Aswan University Journal of Science and Technology Volume 4, issue 1, March 2024

Improving Glaucoma Detection: Harnessing the Power of Ensemble Semantic Segmentation for Optic Disc and Optic Cup with Deep Learning

YASSER AHMED DAHAB
College of Computing and Information Technology, Arab Academy for Science, Technology and Maritime Transport, South Valley Branch, Aswan 81516, Egypt

Abstract

Globally, glaucoma is a leading cause of irreversible visual loss. Due to the lack of symptoms in the early stages of the disease, glaucoma is typically not diagnosed until severe vision loss has occurred. One of the most common ways to diagnose glaucoma is with a comprehensive eye exam. However, a substantial commitment of time, money, and specialist equipment and personnel is necessary to carry out such investigations. Using deep learning optic cup and disc segmentation based on retinal fundus images, this work aims to develop and evaluate the effectiveness of a novel, affordable glaucoma screening tool. The research made use of ensemble learning technique to enhance semantic segmentation models. A number of recognized deep learning architectures, including Unet 3Plus, Deep Lab V3P, PSPNET and UW-Net, are combined in the proposed semantic ensemble segmentation model. Ensemble combination involves merging predictions from several models using a weighted averaging technique that considers the accuracy and dependability of each individual model. Metrics such as accuracy, specificity, sensitivity, area under the curve, intersection over union, dice coefficient, and f1-score are used to evaluate the performance in the study. The proposed model was validated on three publicly available datasets, namely ORIGA, REFUGE and RIM-ONE DL. The experimental results show that the suggested method can estimate the Cup to Disc Ratio CDR for thorough glaucoma screening, and it is on par with the state-of-the-art architecture utilized for optic disc and cup segmentation.

Keywords: Glaucoma diagnosis; segmentation of the optic cup/disc; Deep learning; Ensemble learning

1. Introduction

Glaucoma is a pathological condition characterized by progressive degeneration of nerve fibers leading to gradual impairment of the optic nerve. This impairment of the optic nerve is caused by the increased intraocular pressure. The optic nerve serves as a conduit for transmitting visual impulses from the retina to the brain, where they are then processed as visual images. Glaucoma is one of the most common causes of irreversible vision problems worldwide [1].

It is a group of eye diseases that can potentially cause vision problems and, in severe cases, complete loss of vision. Delaying or not treating glaucoma increases the likelihood that the patient will experience irreversible visual impairment. Recognition of glaucoma can be delayed until significant vision impairment occurs because the early stages are often without obvious symptoms [2]. In order to diagnose glaucoma, a thorough ophthalmologic examination is typically conducted. Analyzing the visual field, measuring intraocular pressure using tonometry, and evaluating the optic disc and optic cup should all be part of a thorough evaluation during this examination [3].

It should be mentioned that undertaking such tests frequently necessitates a substantial financial outlay, in addition to specialized tools and staff. Consequently, this element must be acknowledged and assessed [4]. Furthermore, ophthalmologists are in short supply in rural and developing nations. An automated glaucoma detection system that can

Corresponding author E-mail: ydahab@aast.edu
Received, November 26, 2023, received in revised form December 6, 2023 accepted December 11, 2023
(ASWJST 2022/ printed ISSN: 2735-3087 and on-line ISSN: 2735-3095) https://journals.aswu.edu.eg/stjournal

offer affordable, accurate, quick, and easily understandable diagnoses is essential for processing massive volumes of data and reducing screening-related human error. Automating the glaucoma diagnostic procedure with artificial intelligence algorithms is thus becoming more common.

In recent years, there have been extensive studies on the application of deep learning algorithms for glaucoma diagnosis through semantic segmentation of the optic cup and optic disc.

In order to identify the discrepancy between the optic disc and the orbit, Rakhshanda et al. [5] performed a pixel-wise analysis using a semantic segmentation model. The initial stage in completing this work is to generate feature vectors that will be utilized for object detection (OD), object classification (OC), and background classification. In order to achieve this, a VGG-16 network is trained using supplementary data in addition to the original data set. The Cup-to-Disc Ratio (CDR), an important metric in glaucoma research and diagnosis, can be computed after the segmentation data has been acquired.

The DENet model for glaucoma diagnosis was created by Cheng et al. [6]. The authors' proposed network combines four submodels and employs both global and local fundus image data. The ResNet architecture is used for image classification, while the U-Net model is used to segment disc areas. The segmented disc regions are then subjected to probabilistic screening procedures, and the accuracy of the segmentation is improved by employing a polar transform. These four results were combined to provide the final screening result.

An integrated glaucoma screening paradigm was introduced by the writers of [7]. Their strategy involved making some adjustments to the U-Net model and then replacing the encoder component with a pre-trained VGG-16 model. Finding the fovea, segmenting the optic nerve head (OD) and optic cup (OC), and making glaucoma diagnoses easier are all goals of the study, which employs retinal fundus images. To do this, the model's encoder and decoder layers are connected using hop links, and a special optimization technique is used.

The aim of the study by Junja et al. [8] was to develop a glaucoma diagnostic method based on retinal fundus images. To achieve this, the researchers designed G-Net, a deep CNN architecture for glaucoma diagnosis. The enhanced U-Net architecture in G-Net integrates convolutional neural networks (CNNs) for the purpose of distinguishing between the optic disc (OD) and optic cup (OC) inputs, allowing the CDR calculation. The use of the red channel is deferred to the input images, which are then used for object detection (OD) segmentation, until the pre-processing and cropping operations are complete. Traditional segmentation methods only use two of the red-green-blue (RGB) channels. However, object contour segmentation (OC) uses all three RGB channels. Before being sent to G-Net for glaucoma detection, the collected data, derived from the calculation of the cup-to-disc ratio (CDR), undergoes an enhancement.

