of inception network architecture is to process and extract the various feature maps concurrently in parallel. Whereas the inception module extracts a different piece of information from every convolution or pooling operation the different features are extracted from each operation. Combining all extracted information into a single feature map helps to increase the accuracy of the model as it focuses on multiple features simultaneously. Inception-v3 factorizing the larger convolutions into smaller convolutions the number of parameters involved in a network was reduced which helped to reduce the computational efficiency. The added auxiliary classifiers helped to improve the accuracy of the network. The activation dimension of the network filters was expanded to reduce the grid size efficiently.

DenseNet [26] is a Densely Connected Convolutional Network whose architecture consists of transition layers and dense blocks. Each convolutional layer is linked to each other layer in a feed-forward fashion within the block. This maximum information flow between layers is created by the connection between the output of each layer and the input of the next layer. The usage of the transition layer reduces the spatial dimensionality of the feature maps. DenseNet-121 is a version of DenseNet with 121 layers.

B. VOTING METHODS

Majority voting (MV) - each model makes a prediction (vote) for each test instance and the final output prediction is the one that receives more than half of the votes [31].

Averaging voting (AV) [32] is a method where the final prediction is made by averaging the extracted predictions from multiple models. Average prediction is calculated using the arithmetic mean, which is the sum of the predictions divided by the total number of predictions.

$$y^* = \underset{i}{argmax} \frac{1}{m} \sum_{j=1}^{m} p_{ij}. \tag{1}$$

Refer to (1), m is the number of data points to be averaged, and p_{ij} is the probability of the i class label of the j classifier.

Weighted average voting (WAV) [33] is a method where the initial learners are given different weights to indicate the importance of each model in the prediction. The weighted average for each class 0 or class 1 is calculated by applying multiplication for each prediction by the weight of the classifiers to provide a weighted sum and then dividing the result by the sum of the weights of the classifier. Class 0 represents a background and class 1 represents the area of the OD or OC in case of segmentation of the OD or OC accordingly.

$$y^* = \frac{\sum_{j=1}^{m} w_j x_j}{\sum_{j=1}^{m} w_j}. \tag{2}$$

Refer to (2), w is the weighted average and m is the number of data points to be averaged.

Unanimous voting method (UV) [34] is similar to majority voting but in a unanimous voting case is no requirement of half the votes. Here, all segmentation models are required to predict the same value, otherwise, no prediction is made.

Max voting is one of the simplest ways of combining predictions from multiple machine learning algorithms. In max voting, each base model makes a prediction and votes for each sample. Only the sample with the highest votes is included in the final prediction.

C. PROPOSED ENSEMBLE

Ensemble learning is a meta-approach aiming to improve predictive performance by combining the predictions from multiple models instead of using a single model. In this paper, we developed five deep-learning models for joint OD and OC segmentation based on a modified Attention U-Net. The schema of the proposed method, including mixed training data of different datasets preparation by applying data preprocessing, data augmentation, training of different CNNs modifications, applied ensemble methods, and evaluation strategies are illustrated in Fig. 1. The pre-trained networks, namely ResNet34, ResNet50, MobileNet, Inceptionv3, and DenseNet121 have been used instead of the original encoder of the Attention U-Net as backbones separately. All backbones have weights trained on the 2012 ILSVRC ImageNet dataset for better and faster convergence. For clarity, each modification of the Attention U-Net with the different pre-trained network, namely ResNet34, ResNet50, MobileNet, Inceptionv3, and DenseNet121 as a backbone has been named Model-1, Model-2, Model-3, Model-4, and Model-5 accordingly. The top 3 models that achieved the highest value of the Dice coefficient have been chosen to compile an ensemble model. We applied five different voting methods, namely majority voting, weighted average voting, unanimous voting, max voting, and averaging voting to obtain the final OD and OC segmentation prediction provided by the ensemble model.

