RESEARCH ARTICLE

Automated Tool Support for Glaucoma Identification With Explainability Using Fundus Images

THISARA SHYAMALEE[1], **DULANI MEEDENIYA**[1], **(Senior Member, IEEE)**, **GILBERT LIM**[2], **(Member, IEEE), AND MIHIPALI KARUNARATHNE**[3]

[1]Department of Computer Science and Engineering, University of Moratuwa, Moratuwa 10400, Sri Lanka
[2]SingHealth AI Health Program, SingHealth Group, Singapore 168582
[3]National Eye Hospital, Colombo 01000, Sri Lanka

Corresponding author: Dulani Meedeniya (dulanim@cse.mrt.ac.lk)

ABSTRACT Glaucoma is a progressive eye condition that causes irreversible vision loss due to damage to the optic nerve. Recent developments in deep learning and the accessibility of computing resources have provided tool support for automated glaucoma diagnosis. Despite deep learning's advances in disease diagnosis using medical images, generic convolutional neural networks are still not widely used in medical practices due to the limited trustworthiness of these models. Although deep learning-based glaucoma classification has gained popularity in recent years, only a few of them have addressed the explainability and interpretability of the models, which increases confidence in using such applications. This study presents state-of-the-art deep learning techniques to segment and classify fundus images to predict glaucoma conditions and applies visualization techniques to explain the results to ease understandability. Our predictions are based on U-Net with attention mechanisms with ResNet50 for the segmentation process and a modified Inception V3 architecture for the classification. Attention U-Net with modified ResNet50 backbone obtained 99.58% and 98.05% accuracies for optic disc segmentation and optic cup segmentation, respectively for the RIM-ONE dataset. Additionally, we generate heatmaps that highlight the regions that impacted the glaucoma diagnosis using both Gradient-weighted Class Activation Mapping (Grad-CAM) and Grad-CAM++. Our model that classifies the segmented images achieves accuracy, sensitivity, and specificity values of 98.97%, 99.42%, and 95.59%, respectively, with the RIM-ONE dataset. This model can be used as a support tool for automated glaucoma identification using fundus images.

INDEX TERMS Artificial intelligence, classification, explainability, segmentation, support tool, trustworthiness.

I. INTRODUCTION

Glaucoma is a degenerative eye disease that can lead to permanent blindness. Progressive optic neuropathy, or loss of vision, is the most common symptom of glaucoma. In glaucoma, the optic disc and optic cup of the eye, are destroyed due to permanent damage to the optic nerve due to an increase in fluid or pressure in the eye. This causes the optic cup becomes larger with respect to the optic disc, which increases the cup-to-disk ratio (CDR) [1]. Glaucoma

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

typically advances unnoticed until the optic nerve suffers irreparable damage. Common signs include tunnel vision, blind spots, and blurry vision. Individuals suffer less from irreversible disease progression that results in blindness if glaucoma is detected and treated early [2]. Glaucoma affects between 3-5% of individuals around the world between the ages of 40 and 80. In particular, it was anticipated that 64 million individuals aged 40 to 80 were affected by glaucoma worldwide in 2013, and that figure is projected to rise to 78 million in 2020 and 111.8 million in 2040 [3].

In clinical practice, several techniques are used to diagnose glaucoma. One of the common diagnostic practices is to

examine the dilated eye using ophthalmoscopy. Here, fundus images of the retina are used to assess the optic nerve in the diagnosing process [4], [5]. Generally, ophthalmologists manually review fundus images to identify the signs that lead to glaucoma conditions. However, the entire image-capturing and manual screening procedure consumes time and effort. Additionally, misdiagnosis can happen with practitioners with less experience.

Computer-aided applications can be used as a support tool to identify glaucoma conditions by analyzing the fundus images. The evolution of Artificial Intelligence (AI) and Deep Learning (DL) in the field of medical imaging has provided efficient and effective solutions [6], [7], [8]. However, there are several challenges and gaps in the existing literature. For example, the available annotated datasets for glaucoma detection are often limited in size and may not cover the diverse population and severity levels encountered in clinical practice [9], [10]. Small datasets can hinder the development of robust and generalizable models [11]. The imbalance between normal and glaucomatous cases in datasets poses challenges for training models [12]. In addition, due to the black-box nature of such models, most of the users have trustworthiness issues in using such applications. Thus, there is an emerging requirement to increase confidence to use computational models by incorporating the explainability of the proposed models to build trust among clinicians and ensure transparency in decision-making [13], [14], [15], [16]. Moreover, existing models may not fully leverage domain-specific clinical knowledge. Collaboration between machine learning researchers and ophthalmologists is crucial to incorporate expert insights into the model design and interpretation. Furthermore, deploying glaucoma detection models in real-time or point-of-care settings requires addressing challenges related to computational efficiency, model size, and integration with existing clinical workflows.

In our previous studies [5], [17], [18], we presented a model to segment and classify fundus images considering RIM-ONE and ACRIMA datasets. We applied specific preprocessing steps such as CLAHE and median filter and provided a comparison study among several segmentation and classification models and identified the best-performing models. Extending our work, as a novel contribution, this study utilizes the identified best-performing DL-based models and extends the approach for five different public datasets to identify the generalizability of the model. We use five different datasets namely RIM-ONE [19], ACRIMA [20], REFUGE [21], DRISHTI-GS1 [22], and ORIGA [23] for the proposed computational framework. We apply segmentation with the ResNet50 model combining Attention U-Net architecture and classification with the Inception V3 to predict the images. These models were selected as they have shown high performance in our previous studies. Also, we provide tool support with a web application to identify and explain glaucoma conditions using fundus images.

Additionally, as another contribution towards the trustworthiness of the system, we provide Gradient-weighted Class Activation Mapping (Grad-CAM) [24], and Grad-CAM++ [25], based on visualization explainability for the obtained results as a comparative study. The generated heatmaps give this transparency, defining the regions of interest of the fundus image that impact the model's decision-making process. Consequently, domain specialists such as optometrists, ophthalmologists, and eye surgeons can be informed of the reasoning behind the predicted results. In this study, we aim to make the decision-making process of DL models more visible, thereby increasing the models' dependability. Moreover, we present a web application named GlaucoCare, as a support tool that can be used by medical practitioners [26]. Finally, the results are validated by subject experts and a system usability study (SUS) is conducted to show the usability of the proposed application among domain experts using surveys and interviews. We have made available all the resources in publicly accessible platforms [27]. Therefore, the proposal of a generic approach to segment and classify the fundus images in different datasets, which leads to high performance and provides an explainability mechanism can be considered a novelty of this study.

The paper is structured as follows. Section II explores the related studies and Section III describes the proposed methodology. Section IV analyses the obtained results and Section VI concludes the paper.