Identification of glaucoma was achieved by using preprocessing, segmentation and classification techniques as described by Natarajan et al. described. described [9]. Preprocessing was performed using Contrast Limited Adaptive Histogram Equalization (CLAHE). The regions of interest (ROIs) were then selected using only the data from the green channel. To distinguish between OD and OC, the researchers used the modified kernel fuzzy C-means approach. After segmenting the data, the subsequent procedure included the extraction of GLCM (Gray-Level Cookcurrence Matrix) features. These features were then used as input to a VGG16 classifier, with the aim of determining the extent or severity of glaucoma.

To effectively discriminate between optic discs (ODs) and optic cups (OCs) without using localization or prepost processing techniques, Tabassum et al. [10] presented the use of a dense cup-disc encoder-decoder network. The integration of a decoder into a system facilitates the effective use of data, while encoding is the key factor that enables the smooth reuse of functions. As a result of this phenomenon, the need for upsampling functions can be reduced, allowing a reduction in the number of network nodes without sacrificing performance.

The study by Sreng et al. [11] investigated the use of a two-stage deep learning architecture for glaucoma diagnosis. Many convolutional neural networks work together to improve results. The first phase uses a modified DeepLabv3+ architecture, which is then combined with five different network configurations to partition the object detection area (OD). In addition, a total of eleven pre-trained CNNs were used to extract features from the object detection (OD) region of interest. The study used transfer learning, support vector machines (SVM), or a combination of both methods to achieve the desired result.

To improve the performance of the network, Liu et al. [12] developed a deeply separable convolutional network that has a dense connection at its core and a multi scale image pyramid as an additional component. Image morphology techniques have been employed to improve the segmentation output and to facilitate further processing steps. Optic nerve nucleus identification was performed locally using a convolutional neural network (CNN) and Hough circle detection methods. A segmentation network was trained using the successfully retrieved area of interest. After completing the network training process, the papilla underwent precise segmentation, resulting in its components being clearly separated.

In their study, Zhao et al. [13] presented a simple approach aimed at enhancing the accuracy of fundus image segmentation. This approach achieves improved accuracy through the use of an attentional U-Net model and transfer learning. By incorporating an attention gate mechanism between the encoder and decoder components, the model is able to prioritize and focus its attention on specific areas of interest within the target data. This approach also shortens processing times and minimizes the amount of parameters.

To efficiently segment the optic discs and categorize glaucoma patients into two groups, Ganesh et al. [14] suggested use the GD-Ynet deep learning model. The modified U-Net architecture incorporates inception modules in place of convolutional layers, which are often employed for low-level character recognition. Disc segmentation and ROI recognition are accomplished using the method outlined in the study by utilizing characteristics extracted from contextual activation maps. Images of the optic nerve head are analyzed using a binary classification approach based on a sequence of cumulative alterations in order to diagnose glaucoma.

Panda et al. [15] constructed a deep learning technique for segmentation of the optic disc and optic cup based on fundus images to detect glaucoma using a limited number of photos. The use of post processing approaches, residual learning with jump connections, and patch based training in this approach results in improved edge smoothness and a more precise cup to disc ratio.

Veena et al. [16] conducted a study in which they developed a set of criteria for glaucoma detection. To delineate the orbital region from the optic disc and calculate the cup to disc ratio (CDR), this method used a pair of convolutional neural network (CNN) models. To improve feature extraction and refine the input, the models use a comprehensive set of 39 layers of convolutional neural networks (CNNs). The primary methods used in the image preparation process include morphological techniques to enhance contrast, feature extraction with a Sobel filter, and optic nerve localization through the application of the Watershed algorithm. Data from both models are used in the calculation of the CDR.

Nazir et al. [17] developed a deep learning technique in conjunction with a modified mask RCNN model for autonomously extraction of the optic disc and optic cup. The ground truth annotations were created using a data augmentation technique that intentionally smeared the data to increase the variation of the data set. They used the DenseNet-77 Architecture in MaskRCNN's feature computation technique to improve the precision of identifying the object detection (OD) and object classification (OC) regions in different sample environments.

Wang et al. [18] proposed an asymmetric network for optic nerve head segmentation using U-Net, a unique cross-connected subnetwork and a decoding convolution block. The network effectively combats the negative effects of recurring pooling by accounting for the presence of multi-scale input features. By integrating these elements, it is possible to reduce the data loss that occurs during the compression process. This consequently increases the network's ability to detect and understand morphological irregularities within the important image segments.

Kumar et al. [19] presented an approach that uses morphological approaches to generate the highest quality ground truth data. The U-Net architecture is used to distinguish spatial properties. The architecture has a total of nineteen Convolutional

Neural Network (CNN) layers, with each layer having its own encoder and decoder blocks. The notable aspect therefore lies in the improved predictive capacity of the model, which is specific to the optic disc region.

Fu et al. [20] used a U-Net architecture to perform optic disc segmentation in images of the abnormal retinal background. The proposed methodology uses model driven probability bubbles to accurately identify the exact position of the optic disc and effectively eliminate the interferences caused by slight lesions, resulting in improved segmentation accuracy.

Researchers in [21] used FCNs to develop a deep learning system for segmenting optic disc and cup regions. The dataset known as RIM-ONE was subjected to analysis by the authors using both single-tier and two-tier approaches. The results showed that the accuracy achieved using a single layer was 95.6%, but using two layers the accuracy increased to 96.9%. The results show that the AUC reached a rate of 98% with one layer and 97.8% with two layers.