IV. EXPERIMENT

The three publicly available retinal fundus image datasets REFUGE [35], RIM-ONE [36], and Drishti-GS [37] have
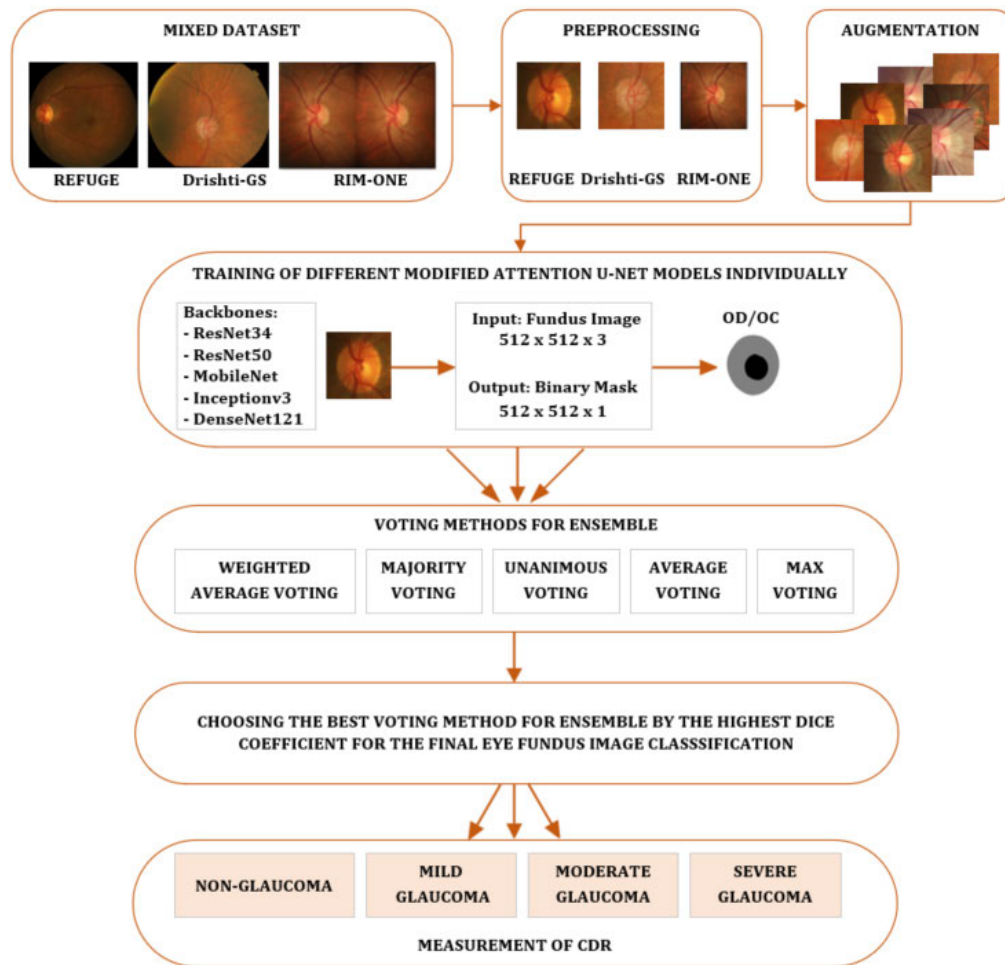
FIGURE 1. The schema of the proposed method.

been preprocessed and used to construct the mixed training dataset to train the original Attention U-NET and five different modifications of Attention U-Net with a different pre-trained networks ResNet34, ResNet50, MobileNet, Inceptionv3, and DenseNet121 as a backbone for joint OD and OC segmentation. Each modification has been tested on three separate test datasets REFUGE, RIM-ONE, and Drishti-GS by evaluating the Dice and IoU. The obtained outputs of segmented OD and OC have been used to calculate the CDR for eye fundus image classification into non-glaucoma, mild glaucoma, moderate glaucoma, and severe glaucoma cases. The CDR was calculated for the output of segmented OD and OC obtained using the ensemble method and compared with the calculated CDR of ground truth test images according to glaucoma stages and non-glaucoma cases.

A. DATASET DESCRIPTION AND PREPROCESSING

The high-quality retinal fundus images used for this work were collected from different publicly available datasets such as REFUGE [35], RIM-ONE [36], and Drishti-GS [37].

REFUGE [35] is a publicly available dataset that was created as part of the Retinal Fundus Glaucoma Challenge by releasing 1200 color fundus photographs stored in JPEG format of female Chinese patients with no information about their age. These images were obtained from clinical studies and several hospitals by selecting only high-quality pictures and anonymization of any personal data. The OD and the OC were manually annotated by seven independent glaucoma specialists with expertise from 5 to 10 years in the field. The dataset is split equally into training, validation, and testing subsets. The training dataset consists of 40 glaucomatous and 360 nonglaucomatous images. The validation and testing datasets have no labeling of glaucomatous and nonglaucomatous images. The eye fundus images of the training set have been captured by a Zeiss Visucam 500 fundus camera of resolution 2124 × 2056 px. The 400 images of the testing dataset and 400 images of the validation dataset have been captured by a Canon CR-2 camera of size 1634 × 1634 px.

Retinal Image Database for Optic Nerve Evaluation (RIM-ONE) [36] is a public dataset consisting of color

stereo retinal fundus images stored in JPG format with corresponding OD and OC ground truth stored in PNG format of 85 nonglaucomatous and 74 glaucomatous eyes. Manual segmentation of the OD and OC has been done by 2 experts in ophthalmology. The pictures were taken by HUC with a non-mydriatic Kowa WX 3D stereo fundus camera with a resolution of 2144 × 1424 px in various Spanish hospitals but there is no information about the age and gender of the patients. For our experiments, the left part of a stereo eye fundus image has been used as it contains the ground truth label.

Drishti-GS [37] is a publicly available dataset containing 101 color images stored in PNG format dividing them into 50 training and 51 testing images. The training dataset contains 32 glaucomatous images and 18 nonglaucomatous images. The test dataset contains 38 glaucomatous images and 13 nonglaucomatous images. The images were collected at Aravind Eye Hospital of females and males 40-80 years of age. All the images were segmented manually by 4 eye experts with clinical experience of 3, 5, 9, and 20 years. All images were taken centered on the OD with a Field-Of-View (FOV) of 30 degrees with a resolution of 2045 × 1752 px.

Analyzing these datasets, it was observed that they have been created by different institutions using different fundus cameras and labeled glaucomatous / nonglaucomatous by experts with clinical experience in this field from 3 to 10, and 20 years. The number of experts involved in the preparation of REFUGE, Drishti-GS, and RIM-ONE datasets are 7, 4, and 2 respectively.