II. BACKGROUND
A. OVERVIEW OF GLAUCOMA

As shown in Figure 1, the fundus image of an eye is an ocular record, which includes the optic disc, optic cup, and blood vessels that capture the appearance of the retina. Generally, fundus images are used to track the progression of eye conditions with glaucoma, by ophthalmologists and medical practitioners. The area where blood vessels and nerve fibers enter the retina is identified as the Optic Disc (OD) also known as the optic nerve head (ONH) in a fundus image. The brightest area of the retina's center, known as the optic cup (OC), is located inside the optic disc. The area between the optic disc and the optic cup is known as the neuroretinal rim and the rim width ratios vary according to the shape of the optic disc [2], [4]. The most reason for identifying glaucoma condition is the cupping issue, where the optic cup is getting enlarged. This is due to high pressure in the eye damaging the optic nerve, which results in the loss of individual nerve cells. This causes a subsequent increase in the size of the cup, also called cupping.

B. OVERVIEW OF EXPLAINABILITY

Generally, the DL-based classification models are difficult to explain and hard to understand by non-experts in that domain. The evolving explainable artificial intelligence (XAI) techniques provide support to explain the reason behind the decisions made by the DL models [16]. XAI enables transparency, interpretability, and explainability for
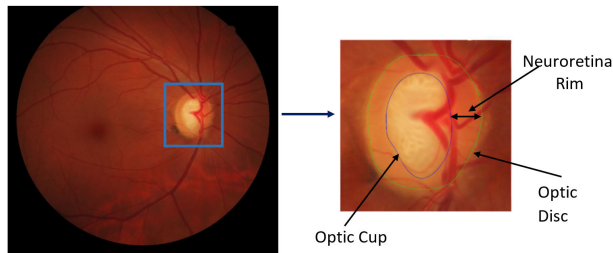
FIGURE 1. Overview of a fundus image.

the results of a DL model visually or textually. For instance, activation maps give a visual interpretation of the results by showing the region of interest of the input image that leads to the classification results [28], [29]. Moreover, Grad-CAM produces a coarse localization map by using the gradients of the output passes to the final convolutional layer to produce a coarse localization [24], [30].

Grad CAM++ is a generalized formulation of GRAD-CAM that solves some of the issues of Grad-CAM, which needs a network modification and estimates smaller weights when the response is small and the area of the response is small [25]. This provides better visual explanations of CNN model predictions in terms of better object localization and explaining occurrences of multiple object instances in a single image. The weights used to generate visual explanations for the corresponding class label are generated from a weighted combination of the positive partial derivatives of the last convolutional layer feature maps with respect to a specific class score.

These heatmaps highlight the significant regions of the image that were focused on making the prediction [1], [31]. For example, such indication can be interpreted as the model has focused more on the middle of the optic disk and a larger optic cup has a high correlation with glaucoma cases. Consequently, the ophthalmologists can understand the reasons for getting such output and it will increase the trustworthiness of the system. Therefore, automated tool support is useful for consistent and reliable decision-making with less time and effort.

C. RELATED STUDIES

A wide range of literature is available on fundus image segmentation and classification for glaucoma identification. Considering the latest studies, Kumar et al. [32], have proposed a model that effectively segments the optic disc and optic cup from retinal fundus images using U-Net++. The segmented images are fed into a hybrid deep learning architecture with ResNet and Gated Recurrent Unit (GRU) layers to detect glaucoma with minimal computational overhead accurately. The proposed solution has shown an accuracy of 98.75%, sensitivity of 99.21%, and a specificity of 98.29%.

In addition, some studies have proposed new CNN architectures to classify fundus images. For example, Sonti and Dhuli [33], have presented KR-NET for the DRiFT2

database result in an accuracy of 96.4%, sensitivity of 95.0%, and specificity of 98.5%. Similarly, Raza et al. [34], have proposed FEDS-Net for segmentation with several datasets showing promising results. Another, parameter-efficient model called AlterNet-K is presented by D'Souza et al. [35], by combining the strengths of ResNets and multi-head self-attention. They have achieved an accuracy of 91.6% over Rotterdam EyePACS AIROGS dataset.

Moreover, several types of research have been published on glaucoma diagnosis based on interpretable techniques. For instance, Liao et al. [36], have presented a novel clinical interpretable DL-based framework named EAMNet, which is based on convolutional network architecture, focusing on classifying fundus images for glaucoma detection. They have used transparent interpretation by highlighting the distinct regions recognized by the network. The experiment was performed on the ORIGA dataset and showed an AUC of 0.88. They have used Class Activation Map (CAM) to highlight important regions of the fundus image. Generally, aspects such as bleeding, notch, and structural variation of blood vessels are also crucial clinical evidence for glaucoma diagnosis. These features are not always visible on the fundus images of glaucoma patients. This approach was focused on both the parameters of the optic disc and cup and some other features to diagnose glaucoma. However, the proposed model may not perform well in blurred images.

In another study, Das et al. [37], have introduced an innovative cascaded attention mechanism (CAM) based model named CA-Net, designed to facilitate effective multistage classification of glaucoma. The proposed CA-Net consists of a backbone network, CAM to obtain refined feature maps, and a classifier. Three datasets, Harvard Dataverse V1, LMG, and RIM-ONE Extended, were used for evaluation and Grad-CAM and Grad-CAM++ techniques were used for visualize the interpretability. The suggested CA-Net model has shown high performance with 92.51%, 93.85% and 91.76% of accuracy, F1-Score and AUC, respectively. Moreover, Deperlioglu et al. [38], have suggested an Explainable AI (XAI) and CNN-based glaucoma screening method using three different datasets namely Drishti-GS, ORIGA-Light, and HRF. CAM was used as the XAI technique to create the heatmap for fundus image analysis. The CNN architecture on the ORIGALight dataset provided the best performance.

Another study by Neto et al. [1], have addressed glaucoma screening based on classification and segmentation using different datasets of fundus images. They have utilized Xception, ResNet152 V2, and Inception ResNet V2, as a comparative study. Additionally, activation maps were generated using the Grad-CAM technique for the interpretability and explainability of the results. They have used U-Net architecture combining Inception ResNet V2 and Inception V3 models as the backbone of their segmentation process. The Xception model with the REFUGE dataset has shown the highest accuracy of 97% in the classification process. Another novel CNN model is presented by Tian et al. [39], for glaucoma classification using Global and Class Attention

Blocks (GC-Net). It utilized Inception V3 as the backbone and incorporated two attention blocks, namely the global attention block (GAB) and the class attention block (CAB). The model takes fundus images around the optic disc as inputs and predicts the categories of glaucoma disease. They have achieved an accuracy of 78.01%, specificity of 94.82%, AUC of 90.68%.