The authors of [22] successfully diagnosed glaucoma using a dataset consisting of 1542 retinal fundus images and multimodal data. The above approach was used by leveraging transfer learning, specifically using the GoogleNet and Inception-V3 models. The results obtained include an accuracy rate of 84.5% and an AUC of 93%.

Although deep learning algorithms have made significant progress in detecting glaucoma, there is a clear lack of study in the area of ensemble semantic segmentation for the optic disc and optic cup. Although individual deep learning models have demonstrated potential in segmenting these structures, the advantages of utilizing ensemble methods to combine many models have not been thoroughly investigated. The current body of research predominantly concentrates on single-model methodologies, disregarding the benefits that can be obtained by harnessing various models and their complimentary capabilities. Moreover, there has been insufficient focus on the comprehensibility of ensemble models in the context of glaucoma detection, impeding the comprehension of the decision-making procedure. The proposed study aims to fill a research gap by examining the efficacy of ensemble semantic segmentation for the optic disc and optic cup using deep learning methods. This will contribute to the existing literature and offer valuable insights into potential enhancements in glaucoma detection.

Ensemble learning is used to overcome the limitations of single semantic segmentation models. It leverages the collective intelligence of multiple models to improve performance and address critical issues. Ensemble learning also improves the accuracy of segmentation, reduces overfitting, and evaluates uncertainty by combining model predictions to demonstrate the reliability of the predictions. Ensemble learning increases the diversity of models, allowing them to capture a wider range of patterns and increasing their ability to withstand noise [23]. This study uses a deep learning semantic ensemble segmentation technique to accurately delineate the optic nerve head and optic cup in retinal fundus images. To diagnose glaucoma, the cup-to-disc ratio is then calculated.

The following are the main contributions of this study:

(i) Propose a semantic ensemble deep learning segmentation model using a probability averaging strategy for the optic cup and optic nerve head is proposed.

(ii) Validate on three public fundus images datasets

(iii) Perform a comparative study between four different semantic segmentation models and the proposed ensemble model for optic cup and optic disc segmentation.

The paper is structured as follows: Section 2 describes the specific details of the proposed method and Section 3 provides the experimental results of the study. The discussions and conclusions can be found in Sections 4 and 5.

2. Materials and Method

2.1 Datasets

The study uses three public datasets, namely ORIGA, REFUGE and RIM-ONE DL. Table 1 depicts the descriptions of the public datasets used in this study.

The ORIGA dataset contains 168 photos of people with glaucoma and 482 photos of healthy people [24]. Disc and cup masks are included in the data.

The REFUGE dataset includes 120 images of individuals diagnosed with glaucoma and 1080 images of healthy individuals with segmented disc and cup. The photos were taken with different cameras and clearly illustrate the striking difference between the training data and the test data [25].

RIM-ONE DL includes both healthy and glaucomatous patient fundi [26]. All photos were saved in PNG format after being cropped to remove unwanted backgrounds and centered on the optic nerve head [26].

Figure 1 shows some example images of fundus glaucoma from the datasets used in the study along with the optic disc and optic disc masks.

Table 1. Overview of the public datasets employed in this study

Data set	Number of images			Ethnicity
	Glaucoma	Normal	Total	
ORIGA	168	482	650	Malay
REFUGE	120	1080	1200	Chinese
RIM-ONE DL	172	313	485	Spanish

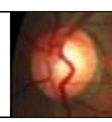
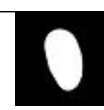
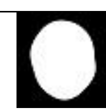
Dataset	ORIGA		REFUGE		RIM-ONE DL		
	Image	Cup Disc	Image	Cup Disc	Image	Cup	Disc
Glaucoma							

Figure 1. Example images of fundus glaucoma and the associated optic disc and optic disc masks

2.2 Preprocessing

The image preprocessing flow in this study is shown in Figure 2. Preprocessing occurs following: The first step is to improve image contrast and reduce the effects of noise and anomalies. The second step, on the other hand, aims to address the space constraints and unbalanced data issues of colored fundus images.

In the first step, two preprocessing techniques are used: median filtering and contrast limited adaptive histogram equalization (CLAHE). By redistributing pixel intensities, CLAHE is used to increase the contrast of an image. Adaptive histogram equalization is then applied to each tile, which are small regions formed from the image. This method guarantees the localization of contrast enhancement, reducing the risk of image noise or anomalies that are over-enhanced [27].

Subsequently, median filtering is employed to lower noise and achieve image smoothing. Intensity values of each pixel are replaced with the median value of the pixels around it within a certain range. This method is especially efficient in reducing salt-and-pepper noise, which may occur in color fundus images as a result of variables including image acquisition or transmission abnormalities. Utilizing median filtering diminishes the noise, leading to an enhanced and dependable image. Implementing these pre-processing steps significantly improves quality and results in more reliable and accurate analysis.

The space limitations in colored fundus images occur because the OC in the colored fundus images lies within the OD, making it difficult to accurately represent this condition using Cartesian coordinates. On the other hand, these images show significant differences in pixel distribution between OC, OD and the background, which causes the problem of unbalanced data. Unbalanced data can lead to training bias and overfitting, reducing the accuracy of OD and OC segmentation.

The polar transformation of the colored fundus photos results in a layered structure that presents the OC, OD and background pixels in an organized manner. This addresses the concerns related to space limitations. In addition, the polar transform flattens the pixel representation around the center of the optic nerve head, resulting in an improved portion of the optic cup in the image. This helps address the problem of imbalanced data [28].

The dataset was partitioned into training, validation, and test sets in the proportions of 80%, 10%, and 10%, respectively. Augmentation was then performed to reduce overfitting and for generalization purposes. The images were flipped vertically and horizontally. To speed up the implementation of segmentation, each image was downsampled to 128x128.