There are various image preprocessing techniques such as Gaussian noise or salt-and-pepper noise used more for low-contrast images [3], variational histogram equalization that helps to improve the quality of image [38], image normalization [4], [20], the region of interest (ROI) extraction to reduce the impact of non-relevant regions [16], [39], image resizing to reduce the computation time [22], [40].

In the image preprocessing stage of this paper, the image normalization between 0 and 1, the region of interest (ROI) extraction, and image resizing have been applied but in future work, the other preprocessing techniques will be investigated and their impact on segmentation results will be evaluated. The region of interest (ROI) was extracted by cropping the central area of the OD. The size of cropped ROIs differed depending on the sizes of the original fundus image, including from 408 × 408 to 616 × 616 px in the REFUGE dataset, from 674 × 674 to 1060 × 1060 px in the Drishti-GS dataset, from 456 × 456 to 890 × 890 px in RIM-ONE dataset. The images have been resized to sizes of 512 × 512 px by applying the bicubic interpolation [22]. The preprocessed eye fundus images have been used to train, validate, and test the different CNNs.

B. IMPLEMENTATION

Combining preprocessed eye fundus images of three different datasets REFUGE, Drishti-GS, and RIM-ONE into mixed dataset the original Attention U-Net and each modification of the Attention U-Net with a different pre-trained network, namely ResNet34, ResNet50, MobileNet, Inceptionv3, and DenseNet121 as a backbone named Model-1, Model-2, Model-3, Model-4, and Model-5 accordingly was trained and validated on a single GPU machine [45] with 1 TB of RAM in the Keras and TensorFlow frameworks by using the Adam optimizer and Dice loss, and applying an early stopping technique to reduce unnecessary training time. Applying the KerasTuner framework, the parameters, i.e. batch size, dropout rate and learning rate for each CNN modification have been searched separately. Various image augmentation techniques, such as image rotation by an angle of rotation from 0° to 45°, zooming by 20%, and horizontal and vertical flipping have been performed during the training by dedicating 20% of images for validation to increase the diversity and prevent overfitting. All trained CNNs have been tested on separate test datasets of REFUGE, Drishti-GS, and RIM-ONE consisting of 50 eye fundus images with different glaucoma stages and non-glaucoma.

C. EVALUATING METRICS

The performance of OD and OC segmentation was evaluated by the Dice coefficient (Dice) and Intersection over Union (IoU). Dice is used in most cases [13], [14], [41], [46] to describe the similarity between the two images.

$$Dice = \frac{2\left|S \bigcap L\right|}{\left|S\right| + \left|L\right|}. \tag{3}$$

Refer to (3), S – the result based on segmentation, L – the ground truth label.

IoU represents the overlapping ratio between the segmentation results and the ground truth mask [8], [12], [42].

$$IoU = \frac{\left|S \bigcap L\right|}{\left|S \bigcup L\right|}. \tag{4}$$

Refer to (4), S – the result based on segmentation, L – the ground truth label.

The CDR is calculated as the ratio of the vertical cup diameter (VCD) to the vertical disc diameter (VDD) [7].

$$CDR = \frac{vertical\ cup\ diameter}{vertical\ disc\ diameter}. \tag{5}$$

V. RESULTS

Original Attention U-NET and five modifications of Attention U-Net with different pre-trained networks ResNet34, ResNet50, MobileNet, Inceptionv3, and DenseNet121 as a backbone named Model-1, Model-2, Model-3, Model-4, and Model-5 accordingly have been trained on a mixed dataset and tested on testing data of REFUGE, Drishti-GS, and RIM-ONE separately. The results of OD and OC segmentation obtained by Dice and IOU are presented in Table 1. The three models Model-1, Model-4, and Model-5 that achieved the highest OD and OC segmentation results by Dice and IoU have been chosen to compile an ensemble model.

The five different voting methods such as majority voting, weighted average voting, unanimous voting, and averaging

TABLE 1. OD and OC segmentation results obtained by different models.

| Model	Backbone	Training time per epoch	Test Dataset	Optic Disc Dice	IoU	Optic Cup Dice	IoU
Attention U-Net (original)	-	1 s	REFUGE	0.944	0.895	0.864	0.761
			Drishti-GS	0.961	0.925	0.897	0.814
			RIM-ONE	0.912	0.838	0.815	0.688
Model-1	ResNet34	370 ms	REFUGE	**0.958**	**0.919**	**0.887**	**0.797**
			Drishti-GS	**0.972**	**0.945**	**0.918**	**0.848**
			RIM-ONE	**0.972**	**0.946**	**0.893**	**0.807**
Model-2	ResNet50	396 ms	REFUGE	0.953	0.911	0.886	0.795
			Drishti-GS	0.972	0.946	0.915	0.843
			RIM-ONE	0.957	0.917	0.891	0.804
Model-3	MobileNet	370 ms	REFUGE	0.947	0.899	0.883	0.791
			Drishti-GS	0.964	0.931	0.905	0.826
			RIM-ONE	0.969	0.940	0.882	0.789
Model-4	Inceptionv3	470 ms	REFUGE	**0.952**	**0.909**	**0.883**	**0.791**
			Drishti-GS	**0.970**	**0.943**	**0.912**	**0.839**
			RIM-ONE	**0.974**	**0.950**	**0.892**	**0.805**
Model-5	DenseNet121	404 ms	REFUGE	**0.958**	**0.920**	**0.889**	**0.800**
			Drishti-GS	**0.973**	**0.948**	**0.912**	**0.838**
			RIM-ONE	**0.974**	**0.949**	**0.898**	**0.815**
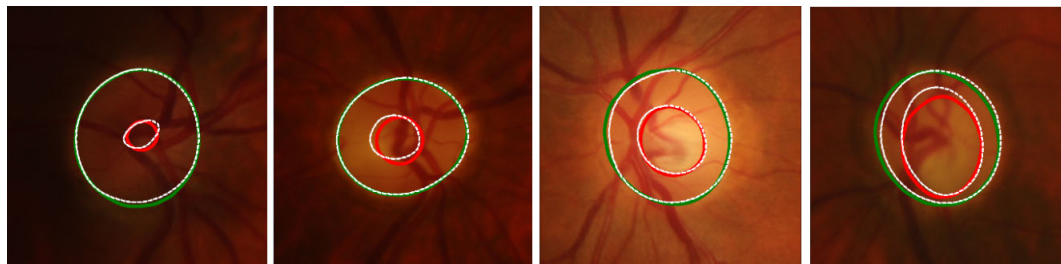
FIGURE 2. The results of the best ensemble method. Dashed circles indicate ground truths of OD and OC. The green circle indicates segmented OD, and the red circle - segmented OC. From the left: non-glaucoma, mild-stage glaucoma, moderate-stage glaucoma, and severe-stage glaucoma.