Another pre-trained CNN and classifier fusion based glaucoma classification approach was proposed by Velpula and Sharma [40]. The suggested model employed five pre-trained CNN models, namely ResNet50, AlexNet, VGG19, DenseNet-201, and Inception-ResNet-v2. The evaluation encompassed four publicly available datasets ACRIMA, RIM-ONE, Harvard Dataverse (HVD), and Drishti. A classifier fusion approach was implemented, consolidating decisions from all CNN models through a maximum voting-based strategy. Across datasets, the proposed model demonstrated remarkable performance. For the ACRIMA dataset, it achieved a perfect area under the curve (AUC) of 1 and an accuracy of 99.57%. The HVD dataset yielded a notable AUC of 0.97 and an accuracy of 85.43%. Drishti and RIM-ONE datasets exhibited accuracy rates of 90.55% and 94.95%, respectively. Model interpretation involved attribution-based methods, including activations and gradient class activation map, as well as perturbation-based methods such as locally interpretable model-agnostic explanations and occlusion sensitivity, generating informative heatmaps for various image sections influencing model predictions.

Another fundus images-based Glaucoma diagnosis and localization approach was proposed by Kim et al. [15]. They have also considered a comparative study using VGG-16, ResNet-152, and Inception-v4, and Grad-CAM was used to localize the glaucomatous areas of the fundus image. The ResNet-152-M model has outperformed the other CNN variants in terms of performance with 96% of accuracy. In addition, the researchers designed a web-based application (Medinoid), providing a decision, diagnostic confidence score, and suspected area for a given input fundus image.

Furthermore, Juneja et al. [41], have presented a deep CNN architecture (CoG-Net) based glaucoma prediction approach. They have used four publicly available datasets namely Drishti-GS, RIM-ONE, REFUGE, and ACRIMA. An activation map has been utilized as an explainable technique to identify the important regions of the fundus image responsible for classification. The suggested CoG-Net model has performed the best result with 95.3% of accuracy.

In our previous studies, first, we conducted a comparative study on classifying fundus images using different Convolutional Neural Networks (CNNs) [17]. We used modified architectures of Inception V3, Visual Geometry Group 19 (VGG19), and Residual Neural Network 50 (ResNet50). The highest accuracy is shown as 98.52% with Inception V3 for the ACRIMA dataset. Then, we applied segmentation based on attention U-Net with different backbone encoders [5].

Here, the best accuracy for optic disk segmentation is achieved as 99.53% for the RIM-ONE dataset.

Among the existing studies, we believe that still the models can be designed to increase the associated performance metrics, as the application is highly dependable. Especially, accuracy, sensitivity, specificity, and trustworthiness play a significant role in the computer-based automated models that interface the medical domain. Accordingly, in this study, we present two XAI methods over our segmentation and classification approach. Here, we aim to show an approach to glaucoma identification that leads to higher performance compared to the existing studies, with explainability features to improve confidence in using such systems. The proposed model is deployed as a web application; thus the practitioners can be used it as a support tool.

III. GALUCOCARE FRAMEWORK
A. DATASET DETAILS
In order to provide a generalized solution, we consider five different fundus image datasets for this study. They are RIM-ONE [19], ACRIMA [20], REFUGE [21], DRISHTI-GS1 [22], and ORIGA [23] datasets. A summary of the datasets is shown in Table 1. Considering these related studies that have used these datasets, RIM-ONE is the most-used dataset in the literature.

TABLE 1. Summary of fundus image datasets.

Dataset	Images	Related Studies
RIM-ONE [12]	942	[42], [43], [44], [45], [46], [47], [48], [49], [50]
ACRIMA [51]	705	[42], [43]
ORIGA [11]	650	[52], [42], [36]
REFUGE [10]	1200	[42]
Drishti-GS [9]	101	[52], [42], [45], [47], [49]

The Retinal Image database for Optic Nerve Evaluation (RIM-ONE) is the publicly available fundus image dataset [19]. It has three releases including monosporic images and stereoscopic images. The first release consists of 169 monosporic images collected from 51 glaucomatous images and 118 normal images. The second release contains 455 monosporic images with 200 glaucomatous images and 255 normal images. The third version includes 318 monosporic images collected from 148 glaucomatous images and 170 normal images. Accordingly, we obtained 942 fundus images, where 399 are glaucomatous and 543 are normal. Moreover, the implementation of the segmentation process requires an image mask of each fundus image. Optic disc and cup masks are consistent in RIM-ONE version 3, whereas optic disc masks are only available in the RIM-ONE version 1 dataset. We created masks from the remaining dataset using www.apeer.com to complete the segmentation procedure. The mask-creation process is described later in this section.

ACRIMA is another publicly available dataset that includes 705 fundus images collected from 396 glaucomatous images and 309 normal images [20]. Two glaucoma

specialists with eight years of combined experience have annotated every image from the ACRIMA database. Further, segmented optic disc masks and optic cup masks are not offered in the ACRIMA database. Therefore, we created masks using www.apeer.com.

The Online Retinal Fundus Image Dataset for Glaucoma Analysis and Research (ORIGA) database includes 650 fundus images with 168 glaucomatous images and 482 non-glaucomatous images collected from the Singapore Malay Eye Study (SiMES) [21]. ORIGA dataset consists of mask images of the optic cup and optic disc for the entire dataset. REFUGE is another large publicly available dataset which composes of 1200 fundus images collected from 1080 non-glaucoma images and 120 glaucomatous images [22]. REFUGE also includes image masks for the entire fundus images in the dataset.

A total of 101 images, with 31 normal images and 70 glaucomatous images, are available in the DRISHTI-GS1 dataset [23]. The dataset was divided into training and testing sets including 50 and 51 images. Image masks are only provided for the training sets and have used www.apeer.com to create image masks for the remaining images. It is a cloud-based application and the associated annotation tool is a part of the APEER platform that supports automated image processing. The APEER application uses a background class that does not contain objects belonging to any class predicted by the model. It is used to differentiate the regions that do not consist of classes which should be segmented and the regions that are not annotated. We can use partial annotations, as the non-annotated regions are disregarded in training the model, it allows partial annotations. In this manual mask generation process, first, we used the brush tool to annotate the masks in the region. We segmented the mask for both optic disk and optic cup separately, for the RIM-ONE, ACRIMA, and DRISHTI-GS1datasets only, which do not contain the corresponding masks. While keeping the property "autofill holes" enabled, the brush tool is used again to draw the background around the annotated objects. It automatically crops the background when other annotations exist and ensures that any of the non-annotated items are not covered by the background class. Figure 2 shows the output masks generated for sample input images using the APEER application.

(a) Input Image 1   (b) Output Image 1   (c) Inpute Image 2   (d) Output Image 2
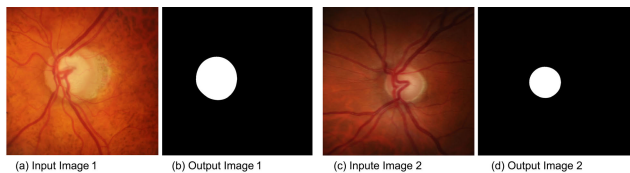
FIGURE 2. Sample inputs and output masks generated using APEER.