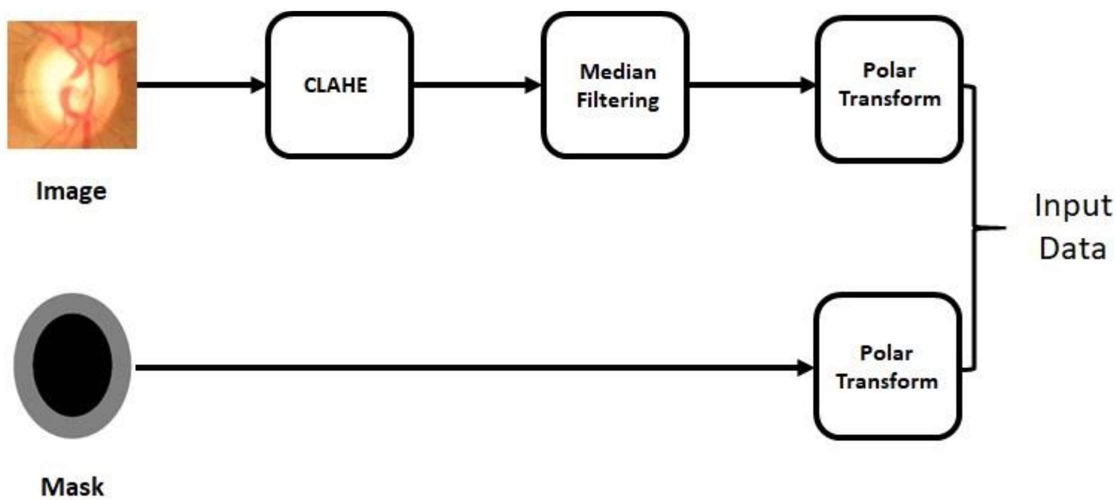


Figure 2. The preprocessing method used in the study

2.3 The model

One distinctive feature of the proposed model is its integration of deep learning techniques for glaucoma detection with ensemble semantic segmentation for the optic disc and optic cup. While previous studies have looked at deep learning models for structure segmentation alone [29], the suggested approach introduces a novel ensemble framework that combines the predictions of multiple models. The accuracy and robustness of the glaucoma detection system can be enhanced by learning from a bigger pool of information through the combination of different models. The suggested model fills a major need in the existing literature by combining ensemble methods and deep learning to improve glaucoma detection in a novel and comprehensive way.

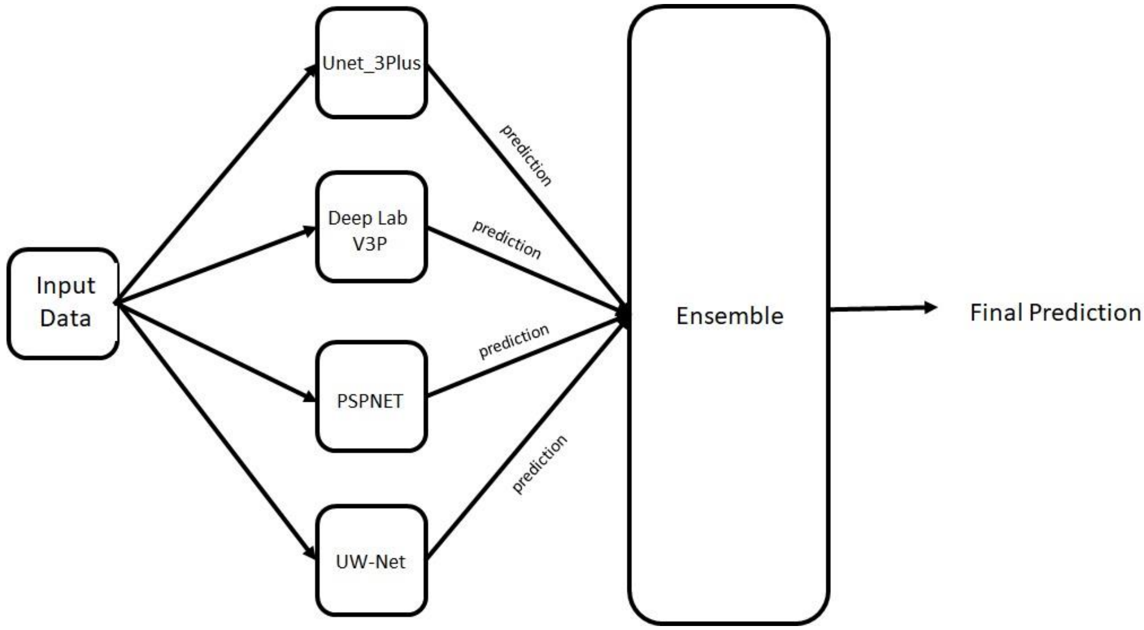


Figure 3. The proposed method

Four distinct semantic segmentation architectures are chosen as base models to provide a strong basis for the ensemble. The architectures include Unet_3Plus [30], Deep Lab V3P [31], PSPNET [32], and UW-Net [33]. Every base model is trained separately on the three datasets included in the research, which consist of annotated images depicting cup and disc regions. After training the base models, the test data from the three datasets is used to provide predictions for optic cup and disc segmentation. Every base model analyzes the incoming images and generates segmentation masks that precisely outline the areas of interest. The projected masks are then used as the foundation for the next ensemble aggregation.