TABLE 2. Results of five voting methods with test instance.

Voting technique	Test Dataset	Optic Disc Dice	IoU	Optic Cup Dice	IoU
Weighted average	REFUGE	0.956	0.916	0.886	0.795
	Drishti-GS	0.972	0.945	0.915	0.843
	RIM-ONE	0.975	0.950	0.895	0.810
Majority	REFUGE	**0.961**	**0.925**	**0.894**	**0.808**
	Drishti-GS	**0.974**	**0.950**	**0.916**	**0.845**
	RIM-ONE	**0.978**	**0.957**	**0.902**	**0.822**
Unanimous	REFUGE	0.952	0.908	0.877	0.781
	Drishti-GS	0.968	0.938	0.901	0.819
	RIM-ONE	0.973	0.948	0.886	0.796
Average	REFUGE	0.956	0.916	0.887	0.796
	Drishti-GS	0.972	0.945	0.914	0.842
	RIM-ONE	0.974	0.949	0.894	0.809
Max	REFUGE	0.957	0.917	0.890	0.801
	Drishti-GS	0.973	0.948	0.922	0.856
	RIM-ONE	0.971	0.943	0.894	0.809

TABLE 3. Summary of OD and OC segmentation.

Model	Test Dataset	Optic Disc Dice	IoU	Optic Cup Dice	IoU
Attention U-Net (original)	REFUGE	0.944	0.895	0.864	0.761
	Drishti-GS	0.961	0.925	0.897	0.814
	RIM-ONE	0.912	0.838	0.815	0.688
Model-5	REFUGE	0.958	0.920	0.889	0.800
	Drishti-GS	0.973	0.948	0.912	0.838
	RIM-ONE	0.974	0.949	0.898	0.815
Proposed Ensemble (Majority voting method)	REFUGE	**0.961**	**0.925**	**0.894**	**0.808**
	Drishti-GS	**0.974**	**0.950**	**0.916**	**0.845**
	RIM-ONE	**0.978**	**0.957**	**0.902**	**0.822**

voting have been applied to obtain the final OD and OC segmentation prediction provided by the ensemble model on each test dataset REFUGE, Drishti-GS, and RIM-ONE separately. The results are presented in Table 2 which indicates that the majority voting method for the predictions of the model was performed with the highest values of Dice and IoU to conclude the final prediction.

Table 3 presents the summary of OD and OC segmentation results by Dice and IoU of trained models, namely original

Attention U-Net, modified Attention U-Net with pre-trained DenseNet121 as a backbone, and proposed ensemble with a majority voting method. In comparison with the single original Attention U-Net, the proposed ensemble method has increased the OD segmentation by Dice by 2%, 2%, and 7% on REFUGE, Drishti-GS, and RIM-ONE respectively. The OC segmentation results by Dice have been increased by 3%, 2%, and 9% on REFUGE, Drishti-GS, and RIM-ONE respectively. In comparison with the single best Attention U-Net modification holding pre-trained DenseNet121 as a backbone, the proposed ensemble method has increased the OD segmentation by Dice by 0.3%, 0.1%, and 0.4% on REFUGE, Drishti-GS, and RIM-ONE accordingly. The OC

TABLE 4. Comparison of the proposed method and different existing methods in OD and OC segmentation.