Initially, we split each of the dataset into 70:15:15 ratio for training, testing and validation, respectively. By considering the number of images in glaucoma and healthy classes, we applied different augmentation techniques separately for each subset to preserve class balancing property. The final

number of training, testing and validation images used for each dataset are stated in Table 2.

TABLE 2. Summary of training, validation, and test images of used datasets.

Dataset	Training Images	Validation Images	Testing Images
RIM-ONE [12]	4512	970	976
ACRIMA [51]	3193	718	724
ORIGA [11]	775	228	231
REFUGE [10]	1380	300	300
DRISHTI [9]	225	90	87

B. PROCESS VIEW

The High-level architecture of the proposed methodology is illustrated in Figure 3. The main components of our approach are the segmentation and classification process, the interpretable approach, and the GlaucoCare Web app. In the field of medical image analysis, numerous pre-processing and feature extraction techniques have been employed [42]. Pseudocode of proposed architecture is shown in Algorithm 1.

Algorithm 1 Pseudo-Code of Overall Process

FUNCTION main():
  Dataset ← load *RIM-ONE*;
  Preprocessed_dataset ← preprocess *(Dataset, 2)*;
  augmented_dataset ← augment *(preprocessed_dataset, 8)*;
  ForAll segmentation_process **do**
    x_a ← imp_model *(U-Net, ResNet)*
    x_b ← load_model *(x_a, augmented_dataset)*
  RETURN CDR, segmented_images
  ForEach classification_process **do**
    y_a ← imp_model *(Inception V3, segmented_images)*;
    y_b ← load_model *(y_a, augmented_dataset)*;
    y_c ← apply_XAI *(y_b, 2)*;
  RETURN class_prediction, Heatmap
End FUNCTION

Mainly, in this study we considered five different datasets to support generalizability and produced explainable visualizations, together with a support tool that can be utilized in practice.

C. IMAGE PRE-PROCESSING

We used two main pre-processing techniques namely contrast limited adaptive histogram equalization (CLAHE) and the median filter. CLAHE is a well-known image quality and contrast enhancement technique [53], [54]. However, because noise degrades image quality, the median filter technique is used to reduce noise while preserving the image's edge features. In order to address the issues of data imbalance, and overfitting, and to increase the number of images for model training, we used various augmentation techniques [55], [56].

D. SEGMENTATION AND CLASSIFICATION PROCESS

Segmentation-based research has focused on fully convolutional networks (FCN). The U-Net architecture is suitable for biomedical image segmentation because it is based on the
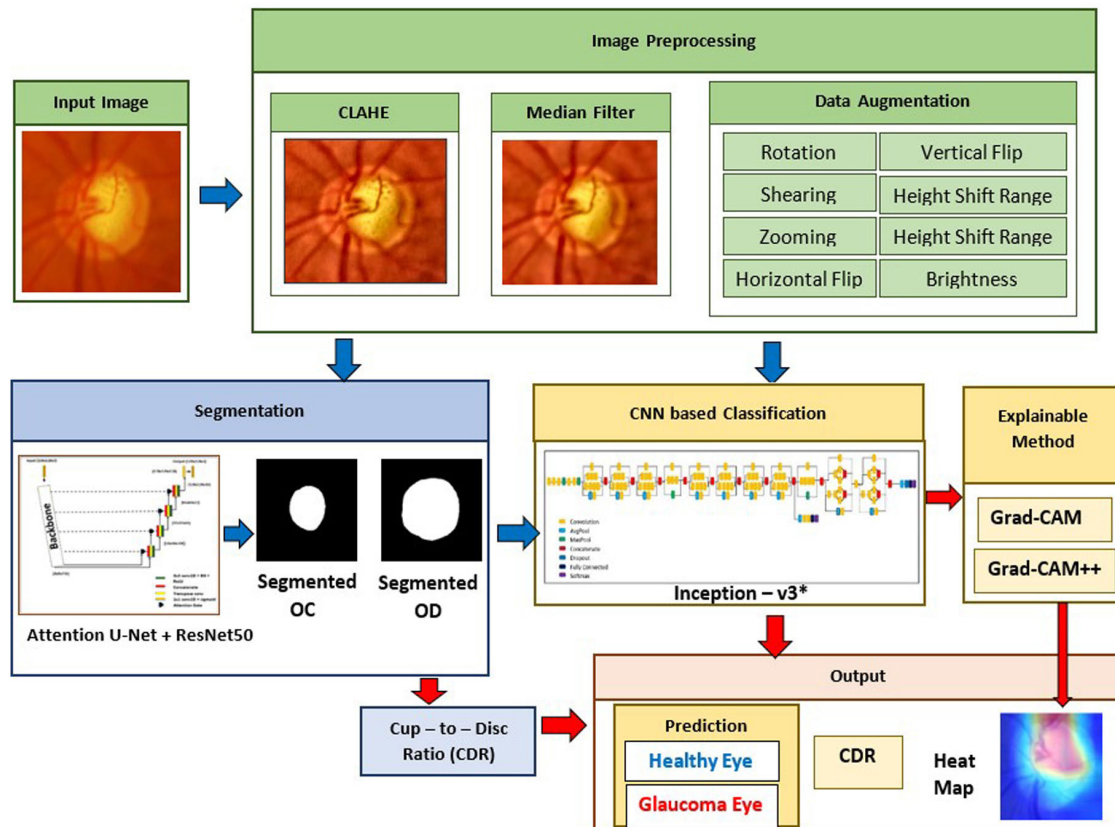
FIGURE 3. Overall process of the proposed approach.

FCN architecture [57]. In our framework, the Attention U-Net model has been slightly modified from the traditional model to achieve the best results. It consists of an encoder, a decoder, and an attention gate (AG) at each level's skip connection. We replaced the original encoder of the attention U-Net with a pre-trained network, namely ResNet50 as the backbone to determine the best segmentation performances [5]. The selection of the ResNet50 model was based on its in-built features. The skip connections in the ResNet50 architecture act as gradient superhighways, allowing the gradient to move freely. In this process, we performed both optic cup segmentation (OCS) and optic disc segmentation (ODS). The images were divided into 70:15:15 training, testing, and validation sets during the U-Net segmentation process.