Subsequently, several ensemble procedures are explored, such as voting, averaging, weighted averaging, and stacking. The ensemble aggregation stage entails implementing the chosen ensemble method to merge the predictions generated by the basis models. In a vote-based ensemble, the ultimate forecast for each pixel is decided by majority voting among the basic models. In contrast, averaging-based ensembles use the mean or median of the pixel-wise predictions from the underlying models to get the ultimate segmentation prediction. A performance assessment is conducted on the test data of the three datasets to assess the efficacy of the ensemble process. The ensemble's performance is assessed by comparing it to each individual baseline model, in order to quantify the improvement made via the use of the ensemble approach. Figure 3 shows the proposed methodology.

2.4 Evaluation Metrics

Table 2 shows the evaluation metrics used in this study along with a description of each metric. These metrics are calculated based on the confusion matrix and the corresponding values of true positive (TP), true negative (TN), false positive (FP), and false negative (FN).

Table 2. Evaluation metrics used for segmentation of the optic disc and optic cup

Metric	Formula	Description
Accuracy (ACC)	$\frac{TP+TN}{TP+TN+FP+FN}$	measures the correctly identiefied cases
Specificity (SPE)	$\frac{TN}{TN+FP}$	quantifies the accuracy of Identifying instances without glaucoma.
Sensitivity (SEN)	$\frac{TP}{TP+FN}$	quantifies the proportion of Accurately diagnosed instances of glaucoma.
Intersection over Union (IoU)	$\frac{TP}{TP+FP+FN}$	measures the degree of overlap between the predicted segmented region and the ground truth annotated region from a dataset.
Dice Coefficient (DSC)	$\frac{2\ X\ TP}{2\ X\ TP+FP+FN}$	assesses the effectiveness of segmentation algorithms
F1-Score	$\frac{2\ X\ \text{Sensitivity X Specificity}}{\text{Sensitivity} + \text{Specificity}}$	figures out how well and how efficiently segmentation works.

3. Experiments and Results

3.1 Experimental setup

The experiments were carried out using Google Colab with T4 GPU, Python 3.10.12, and Tensorflow 2.14. The following setup was used for models evaluation: Nadam optimizer, the batch size is set to 16, ReduceLROnPlateau is used with a factor of 10 in a patience of 15 epochs. Early stopping is applied with a patience of 30 epochs. The used hyperparameters are listed in Table 3.

Table 3. The hyperparameters

Batch Size	Loss Function	Learning Rate	Optimizer	Epochs
16	Categorical crossentropy	0.00001	Nadam	100

3.2 Performance Comparisons

A grid search was performed to obtain the best matching weights of the predictions. The ensemble semantic segmentation model underwent testing on three distinct datasets including colored fundus images. The model demonstrated high efficacy in precisely detecting the optic nerve head and optic cup. The model's performance was assessed using quantitative assessment metrics including accuracy, sensitivity, specificity, area under the curve (AuC), intersection over union (IoU), Dice coefficient, and F1 score. The performance assessment findings of the three datasets are shown in Tables 4–6. The performance comparison with other methods is shown in Tables 7–8.

Among the four models evaluated on the Origa dataset, PSPNET demonstrates superior performance in both optic disc and optic disc segmentation. Conversely, according to the Refuge dataset, Unet_3Plus has superior performance in segmenting the optic nerve head. UW-Net demonstrates superior performance in optic nerve cup segmentation on the same dataset. Unet_3Plus demonstrates superior performance in optic disc and optic cup segmentation on the RIM ONE DL dataset.

The absence of a better segmentation model for the three datasets is evident, highlighting the need of using ensemble learning. The research used ensemble learning techniques to enhance the overall performance. The proposed semantic ensemble segmentation model demonstrates significantly superior performance in comparison to the individual segmentation models across the three datasets used.

In summary, the quantitative assessment metrics of the segmentation findings provide strong support for the efficiency and possible clinical significance of the ensemble semantic segmentation strategy for detecting glaucoma. The model's excellent sensitivity, specificity, F1 score, and accuracy values indicate its proficiency in detecting the optic nerve head and accurately segmenting the optic nerve cup. This capability is essential for early diagnosis and treatment planning.

The findings underscore the significance of this study and its capacity to enhance glaucoma identification and patient outcomes. To ensure the reliability and applicability of the suggested semantic ensemble segmentation methodology, it is essential to conduct further examination and verification using bigger and more varied datasets. Additionally, comparing the approach with current cutting-edge approaches is necessary.

Table 4. Results for the test set in the ORIGA dataset

Method	Optic Disc Segmentation							Optic Cup Segmentation						
	Acc	Spe	Sen	Auc	Iou	DSC	F1	Acc	Spe	Sen	Auc	Iou	DSC	F1
Unet_3Plus	0.970	0.981	0.947	0.964	0.935	0.951	0.953	0.932	0.890	0.965	0.928	0.870	0.940	0.970
Deep Lab V3P	0.975	0.992	0.936	0.964	0.942	0.956	0.958	0.925	0.841	0.977	0.910	0.852	0.939	0.975
PSPNET	0.979	0.985	0.965	0.975	0.950	0.962	0.963	0.944	0.927	0.957	0.941	0.891	0.953	0.979
UW-Net	0.979	0.988	0.957	0.973	0.952	0.964	0.966	0.943	0.919	0.957	0.938	0.888	0.950	0.979
Ensemble	0.993	0.997	0.972	0.985	0.971	0.982	0.982	0.963	0.949	0.984	0.962	0.918	0.968	0.983

Notes: Blue represents the best performance, green represents the second best performance.