Method	Model	Training dataset	Test dataset	Optic Disc Dice	Optic Disc IoU	Optic Cup Dice	Optic Cup IoU
X. Zhang et al. [43] (2022)	CAE-BMAL	REFUGE	REFUGE	0.963	-	0.879	-
			Drishti-GS	0.962	-	0.857	-
			RIM-ONE3	0.898	-	0.791	-
A. Zhao et al. [15] (2023)	Paired-Box RPN	REFUGE	REFUGE	0.959	-	0.903	-
		ORIGA	ORIGA	0.963	-	0.893	-
Z. Li et al. [16] (2023)	TUNet	REFUGE	REGUGE	0.961	-	0.901	-
		Drishti-GS	Drishti-GS	0.973	-	0.903	-
		RIM-ONE	RIM-ONE	0.969	-	0.862	-
R. Bhattacharya et al. [17] (2023)	PY-Net	REFUGE	REFUGE	0.965	0.932	0.885	-
		Drishti-GS	Drishti-GS	0.971	0.944	0.876	-
		RIM-ONE	RIM-ONE	0.961	0.926	0.874	-
L. Jiang et al. [39] (2023)	BEAC-Net	Drishti-GS	Drishti-GS	0.861	0.839	0.809	0.763
		RIM-ONE	RIME-ONE	0.858	0.839	0.733	0.663
			66 Vision-Tech	0.827	0.814	0.806	0.789
Y. Yi et al. [44] (2023)	C2FTFNet	REFUGE	REFUGE	0.969	-	0.908	-
		Drishti-GS	Drishti-GS	0.976	0.954	0.920	0.854
		DRIONS-DB	DRIONS-DB	0.968	0.936	-	-
W. Zhou et al. [40] (2023)	EARDS	REFUGE	REFUGE	0.955	0.915	0.887	0.802
		Drishti-GS	Drishti-GS	0.974	0.950	0.916	0.849
J. Wang et al. [20] (2023)	EE-Unet	REFUGE	Drishti-GS	0.962	0.885	0.923	0.815
		GAMMA	RIM-ONE	0.956	0.880	0.864	0.762
Proposed	Ensemble	Mixed of REFUGE, Drishti-GS, and RIM-ONE	REFUGE	0.961	0.925	0.894	0.808
			Drishti-GS	0.974	0.950	0.916	0.845
			RIM-ONE	0.978	0.957	0.902	0.822

TABLE 5. % of truth CDR and CDR calculated using segmented OD and OC by ensemble on each test dataset separately.

CDR by glaucoma stages	Test dataset	Amount of images	% of correctly classified CDR using ensemble
(0–0.3] Non-glaucoma	REFUGE	4	50
	Drishti-GS	–	–
	RIM-ONE	7	86
(0.3–0.4] Mild glaucoma	REFUGE	7	57
	Drishti-GS	–	–
	RIM-ONE	7	86
(0.4–0.7] Moderate glaucoma	REFUGE	34	92
	Drishti-GS	10	90
	RIM-ONE	23	87
Above 0.7 Severe glaucoma	REFUGE	5	100
	Drishti-GS	40	95
	RIM-ONE	13	92

segmentation results by Dice have been increased by 0.5%, 0.4%, and 0.7% on REFUGE, Drishti-GS, and RIM-ONE respectively.

We compared the obtained OD and OC segmentation results by our proposed ensemble approach with the results obtained by the other authors (Table 4). The proposed ensemble approach when a mixed dataset was used to train different deep learning models outperformed the results of other authors when models were trained and tested on a single dataset.

Table 5 presents the percentage of correctly classified outputs obtained using the ensemble method of segmented OD and OC for test images into four different glaucoma stages and non-glaucoma cases according to calculated CDR in comparison with their ground truth CDR.

The visual samples obtained using the proposed ensemble method are presented in Fig. 2. From the left, it visualizes non-glaucoma, mild glaucoma, moderate glaucoma, and

severe glaucoma cases. The green and red circles indicate the segmented OD and OC accordingly, the white circles indicate their ground truths.

VI. CONCLUSION AND DISCUSSION

In this paper, we have proposed an ensemble based on a modified Attention U-Net with pre-trained CNNs ResNet34, ResNet50, MobileNet, Inceptionv3, DenseNet121 as a backbone for mild glaucoma, moderate glaucoma, severe glaucoma and non-glaucoma classification using eye fundus images. The proposed method was trained on a mixed dataset consisting of images from REFUGE, Drishti-GS, and RIM-ONE and tested on test images of each dataset separately. The five voting techniques, namely weighted average voting, majority voting, unanimous voting, average voting, and max voting have been applied to obtain the final joint OD and OC segmentation results. The performance of the proposed ensemble method has been evaluated by Dice and IoU, and the calculated CDR has been used for eye image classification into different glaucoma stages and non-glaucoma cases. Our experimental results conclude that:

- OD and OC segmentation results are more accurate when applying an ensemble of CNNs than a single model. Applying the proposed CNNs ensemble method the OD segmentation by Dice has been increased by 2%, 2%, and 7% on REFUGE, Drishti-GS, and RIM-ONE respectively. The OC segmentation results by Dice have been increased by 3%, 2%, and 9% on REFUGE, Drishti-GS, and RIM-ONE accordingly.

- Comparing the applied five voting methods, the most accurate OD and OC segmentation results are obtained when the majority voting method is applied to combine the predictions of three different models.

- Severe-stage glaucoma cases are classified at 92%, 95%, and 100% in the RIM-ONE, Drishti-GS, and REFUGE datasets.
- Moderate-stage glaucoma cases are classified at 87%, 90%, and 92% in the RIM-ONE, Drishti-GS, and REFUGE datasets.
- Mild-stage glaucoma cases are classified at 86%, and 57% in the RIM-ONE, and REFUGE datasets. In cases where there is a mistake in classifying cases of mild-stage glaucoma, these are classified as moderate-stage glaucoma.
- Non-glaucoma cases are classified at 86%, and 50% in the RIM-ONE, and REFUGE datasets. In cases where there is a mistake in classifying cases of non-glaucoma, these are classified as mild-stage glaucoma.