The primary objective of the classification process is to generate accurate results. Therefore, we used the pre-processed original images together with the segmented masks generated by the ResNet50 backbone as input images for the classification process, which offers the highest accuracy in the segmentation process. We have used modified Inception V3 architecture to classify the segmented results. The modified Inception V3 model was selected as it has shown high performance in our previous study which was also based on Glaucoma detection [17]. That study, we conducted a comparative study on classifying fundus images using different Convolutional Neural Networks

(CNNs). We used modified architectures of Inception V3, VGG19, and ResNet50. The highest classification accuracy of 98.52% is given using the modified Inception V3 model. Instead of the top dense layers, the modified Inception V3 model has three additional layers. For instance, the global average pooling (GAP) layer, was used to minimize the parameters. A 512-unit dense layer and a Softmax layer were added. In the Softmax layer, two neurons were used to indicate the binary class categorization to distinguish between Glaucoma and Normal classes. As part of our future plan, we intend to enhance this tool for detect types of glaucoma disease, including open-angle glaucoma and closed-angle glaucoma. To facilitate this expansion, we employed a SoftMax layer in our model. Finally, we used a dropout layer with a rate of 0.7 to reduce model overfitting. Additionally, the auxiliary classifier of the Inception V3 model handled the vanishing gradient problem. The model is trained with the Adam optimizer. In the segmentation process, we performed fine-tuning on the pre-trained ResNet50 by initially modifying the last two layers, specifically the dense and SoftMax layers. Subsequently, we replaced the fully connected layer with another densely connected layer featuring 256 units, integrated into the same pre-trained network. ResNet50 architecture is trained for 150 epochs at a learning rate of 0.001 with the stochastic gradient descent (SGD) optimizer [5].

In the training of the model, several hyper-parameters played a pivotal role in shaping its performance. The learning rate, which determines the step size during optimization, was carefully tuned to strike a balance between convergence speed and stability. The batch size, representing the number of data points processed in each iteration, and the number of epochs, indicating the total passes through the dataset, were optimized to efficiently train the model while managing computational resources. The choice of loss function, crucial for guiding the optimization process, was a strategic decision based on the nature of the task, whether regression or classification. Notably, two prominent optimizers, Stochastic Gradient Descent (SGD) and Adam, were employed to update the model parameters during training. The use of multiple optimizers allowed for a comparative analysis of their impact on the model's convergence and final performance, providing insights into the dynamics of the optimization process.

The prediction of glaucoma and normal classes is mainly based on this segmentation and classification process. Additionally, we computed the CDR using the segmented images, and the obtained results complied with the results of the classification model. Therefore, the two outputs namely the prediction class and CDR value, reflect the same outcome. In this study, our contribution is focused on the explainability and tool support for glaucoma identification, which increases the usability and trustworthiness of the proposed approach, as described in the following section.

The mathematical background behind the used theoretical concepts can be defined as follows. The learning rate is a hyperparameter that determines the step size at each iteration while moving toward a minimum of a loss function during training. It influences the convergence and stability of the training process. If the learning rate is too high, the optimization algorithm might overshoot the minimum and fail to converge. On the other hand, if the learning rate is too low, the algorithm may take a long time to converge, or it might get stuck in a suboptimal solution. The learning rate is multiplied by the gradient of the loss with respect to the model parameters to determine the size of the parameter updates. The update rule for a parameter w at iteration t can be expressed using the learning rate as in equation (1), where $\nabla L\left(w_t\right)$ is the gradient of the loss function with respect to the parameter (w_t) and α is the learning rate.

$$w_{t+1} = w_t - \alpha . \nabla L(w_t) \qquad (1)$$

Stochastic Gradient Descent (SGD) is a fundamental optimization algorithm that updates the model parameters using a small, randomly chosen subset of the training data at each iteration. The update rule for the SGD optimizer is based on the gradient of the loss function with respect to the model parameters as given in equation (2), where (Θ_t) represents the model parameters at iteration t, α is the learning rate, determining the size of the step taken during the optimization process and $\nabla J_t\left(\Theta_t\right)$ is the gradient of the loss function (J_t)

with respect to the model parameters at time t.

$$\Theta_{t+1} = \Theta_t - \alpha . \nabla J_t \Theta_t \qquad (2)$$

The SoftMax activation function transforms the raw output scores (logits) into a probability distribution over multiple classes. It is defined for each class i in the output layer and is computed as in equation (3), where Z_i is the raw score (logit) associated with class i and K is the total number of classes.

$$Softmax(Z)_i = e^{z_i} / \sum_{i=1}^{k} e^{z_j} \qquad (3)$$

Global Average Pooling (GAP) layer is used for spatial dimension reduction. It replaces the traditional fully connected layers in the final part of a CNN while reducing the spatial dimensions (width and height) to a single value per feature map. This helps reduce the total number of parameters in the network, mitigates overfitting, and enhances model interpretability. This computes the average value of each feature map across its entire spatial dimensions as in equation (4), for a feature map F with dimensions $H \times W \times C$, where H is the height, W is the width, C is the number of channels, $F_{i,j,c}$ is the activation at position (i,j) in channel c of the feature map and $GAP(F)_C$ is the averaged value for channel c.

$$GAP(F)_C = (1/H * W) \sum_{i=1}^{H} \sum_{j=1}^{w} F_{i,j,c} \qquad (4)$$

Dropout is a regularization technique that prevents overfitting. It randomly "drops out" (sets to zero) a proportion of the neurons or units during training, forcing the model to learn more robust and generalized representations. This helps prevent the model from becoming too dependent on specific neurons and improves its ability to generalize to new, unseen data. The dropout operation is applied independently to each neuron in the layer during training as in equation (5), where x is the input value (activation) to the dropout layer and p is the dropout rate, representing the probability of setting a neuron to zero during training.

$$Dropout(x) = \begin{cases} x * 1/(1-p), & \text{with probability 1-p} \\ 0, & \text{with probability p} \end{cases}$$

$$(5)$$

E. EXPLAINABLE APPROACH
A clinically interpretable deep learning model has been suggested to obtain an accurately automated glaucoma diagnosis while providing a more transparent interpretation by highlighting different regions to aid the diagnosis [58]. Generally, Deep Neural Networks are opaque in nature and their prediction cannot be detected by humans. This can be addressed by incorporating explainable and interpretable techniques into the classification models [24], [25], [56].