Table 5. Results for the test set in the REFUGE dataset

Method	Optic Disc Segmentation							Optic Cup Segmentation						
	Acc	Spe	Sen	Auc	Iou	DSC	F1	Acc	Spe	Sen	Auc	Iou	DSC	F1
Unet_3Plus	0.975	0.981	0.967	0.974	0.950	0.970	0.058	0.970	0.961	0.941	0.968	0.954	0.907	0.053
Deep Lab V3P	0.959	0.991	0.913	0.952	0.918	0.945	0.099	0.948	0.949	0.886	0.974	0.929	0.882	0.066
PSPNET	0.968	0.978	0.951	0.965	0.935	0.956	0.076	0.960	0.956	0.937	0.934	0.950	0.899	0.058
UW-Net	0.972	0.978	0.962	0.970	0.944	0.965	0.066	0.966	0.962	0.947	0.969	0.958	0.912	0.049
Ensemble	0.984	0.989	0.979	0.987	0.969	0.985	0.049	0.986	0.960	0.964	0.987	0.969	0.934	0.041

Notes: Blue represents the best performance, green represents the second best performance.

Table 6. Results for the test set in the RIM ONE DL dataset

Method	Optic Disc Segmentation							Optic Cup Segmentation						
	Acc	Spe	Sen	Auc	Iou	DSC	F1	Acc	Spe	Sen	Auc	Iou	DSC	F1
Unet_3Plus	0.968	0.979	0.951	0.965	0.935	0.956	0.076	0.960	0.935	0.950	0.905	0.927	0.864	0.179
Deep Lab V3P	0.956	0.870	0.990	0.931	0.900	0.970	0.058	0.970	0.895	0.953	0.776	0.865	0.785	0.289
PSPNET	0.970	0.938	0.983	0.961	0.930	0.979	0.041	0.979	0.931	0.951	0.889	0.920	0.856	0.190
UW-Net	0.962	0.911	0.983	0.947	0.913	0.974	0.059	0.974	0.910	0.953	0.821	0.887	0.813	0.250
Ensemble	0.983	0.974	0.989	0.978	0.952	0.976	0.0431	0.981	0.939	0.962	0.899	0.939	0.889	0.149

Notes: Blue represents the best performance, green represents the second best performance.

Table 7. Comparison with other methods using the RIM ONE dataset

Method	Optic Disc Segmentation					Optic Cup Segmentation				
	Acc	Spe	Sen	Iou	DSC	Acc	Spe	Sen	Iou	DSC
[13]	0.996	0.999	0.924	0.887	0.940	0.997	0.999	0.813	0.724	0.840
[10]	0.996	0.997	0.973	0.91	0.958	-	-	-	-	-
Ensemble	0.983	0.974	0.989	0.952	0.976	0.981	0.939	0.962	0.939	0.889

Notes: Blue represents the best performance

Table 8. Comparison with [12] using the REFUGE dataset

Method	Optic Disc Segmentation					Optic Cup Segmentation				
	Acc	Spe	Sen	Iou	DSC	Acc	Spe	Sen	Iou	DSC
[12]	-	-	0.981	0.924	0.960	-	-	0.921	0.807	0.89
Ensemble	0.984	0.989	0.979	0.969	0.985	0.986	0.960	0.964	0.969	0.934

Notes: Blue represents the best performance

4. Discussion

Research shows that semantic ensemble segmentation method successfully defines optic nerve head and optic cup with high precision. When multiple models are combined, their complementary qualities are used to increase segmentation performance. This is in contrast to when models are used individually. In order to accurately detect glaucoma, the ensemble technique is necessary since it effectively captures many features and makes segmentation more robust. The suggested method's potential for usage in a therapeutic context is a major edge. Diagnosing and tracking glaucoma disease requires exact separation of the optic disc and optic cup. Increased accuracy in glaucoma identification is a direct result of the suggested semantic segmentation method, which in turn allows ophthalmologists to make better, more timely decisions regarding patient care. Improving patient outcomes and reducing the likelihood of visual impairments associated with glaucoma can be achieved through this.

The segmentation findings are also improved in quality and reliability by using preprocessing methods including CLAHE, median filtering, and polar transformation. By enhancing contrast, Contrast Limited Adaptive Histogram Equalization (CLAHE) makes small features more visible. The use of median filtering, on the other hand, leads to cleaner, more accurate segmentation by effectively reducing noise and artifacts. On the other hand, the issue of data imbalance can be efficiently resolved by applying polar transform on retinal fundus images. Semantic ensemble segmentation, when combined with various preprocessing methods, makes the system more effective and robust.

Enhanced precision is a main benefit of deep learning ensembles. Combining various models improves segmentation accuracy by reducing the impact of individual model faults and increasing pattern capture. A further benefit of ensemble methods is their enhanced resistance to input data noise and fluctuations. More accurate segmentations and resistance to data uncertainties are both achieved by ensemble deep learning, which takes advantage of the relative merits of multiple models. Furthermore, ensemble approaches improve generalizability by combining forecasts from models trained on diverse data subsets or with independent architectures. With this more comprehensive picture of the patterns at work, performance on previously unseen data is significantly enhanced.

But there are drawbacks to think about as well. Ensemble deep learning is more computationally complex because it trains and maintains several models, which can be demanding on resources and time. Due to the necessity to store the predictions of numerous models during inference, memory utilization significantly increases. Ensemble methods can be quite sophisticated, which means there may be more processing overhead and more complicated integration procedures needed. Also, if the models in the ensemble are really correlated, overfitting could happen. At last, there is a lack of insight into the decision-making process due to the difficulty in understanding ensemble deep learning models. In conclusion, when it comes to semantic segmentation, ensemble deep learning has clear benefits in terms of accuracy and resilience. However, when putting ensemble methods into reality, one must carefully consider the computational needs, model complexity, and the possibility of overfitting.

One must acknowledge the research's shortcomings. Like any deep learning-based methodology, the suggested semantic ensemble segmentation strategy relies heavily on a big, diverse, and well annotated dataset to work.