The mathematical assessment of the optic nerve head and disc excavation will have significant practical importance for ophthalmologists not only in the early diagnosis of glaucoma but also for the diagnosis of optic nerve atrophy due to other reasons.

The proposed approach was applied to publicly available high-contrast images but in future work, it will be applied to private datasets, once its preparation is done, containing low-contrast images taken by different handheld eye fundus cameras and in the development of the tools for various stages of glaucoma detection.

Analyzing datasets it was noticed that in some datasets the fundus images were labeled by seven experts, and in others by two or four experts, therefore future investigation would be needed to evaluate the impact of CNN's segmentation results when the images are labeled by different experts.

Analyzing the obtained results and discussing the ensemble error cases with the independent ophthalmologists and the ophthalmologist from our research team, it was noticed, that image labeling quality highly impacts the segmentation results of CNNs. According to experts, Dice and IoU metrics are not sufficient for the quality of results evaluation and the ophthalmologist should be involved in the evaluation process of the OD and OC segmentation results obtained by the CNNs.

ACKNOWLEDGMENT

REFERENCES
[1] GBD 2019 Blindness and Vision Impairment Collaborators, "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 Global Health*, vol. 9, no. 2, pp. e144–e160, Feb. 2021, doi: 10.1016/s2214-109x(20)30489-7.

[2] Y.-C. Tham, X. Li, T. Y. Wong, H. A. Quigley, T. Aung, and C.-Y. Cheng, "Global prevalence of glaucoma and projections of glaucoma burden through 2040," *Ophthalmology*, vol. 121, no. 11, pp. 2081–2090, Nov. 2014, doi: 10.1016/j.ophtha.2014.05.013.

[3] P. Elangovan and M. K. Nath, "Glaucoma assessment from color fundus images using convolutional neural network," *Int. J. Imag. Syst. Technol.*, vol. 31, no. 2, pp. 955–971, Jun. 2021, doi: 10.1002/ima.22494.

[4] A. Sánchez-Morales, J. Morales-Sánchez, O. Kovalyk, R. Verdú-Monedero, and J.-L. Sancho-Gómez, "Improving glaucoma diagnosis assembling deep networks and voting schemes," *Diagnostics*, vol. 12, no. 6, p. 1382, Jun. 2022, doi: 10.3390/diagnostics12061382. [Online]. Available: https://www.mdpi.com/2075-4418/12/6/1382

[5] M. Alghamdi and M. Abdel-Mottaleb, "A comparative study of deep learning models for diagnosing glaucoma from fundus images," *IEEE Access*, vol. 9, pp. 23894–23906, 2021, doi: 10.1109/ACCESS.2021.3056641.

[6] T. Saba, S. T. F. Bokhari, M. Sharif, M. Yasmin, and M. Raza, "Fundus image classification methods for the detection of glaucoma: A review," *Microsc. Res. Technique*, vol. 81, no. 10, pp. 1105–1121, Oct. 2018, doi: 10.1002/jemt.23094.

[7] S. Virbukaitė and J. Bernatavičienė, "Deep neural networks application for cup-to-disc ratio estimation in eye fundus images," *Proc. 18th Conf. Comput. Sci. Intell. Syst.*, pp. 1191–1195, Sep. 2023.

[8] J. Kim, L. Tran, T. Peto, and E. Y. Chew, "Deep learning and ensemble method for optic disc and cup segmentation," in *Proc. IEEE Conf. Comput. Intell. Bioinf. Comput. Biol. (CIBCB)*, Ottawa, ON, Canada, Aug. 2022, pp. 1–8.

[9] J. Civit-Masot, M. J. Domínguez-Morales, S. Vicente-Díaz, and A. Civit, "Dual machine-learning system to aid glaucoma diagnosis using disc and cup feature extraction," *IEEE Access*, vol. 8, pp. 127519–127529, 2020, doi: 10.1109/ACCESS.2020.3008539.

[10] A. U. Rehman, I. A. Taj, M. Sajid, and K. S. Karimov, "An ensemble framework based on deep CNNs architecture for glaucoma classification using fundus photography," *Math. Biosci. Eng.*, vol. 18, no. 5, pp. 5321–5346, Jun. 2021, doi: 10.3934/mbe.2021270.

[11] V. Kurilová, S. Rajcsányi, Z. Rábeková, J. Pavlovičová, M. Oravec, and N. Majtánová, "Detecting glaucoma from fundus images using ensemble learning," *J. Electr. Eng.*, vol. 74, no. 4, pp. 328–335, Aug. 2023, doi: 10.2478/jee-2023-0040.

[12] M. Y. S. Ali, M. Abdel-Nasser, M. Jabreel, A. Valls, and M. Baget, "Segmenting the optic disc using a deep learning ensemble model based on OWA operators," in *Proc. 23rd Int. Conf. Catalan-Associat. Artif. Intell. (CCIA)*. Electr Network, 2021, pp. 305–314.

[13] Y. Jiang, F. Wang, J. Gao, and S. Cao, "Multi-path recurrent U-Net segmentation of retinal fundus image," *Appl. Sci.*, vol. 10, no. 11, p. 3777, May 2020, doi: 10.3390/app10113777.