In order to make CNN-based models more transparent and understandable, we applied both Grad-CAM [24] and Grad-CAM++ [25], which generates a coarse localization map by
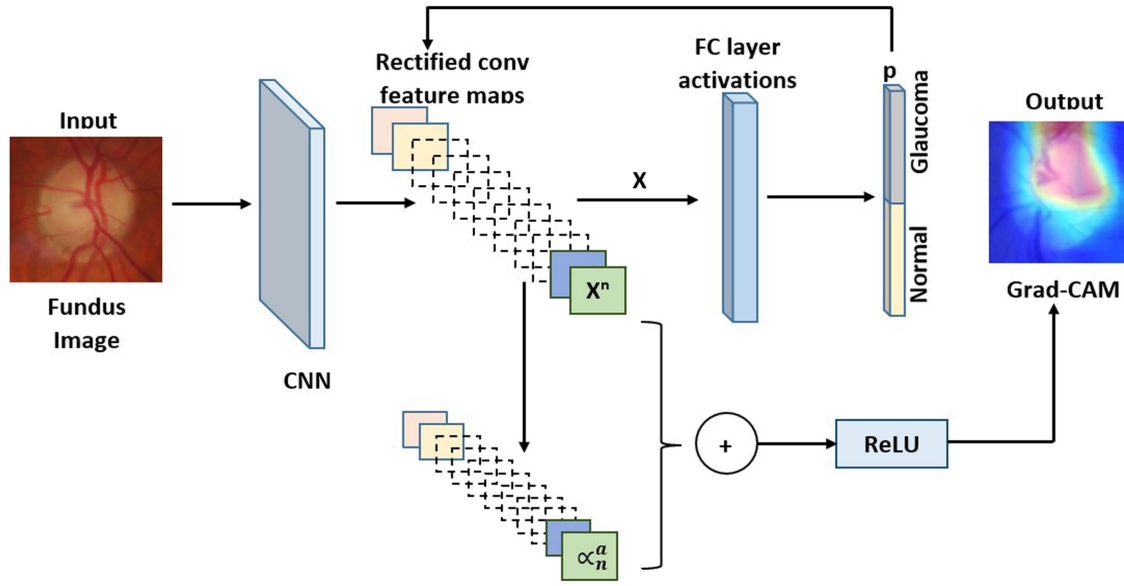
FIGURE 4. Overview of the Grad-CAM process.

using the gradients of any target concept flowing into the final convolutional layer to highlight key areas in the image for concept prediction. Figure 4 shows the overall process view of the Grad-CAM technique. Grad-CAM incorporates gradient weights to generalize the Class Activation Mapping (CAM) technique [24]. The most relevant regions discovered by the suggested CNN models are visually represented by this algorithm. CAM displays activations in the feature maps of the final convolutional layer using the expected output and backpropagation to demonstrate what CNN looks at when recognizing the input. Grad-CAM is combined with smooth visualizations in the recommended method to produce a high-resolution class-discriminative visualization. Grad-CAM has addressed the issues with Class Activation Maps [28], the most popular interpretable technique, such as the need for re-training the model and the limitations of having a specific architecture. The output of Grad-CAM can be overlaid on the input image.

In order to create the final localization map of the glaucomatous regions ($L_{Grad-CAM}^{C}$), first, compute the gradients of the score class 'c' (y^c) with concerning the feature maps (A^k) of the final convolutional layer ($\frac{\partial y^c}{\partial A_{ij}^k}$), having width w and height h, where $k \epsilon \{1, 2, \ldots\ldots, K\}$ and K indicates the total feature maps. The result ($\frac{\partial y^c}{\partial A_{ij}^k}$) is then fed to the global-average pooling procedure over the width and height dimensions (i and j), which yields the weights (W_k^c) for the glaucoma class as stated in (6) [24].

$$W_k^c = \frac{1}{w * h} \sum_i \sum_j (\frac{\partial y^c}{\partial A_{ij}^k}) \quad (6)$$

The Rectified Linear Unit (ReLU) only selects the positive activations after receiving an input consisting of a linear combination of the weights and the relevant feature maps. The final localization map of the glaucomatous regions is obtained using (7).

$$L_{Grad-CAM}^{C} = ReLU(\sum_k W_k^c A_{ij}^k) \quad (7)$$

Table 3 lists the processing steps of Grad-CAM and Grad-CAM++ [24], [25].

IV. RESULTS ANALYSIS AND DISCUSSION
A. SEGMENTATION AND CLASSIFICATION RESULTS
We implemented the models with a 1×10^{-4} learning rate, 150 epochs, and batch size of 8 with ADAM optimizer. For RIM-ONE and ACRIMA datasets we have used a dropout of 0.7 and 512 dense layers. For ORIGA, REFUGE, and DRISHTI datasets, we used a dropout of 0.8 and 1024 dense layers to obtain high results, after several experiments with hyperparameter tuning.

To evaluate the effectiveness of the fundus image segmentation process, we used a variety of metrics. The terms true-positive (TP), false-positive (FP), false-negative (FN), and true-negative (TN) instances can be defined as follows.

- TP: Glaucoma images correctly predicted as Glaucoma by the model
- FP: Normal images incorrectly predicted as Glaucoma by the model
- FN: Glaucoma images incorrectly predicted as Normal by the model
- TN: Normal images correctly predicted as Normal by the model

Our predictions are derived from the U-Net with ResNet50 based segmentation and a modified Inception V3 for classification utilizing the RIM-ONE dataset. To demonstrate the generalizability of our study, as illustrated in Table 4, the evaluation encompassed diverse datasets, including

TABLE 3. Process of Grad-CAM and Grad-CAM++.

Grad-CAM	Grad-CAM++
A. Find the class score (y^c) using the trained model. $y^c = \sum_k W_k^c \sum_i \sum_j A_{ij}^k$	A. Find the class score (y^c) using the trained model. $y^c = \sum_k \left(\left\{ \sum_{a,b} \alpha_{ab}^{kc}.ReLU \left(\frac{\partial y^c}{\partial A_{ab}^k} \right) \right\} \left[\sum_{i,j} A_{ij}^k \right] \right)$
	B. Calculate positive partial derivatives of (y^c) w.r.t convolution layers outputs(feature maps) without ReLU.
B. Calculate partial derivatives of (y^c) w.r.t. convolutional layers outputs (feature maps). $\left(\frac{\partial y^c}{\partial A_{ij}^k} \right)$	$\frac{\partial y^c}{\partial A_{ij}^k} = \sum_{a,b} \alpha_{ab}^{kc}.\frac{y^c}{\partial A_{ab}^k} + \sum_{a,b} A_{ab}^k \left\{ \alpha_{ab}^{kc}.\frac{\partial^2 y^c}{\left(\partial A_{jk}^k \right)^2} \right\}$ $\alpha_{ij}^{kc} = \frac{\frac{\partial^2 y^c}{(\partial A_{ij}^k)^2}}{2\frac{\partial^2 y^c}{(\partial A_{ij}^k)^2} + \sum_{ab} A_{ab}^k \frac{\partial^3 y^c}{(\partial A_{ij}^k)^3}}$
C. Get a combination of weights and convolution layer outputs, by Passing through GlobalAveragePooling. $W_k^c = \frac{1}{w*h} \sum_i \sum_j \left(\frac{\partial y^c}{\partial A_{ij}^k} \right)$	C. Calculate Alpha (a pixel-wise weight). $y^c = \sum_k W_k^c \sum_{i,j} A_{ij}^k$
D. Passing through ReLU. $L_{Grad-CAM}^C = ReLU(\sum_k W_k^c A_{ij}^k)$	D. Find the weights by multiplying Alpha and positive partial derivatives. $W_k^c = \sum_i \sum_j \alpha_{ij}^{kc}.ReLU(\frac{\partial y^c}{\partial A_{ij}^k})$
E. Return GradCAM	E. Get the combination of weights and convolutional layers outputs passing through ReLU. $L_{Grad-CAM++}^c = ReLU(\sum_k W_k^c A_{ij}^k)$ F. Return GradCAM++.