Despite research on several color fundus image datasets, the suggested methodology may not be applicable due to variations in patient demographics, acquisition methods, or picture quality. More study with more diverse datasets and in real-world clinical situations is needed to validate and enhance the efficacy of the suggested strategy.

The computational complexity of training and deploying many models is another barrier. In comparison to training and inferring with individual models, ensemble techniques demand more computer resources and time. However, these challenges can be reduced through advancements in technology and optimization methods, which will make the practical use of semantic ensemble segmentation more likely.

Several potential directions for further research have been identified. Improving segmentation performance could be achieved by investigating different ensemble strategies. To further investigate the semantic ensemble segmentation method's potential therapeutic efficacy, it should be applied to more eye diseases and anatomical structures. A more complete and accurate evaluation of glaucoma progression may be possible with the integration of longitudinal data with multimodal imaging modalities like optical coherence tomography (OCT) or visual field testing.

5. Conclusions

Many different lines of inquiry might be considered for future studies. To enhance segmentation performance, trying out other ensemble methods could be a good idea. Investigating the semantic ensemble segmentation method's potential for use with additional eye disorders and anatomical features may also increase its therapeutic value. A more complete and accurate evaluation of glaucoma progression may be possible with the integration of longitudinal data with multimodal imaging modalities like optical coherence tomography (OCT) or visual field screening.

Ensemble methods enhance segmentation accuracy and resilience by effectively combining the benefits and compensating for the shortcomings of individual models. The quality and reliability of segmentation findings are enhanced by using preprocessing methods such as polar transformation, median filtering, and Contrast Limited Adaptive Histogram Equalization (CLAHE), according to research. The suggested methodology has shown promising results in experimental evaluations carried out on several datasets of color fundus images, suggesting that it may have therapeutic implications. There are substantial implications for early detection and treatment strategies connected to improved glaucoma identification using semantic ensemble segmentation. This could result in better patient outcomes and a decrease in the prevalence of visual impairment caused by glaucoma. Exploring various ensemble approaches, testing the suggested methodology on different datasets, and gauging its efficacy in actual clinical contexts are all potential avenues for future study in this field. To sum up, this study presents fresh chances to enhance glaucoma diagnosis and treatment by utilizing deep learning and semantic ensemble segmentation.

References

[1] Kwon, Y.H.; Fingert, J.H.; Kuehn, M.H.; Alward, W.L. Primary Open-Angle Glaucoma. N. Engl. J. Med. 2009, 360, 1113–1124.

[2] E. Dervisevic, S. Pavljasevic, A. Dervisevic, S.S. Kasumovic, Challenges in early glaucoma detection, Med Arch. 70 (2016) 203–207.

[3] Huang, X.; Sun, J.; Gupta, K.; Montesano, G.; Crabb, D.P.; Garway-Heath, D.F.; Brusini, P.; Lanzetta, P.; Oddone, F.; Turpin, A.; et al. Detecting Glaucoma from Multi-Modal Data Using Probabilistic Deep Learning. Front. Med. 2022, 9, 923096.

[4] Nath, M.K.; Dandapat, S. Techniques of Glaucoma Detection From Color Fundus Images: A Review. Int. J. Image Graph. Signal Process. 2012, 4, 44–51.

[5] Imtiaz, R.; Khan, T.M.; Naqvi, S.S.; Arsalan, M.; Nawaz, S.J. Screening of Glaucoma Disease from Retinal Vessel Images Using Semantic Segmentation. Comput. Electr. Eng. 2021, 91, 107036.

[6] Fu, H.; Cheng, J.; Xu, Y.; Zhang, C.; Wong, D.W.K.; Liu, J.; Cao, X. Disc-Aware Ensemble Network for Glaucoma Screening from Fundus Image. IEEE Trans. Med. Imaging 2018, 37, 2493–2501.

[7] Pascal, L.; Perdomo, O.J.; Bost, X.; Huet, B.; Otálora, S.; Zuluaga, M.A. Multi-Task Deep Learning for Glaucoma Detection from Color Fundus Images. Sci. Rep. 2022, 12, 1–10.

[8] Juneja, M.; Singh, S.; Agarwal, N.; Bali, S.; Gupta, S.; Thakur, N.; Jindal, P. Automated Detection of Glaucoma Using Deep Learning Convolution Network (G-Net). Multimed. Tools Appl. 2020, 79, 15531–15553.

[9] Natarajan, D.; Sankaralingam, E.; Balraj, K.; Thangaraj, V. Automated Segmentation Algorithm with Deep Learning Framework for Early Detection of Glaucoma. Concurr. Comput. Pract. Exp. 2021, 33, e6181.

[10] Tabassum, M.; Khan, T.M.; Arsalan, M.; Naqvi, S.S.; Ahmed, M.; Madni, H.A.; Mirza, J. CDED-Net: Joint Segmentation of Optic Disc and Optic Cup for Glaucoma Screening. IEEE Access 2020, 8, 102733–102747.

[11] Sreng, S.; Maneerat, N.; Hamamoto, K.; Win, K.Y. Deep Learning for Optic Disc Segmentation and Glaucoma Diagnosis on Retinal Images. Appl. Sci. 2020, 10, 4916.

[12] Liu, B.; Pan, D.; Song, H. Joint Optic Disc and Cup Segmentation Based on Densely Connected Depthwise Separable Convolution Deep Network. BMC Med. Imaging 2021, 21, 14.

[13] Zhao, X.; Wang, S.; Zhao, J.; Wei, H. Application of an Attention U-Net Incorporating Transfer Learning for Optic Disc and Cup Segmentation. Signal Image Video Process. 2020, 15, 913–921.