[14] Q. Zhu, X. Chen, Q. Meng, J. Song, G. Luo, M. Wang, F. Shi, Z. Chen, D. Xiang, L. Pan, Z. Li, and W. Zhu, "GDCSeg-Net: General optic disc and cup segmentation network for multi-device fundus images," *Biomed. Opt. Exp.*, vol. 12, no. 10, p. 6529, 2021, doi: 10.1364/boe.434841.

[15] A. Zhao, H. Su, C. She, X. Huang, H. Li, H. Qiu, Z. Jiang, and G. Huang, "Joint optic disc and cup segmentation based on elliptical-like morphological feature and spatial geometry constraint," *Comput. Biol. Med.*, vol. 158, May 2023, Art. no. 106796, doi: 10.1016/j.compbiomed.2023.106796.

[16] Z. Li, C. Zhao, Z. Han, and C. Hong, "TUNet and domain adaptation based learning for joint optic disc and cup segmentation," *Comput. Biol. Med.*, vol. 163, Sep. 2023, Art. no. 107209, doi: 10.1016/j.compbiomed.2023.107209.

[17] R. Bhattacharya, R. Hussain, A. Chatterjee, D. Paul, S. Chatterjee, and D. Dey, "PY-net: Rethinking segmentation frameworks with dense pyramidal operations for optic disc and cup segmentation from retinal fundus images," *Biomed. Signal Process. Control*, vol. 85, Aug. 2023, Art. no. 104895, doi: 10.1016/j.bspc.2023.104895.

[18] J. Zilly, J. M. Buhmann, and D. Mahapatra, "Glaucoma detection using entropy sampling and ensemble learning for automatic optic cup and disc segmentation," *Comput. Med. Imag. Graph.*, vol. 55, pp. 28–41, Jan. 2017, doi: 10.1016/j.compmedimag.2016.07.012.

[19] P. Elangovan and M. K. Nath, "En-ConvNet: A novel approach for glaucoma detection from color fundus images using ensemble of deep convolutional neural networks," *Int. J. Imag. Syst. Technol.*, vol. 32, no. 6, pp. 2034–2048, Nov. 2022, doi: 10.1002/ima.22761.

[20] J. Wang, X. Li, and Y. Cheng, "Towards an extended EfficientNet-based U-Net framework for joint optic disc and cup segmentation in the fundus image," *Biomed. Signal Process. Control*, vol. 85, Aug. 2023, Art. no. 104906, doi: 10.1016/j.bspc.2023.104906.

[21] O. Oktay, J. Schlemper, L. L. Folgoc, M. Lee, M. Heinrich, K. Misawa, K. Mori, S. McDonagh, N. Y Hammerla, B. Kainz, B. Glocker, and D. Rueckert, "Attention U-Net: Learning where to look for the pancreas," 2018, *arXiv: 1804.03999*.

[22] S. Virbukaitė and J. Bernatavičienė, "Impact of eye fundus image preprocessing on key objects segmentation for glaucoma identification," *Nonlinear Anal., Model. Control*, vol. 29, no. 1, pp. 96–110, Nov. 2023, doi: 10.15388/namc.2024.29.33669.

[23] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," 2015, *arXiv:1512.03385*.

[24] A. G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto, and H. Adam, "MobileNets: Efficient convolutional neural networks for mobile vision applications," 2017, *arXiv:1704.04861*.

[25] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," 2015, *arXiv:1512.00567*.

[26] G. Huang, Z. Liu, L. van der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," 2016, *arXiv:1608.06993*.

[27] M. M. Fraz, P. Remagnino, A. Hoppe, B. Uyyanonvara, A. R. Rudnicka, C. G. Owen, and S. A. Barman, "An ensemble classification-based approach applied to retinal blood vessel segmentation," *IEEE Trans. Biomed. Eng.*, vol. 59, no. 9, pp. 2538–2548, Sep. 2012, doi: 10.1109/TBME.2012.2205687.

[28] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," 2015, *arXiv:1505.04597*.

[29] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 1–9.

[30] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in *Proc. 32nd Int. Conf. Mach. Learn.*, 2015, pp. 448–456.

[31] S. Sabzi, R. Pourdarbani, D. Kalantari, and T. Panagopoulos, "Designing a fruit identification algorithm in orchard conditions to develop robots using video processing and majority voting based on hybrid artificial neural network," *Appl. Sci.*, vol. 10, no. 1, p. 383, Jan. 2020, doi: 10.3390/app10010383.

[32] J. M. Montgomery, F. M. Hollenbach, and M. D. Ward, "Improving predictions using ensemble Bayesian model averaging," *Political Anal.*, vol. 20, no. 3, pp. 271–291, 2012, doi: 10.1093/pan/mps002.

[33] G. R. Latif-Shabgahi, "A novel algorithm for weighted average voting used in fault tolerant computing systems," *Microprocessors Microsyst.*, vol. 28, no. 7, pp. 357–361, Sep. 2004, doi: 10.1016/j.micpro.2004.02.006.

[34] G. V. Vlasceanu, R.-M. Paraschiv, C. Artene, and C.-A. Boiangiu, "Voting-based edge detection," in *Proc. 18th RoEduNet Conf., Netw. Educ. Res. (RoEduNet)*, Oct. 2019, pp. 1–5.