ACRIMA, ORIGA, REFUGE, and DRISHTI. It shows that, the model developed for RIM-ONE dataset performed well for other datasets as well. Since the proposed approach of performing segmentation using the Attention U-Net with modified ResNet50 and the classification with the modified Inception v3 model, performs well for all the five datasets separately. Thus, the proposed solution is generalized across different datasets.

Table 4 shows the obtained performance metrics for ODS, OCS, and classification for the five datasets. The corresponding equations are shown in (8), (9), (10), (11) and, (12). Here, sensitivity denotes the probability of predicting glaucoma positive for a person with glaucoma. Specificity states the probability of predicting glaucoma negative for a person with a normal eye condition. The Dice coefficient (DC), also known as (F1-score), measures the performance of the segmentation. Studies have shown that, a DC value higher than 88% is considered a good fit between the predicted and actual segmentation [59]. This is computed by dividing the total number of active pixels in both masks by the number of active pixels at the intersection of the actual and predicted masks. The Jaccard coefficient (JC), also known as Intersection-Over-Union (IoU) measures the percentage overlap between the target mask and prediction output. This is an important metric in segmentation and higher values indicate greater similarity between the two images. Both DC and JC are positively correlated. The Area Under the Curve (AUC) shows the strength of the classifier to differentiate between classes. It is used as a summary of the ROC (receiver operating characteristic) curve. As shown in Table 4, higher AUC values indicate the better diagnostic prediction performance of the model.

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \quad (8)$$

$$Sensitivity = \frac{TP}{TP + FN} \quad (9)$$

$$Specificity = \frac{TN}{TN + FP} \quad (10)$$

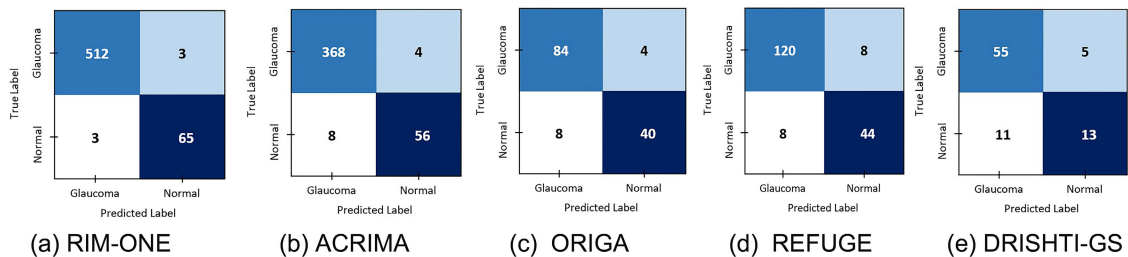$$Dice\ Coefficient(DC) = \frac{2*Area\ of\ Overlap}{Total\ pixels\ combined} \quad (11)$$

$$IoUorJaccard\ Index(JC) = \frac{Area\ of\ Overlap}{Total\ pixels\ combined} \quad (12)$$

Moreover, Figure 5 shows the obtained learning curves for the classification of the five different datasets. Generally, in best-fit models, the validation loss is slightly higher than the training loss, where both learning curves are relatively close to each other. The learning curves start with a somewhat high training and validation loss and steadily lowers. The minimum loss values are shown by the RIM-ONE dataset, where the training and validation loss are 0.0697 and 0.0723, respectively. Furthermore, the training and validation accuracy curves are relatively similar in all the datasets. The highest accuracy is shown by the RIM-ONE dataset, where the training and validation accuracies are 99.16% and 98.97%, respectively. Since the learning curves in Figure 5, do not show signs of underfitting and overfitting issues, the proposed approach performs well in training and testing image data.

Additionally, Figure 6 shows the confusion matrix for the predictions related to the five datasets. Here, the region's top-left, top-right, bottom-left, and bottom-right indicate TP, FP, FN, and TN instances, respectively. All the datasets have shown the highest values for the class of glaucoma images that are correctly predicted as Glaucoma by the model. The second highest values are shown for the class of normal images that are correctly predicted as normal. The lowest values represent the classes that are incorrectly predicted. Therefore, the approach has shown high results for all five datasets.

TABLE 4. Segmentation and classification results on different datasets.

Dataset	Type	Accuracy	Sensitivity	Specificity	DC	JC	AUC
RIM-ONE [12]	ODS	99.58%	99.63%	99.34%	99.76%	99.51%	
	OCS	97.52%	98.37%	93.56%	98.49%	97.03%	
	Classification	98.97%	99.41%	95.58%			99%
ACRIMA [51]	ODS	99.30%	99.54%	96.49%	99.62%	99.24%	-
	OCS	98.32%	98.78%	93.10%	99.08%	98.19%	-
	Classification	97.24%	97.87%	93.33%	-	-	98%
ORIGA [11]	ODS	88.59%	92.02%	72.50%	93.01%	86.93%	-
	OCS	88.15%	92.14%	67.56%	92.87%	86.69%	-
	Classification	91.17%	91.30%	90.90%	-	-	90%
REFUGE [10]	ODS	94.00%	96.03%	83.33%	96.41%	93.07%	-
	OCS	94.33%	96.00%	86.00%	96.57%	93.38%	-
	Classification	91.11%	93.75%	84.61%	-	-	93%
DRISHTI [9]	ODS	81.11%	83.07%	76.00%	86.40%	76.05%	-
	OCS	80.00%	84.61%	68.00%	85.93%	75.34%	-
	Classification	80.95%	83.33%	72.22%	-	-	83%