[14] Ganesh, S.S.; Kannayeram, G.; Karthick, A.; Muhibbullah, M. A Novel Context Aware Joint Segmentation and Classification Framework for Glaucoma Detection. Comput. Math. Methods Med. 2021, 2021, 2921737.

[15] Panda, R.; Puhan, N.B.; Mandal, B.; Panda, G. GlaucoNet: Patch-Based Residual Deep Learning Network for Optic Disc and Cup Segmentation Towards Glaucoma Assessment. SN Comput. Sci. 2021, 2, 99.

[16] Veena, H.N.; Muruganandham, A.; Senthil Kumaran, T. A Novel Optic Disc and Optic Cup Segmentation Technique to Diagnose Glaucoma Using Deep Learning Convolutional Neural Network over Retinal Fundus Images. J. King Saud Univ. Comput. Inf. Sci. 2022, 34, 6187–6198.

[17] Nazir, T.; Irtaza, A.; Starovoitov, V. Optic Disc and Optic Cup Segmentation for Glaucoma Detection from Blur Retinal Images Using Improved Mask-RCNN. Int. J. Opt. 2021, 2021, 6641980.

[18] Wang, L.; Gu, J.; Chen, Y.; Liang, Y.; Zhang, W.; Pu, J.; Chen, H. Automated Segmentation of the Optic Disc from Fundus Images Using an Asymmetric Deep Learning Network. Pattern Recognit. 2021, 112, 107810.

[19] Kumar, E.S.; Bindu, C.S. Two-Stage Framework for Optic Disc Segmentation and Estimation of Cup-to-Disc Ratio Using Deep Learning Technique. J. Ambient Intell. Humaniz. Comput. 2021, 1–13.

[20] Fu, Y.; Chen, J.; Li, J.; Pan, D.; Yue, X.; Zhu, Y. Optic Disc Segmentation by U-Net and Probability Bubble in Abnormal Fundus Images. Pattern Recognit. 2021, 117, 107971.

[21] Hu, Q.I.Z.; Hen, X.I.C.; Eng, Q.I.M.; Iahuan, J.; Ong, S.; Uo, G.A.L.; Ang, M.E.N.G.W.; Hi, F.E.I.S.; Hongyue, Z.; Hen, C.; et al. GDCSeg-Net: General Optic Disc and Cup Segmentation Network for Multi-Device Fundus Images. Biomed. Opt. Express 2021, 12, 6529–6544.

[22] Jin, B.; Liu, P.; Wang, P.; Shi, L.; Zhao, J. Optic Disc Segmentation Using Attention-Based U-Net and the Improved Cross-Entropy Convolutional Neural Network. Entropy 2020, 22, 844.

[23] Mohammed, A.; & Kora, R). A comprehensive review on ensemble deep learning: Opportunities and challenges. Journal of King Saud University-Computer and Information Sciences 2023, 35, 757–774.

[24] Zhang, Z.; Yin, F. S.; Liu, J.; Wong, W. K.; Tan, N. M.; Lee, B. H.; Wong, T. Y.. Origa-light: An online retinal fundus image database for glaucoma analysis and research. Annual international conference of the IEEE engineering in medicine and biology 2010, August, 3065–3068.

[25] Orlando, J.I.; Fu, H.; Barbossa Breda, J.; van Keer, K.; Bathula, D.R.; Diaz-Pinto, A.; Fang, R.; Heng, P.A.; Kim, J.; Lee, J.H.; et al. REFUGE Challenge: A Unified Framework for Evaluating Automated Methods for Glaucoma Assessment from Fundus Photographs. Med. Image Anal. 2020, 59, 101570.

[26] Batista, F. J. F.; Diaz-Aleman; T., Sigut, J.; Alayon, S.; Arnay, R.; Angel-Pereira, D. Rim-one dl: A unified retinal image database for assessing glaucoma using deep learning. *Image Analysis & Stereology 2020*, *39*(3), 161–167.

[27] Reza, A. M. Realization of the contrast limited adaptive histogram equalization (CLAHE) for real-time image enhancement. Journal of VLSI signal processing systems for signal, image and video technology 2004, 38, 35–44.

[28] Fu H; Cheng J; Xu Y; Wong DWK; Liu J; Cao X. Joint optic disc and cup segmentation based on multi-label deep network and polar transformation. IEEE Trans Med Imaging 2018, 37(7), 1597–605.

[29] Heba G. A. Yahia, Ashraf Mohamed Hemeida, Hameda A. Sennary. Image Segmentation Using Hybrid Optimization Algorithms: A Review. Aswan University Journal of Sciences and Technology 2022, 2.2, 86–104.

[30] Huang, H., Lin, L., Tong, R., Hu, H., Zhang, Q., Iwamoto, Y. & Wu, J. Unet 3+: A full-scale connected unet for medical image segmentation. ICASSP 2020-2020 IEEE international conference on acoustics, speech and signal processing (ICASSP), 2020, 1055–1059.

[31] Chen L-C, Papandreou G, Schroff F, Adam H. Rethinking atrous convolution for semantic image segmentation. 2017. arXiv:1706.05587.

[32] Zhao H, Shi J, Qi X, Wang X, Jia J. Pyramid scene parsing network. Proceedings of the IEEE conference on computer vision and pattern recognition 2017, 2881–2890.

[33] Pal, D., Reddy, P. B., & Roy, S. Attention UW-Net: A fully connected model for automatic segmentation and annotation of chest X-ray. Computers in Biology and Medicine, 2022, 150, 106083.

[34] Hussein, Aida, et al. Medical Image Segmentation Using Deep Learning. Aswan University Journal of Sciences and Technology 2023, 3.1, 87–108.