[35] J. I. Orlando et al., "REFUGE challenge: A unified framework for evaluating automated methods for glaucoma assessment from fundus photographs," *Med. Image Anal.*, vol. 59, Jan. 2020, Art. no. 101570, doi: 10.1016/j.media.2019.101570.

[36] Medical Image Analysis Group. (2023). *RIME-ONE V.3 Datase*. [Online]. Available: http://medimrg.webs.ull.es/research/retinal-imaging/rim-one

[37] J. Sivaswamy, S. R. Krishnadas, G. D. Joshi, M. Jain, and A. U. S. Tabish, "Drishti-GS: Retinal image dataset for optic nerve head(ONH) segmentation," in *Proc. IEEE 11th Int. Symp. Biomed. Imag. (ISBI)*, Apr. 2014, pp. 53–56.

[38] D. Vijayalakshmi and M. K. Nath, "A systematic approach for enhancement of homogeneous background images using structural information," *Graph. Models*, vol. 130, Dec. 2023, Art. no. 101206, doi: 10.1016/j.gmod.2023.101206.

[39] L. Jiang, X. Tang, S. You, S. Liu, and Y. Ji, "BEAC-Net: Boundary-enhanced adaptive context network for optic disk and optic cup segmentation," *Appl. Sci.*, vol. 13, no. 18, p. 10244, Sep. 2023, doi: 10.3390/app131810244.

[40] W. Zhou, J. Ji, Y. Jiang, J. Wang, Q. Qi, and Y. Yi, "EARDS: EfficientNet and attention-based residual depth-wise separable convolution for joint OD and OC segmentation," *Frontiers Neurosci.*, vol. 17, pp. 1–15, Mar. 2023, doi: 10.3389/fnins.2023.1139181.

[41] A. Tulsani, P. Kumar, and S. Pathan, "Automated segmentation of optic disc and optic cup for glaucoma assessment using improved UNET++ architecture," *Biocybern. Biomed. Eng.*, vol. 41, no. 2, pp. 819–832, Apr. 2021, doi: 10.1016/j.bbe.2021.05.011.

[42] S. Tadisetty, R. Chodavarapu, R. Jin, R. J. Clements, and M. Yu, "Identifying the edges of the optic cup and the optic disc in glaucoma patients by segmentation," *Sensors*, vol. 23, no. 10, p. 4668, May 2023, doi: 10.3390/s23104668.

[43] X. Zhang, J. Song, C. Wang, and Z. Zhou, "Convolutional autoencoder joint boundary and mask adversarial learning for fundus image segmentation," *Frontiers Hum. Neurosci.*, vol. 16, pp. 1–12, Dec. 2022, doi: 10.3389/fnhum.2022.1043569.

[44] Y. Yi, Y. Jiang, B. Zhou, N. Zhang, J. Dai, X. Huang, Q. Zeng, and W. Zhou, "C2FTFNet: Coarse-to-fine transformer network for joint optic disc and cup segmentation," *Comput. Biol. Med.*, vol. 164, Sep. 2023, Art. no. 107215, doi: 10.1016/j.compbiomed.2023.107215. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0010482523006807

[45] ITAPC. *Information Technology Research Center of Vilnius University*. Accessed: Jan. 20, 2024. [Online]. Available: https://mif.vu.lt/lt3/en/about/structure/it-research-center

[46] S. Lu, H. Zhao, H. Liu, H. Li, and N. Wang, "PKRT-Net: Prior knowledge-based relation transformer network for optic cup and disc segmentation," *Neurocomputing*, vol. 538, Jun. 2023, Art. no. 126183, doi: 10.1016/j.neucom.2023.03.044. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0925231223002953

SANDRA VIRBUKAITĖ (Student Member, IEEE) received the B.S. degree in engineering informatics and the M.S. degree in applied statistics from Vilnius Gediminas Technical University, Lithuania, in 2009 and 2011, respectively. She is currently pursuing the Ph.D. degree in engineering informatics with Vilnius University, Lithuania. Her current research interest includes artificial intelligence-aided glaucoma identification from eye fundus images.

JOLITA BERNATAVIČIENĖ (Member, IEEE) received the master's degree in informatics from Vilnius Pedagogical University, in 2004, and the Ph.D. degree in computer science from the Institute of Mathematics and Informatics, Vilnius Gediminas Technical University, in 2008. She is currently a Senior Researcher with the Cognitive Computing Group, Institute of Data Science and Digital Technologies, Vilnius University. Her research interests include databases, data mining, neural networks, image analysis, visualization, decision support systems and internet technologies, and high-performance computing. She supervises three Ph.D. students and has written more than 60 articles, 18 of which are in the CA WoS Database.
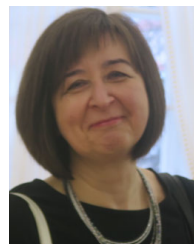
DAIVA IMBRASIENĖ received the Doctor of Medicine degree from Kaunas Medical Academy, in 1996. She has been an Associate Professor, since 2009. Her research interests include ultrasonography of the eyes and orbit, examination of retinal blood vessels, analysis of the eye fundus, intraocular pressure in athletes during rest and physical load, search for diagnostic algorithms in digital images of the eye fundus, eye examinations for individuals with multiple sclerosis, diagnosis, and prevention of chronic non-infectious diseases, use of imaging technologies in disease diagnostics, E-medicine, remote transmission of diagnostic images, and telemedicine.

● ● ●