(a) RIM-ONE (b) ACRIMA (c) ORIGA (d) REFUGE (e) DRISHTI-GS
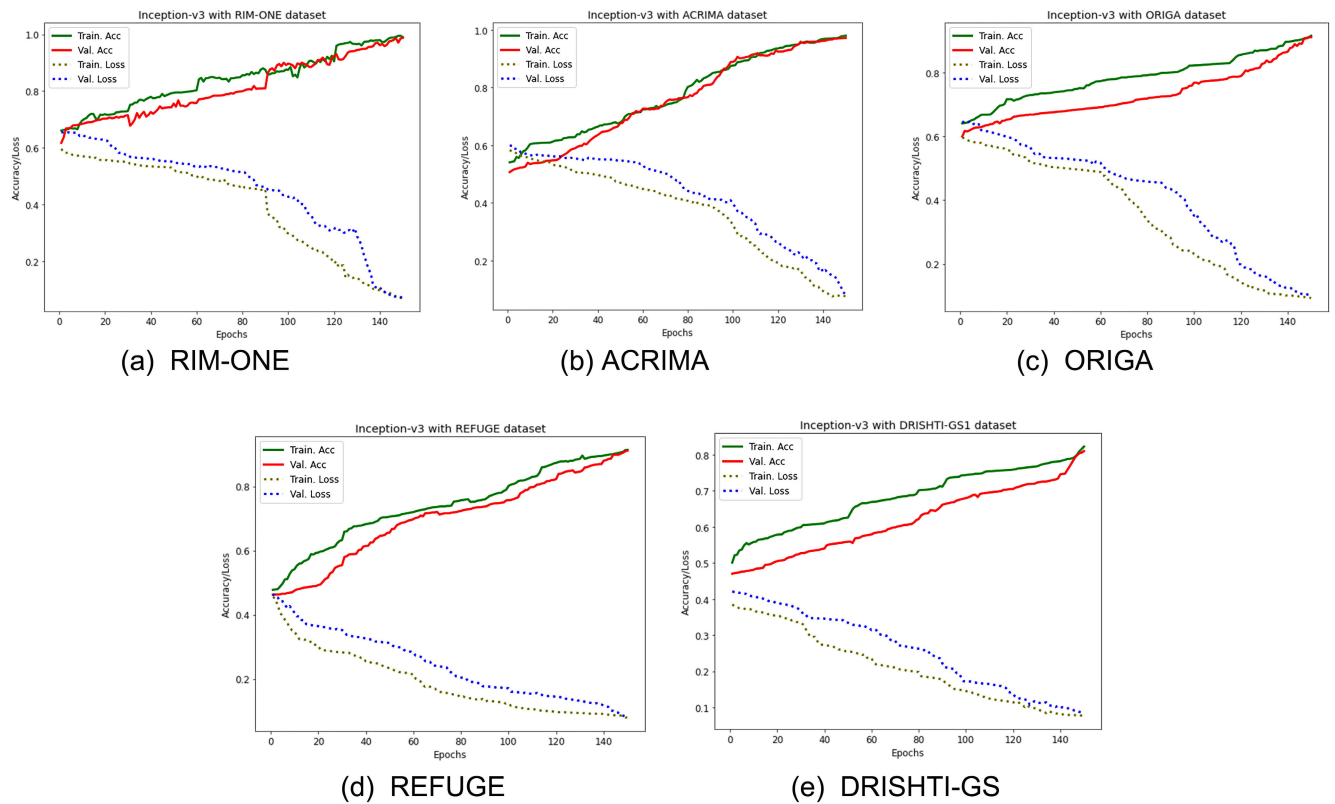
FIGURE 5. Learning curves of the models for different datasets.

(a) RIM-ONE (b) ACRIMA (c) ORIGA (d) REFUGE (e) DRISHTI-GS

FIGURE 6. Confusion matrix of the models for different datasets.

Furthermore, the ROC (receiver operating characteristic) curve shows the performance of the classification at all classification thresholds. It gives the trade-off between sensitivity and specificity. As shown in Figure 7, all the ROC curves of the five datasets are closer to the top-left corner. Therefore, we can say that the classifiers performed better
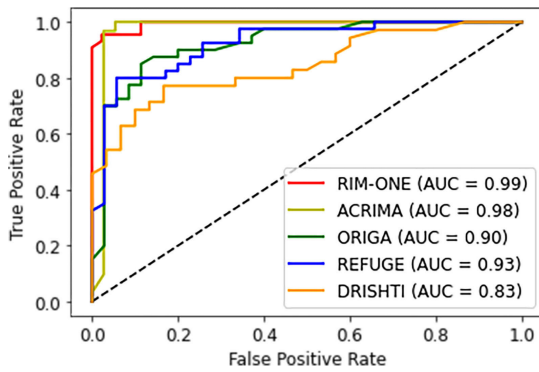
FIGURE 7. ROC for different datasets.

with a more accurate test, where RIM-ONE and ACRIMA datasets perform well than the other datasets.

B. ABLATION STUDY

The impact of each component is further verified through ablation study and Table 5 shows the outcomes of three distinct experiments conducted in the course of this study, encompassing classification-only [17], segmentation-only [5], and a combined segmentation followed by classification [18] approach. The classification-only results were derived from the Inception V3 model applied to the RIM-ONE dataset. Likewise, the segmentation-only outcomes were obtained through the utilization of the attention U-Net with the ResNet50 model, using the RIM-ONE dataset. Notably, the combined segmentation followed by classification results were generated by employing the attention U-Net with ResNet50 for segmentation and the Inception V3 for classification.

TABLE 5. Results comparison of three experiments.

Performance	Classification only %	Segmentation only%	Segmentation followed by classification%
Accuracy	96.56	97.37	98.97
Precision	97.22	-	99.41
Recall	98.31	96.53	99.42
F1-Score	97.76	-	99.40
Sensitivity	98.31	96.53	99.42
Specificity	90.90	97.97	95.59
AUC	98	-	99

A comprehensive evaluation of the results indicates that the segmentation followed by the classification process demonstrated superior performance when compared to the individual processes of classification and segmentation only. It is evident that the model achieved accuracy, precision, recall, F1-score, sensitivity, specificity, and AUC values of 98.97%, 99.42%, 99.40%, 99.42%, 95.59%, and 99%, respectively.

C. EXPLAINABILITY RESULTS

We deployed the proposed computational framework with the segmentation, classification, and explainability of GlaucoCare, a freely available web application for glaucoma diagnosis. In order to increase the trustworthiness of the

application and confidence in use, we included segmentation masks and explainability visualizations. The heatmaps generated using Grad-CAM and Grad-CAM++ techniques highlight the important regions on the input image that were considered to make the predictions. The cupping zone at the center of the OD strongly correlated with glaucoma cases. Generally, the suspicion and likelihood that the patient has glaucoma increase with a high cup-to-dis ratio. For the healthy eye, there are no visible cupping issues and no signs of glaucoma. Therefore no glaucomatous areas to highlight and visualize. The Glaucoma eye mainly highlights the optic nerve where the cupping issue appears. Most of the time, the red color spots were localized at the optic nerve head (ONH), which has the greatest impact on the classification. The importance is indicated by the emphasized shades in the order, from the most important to the least important such that red, orange, yellow, green, and blue.As shown in Figure 8, we compared the two explainability visualizations using Grad-CAM and Grad-CAM++. Generally, Grad-CAM++, built on Grad-CAM, provides better visual explanations of CNN model predictions, in terms of better object localization, explaining occurrences of multiple object instances in a single image, and better explanation-based knowledge distillation than Grad-CAM.

FIGURE 8. Set of input images and the corresponding heatmaps.

D. GLAUCOCARE WEB APPLICATION

GlaucoCare web application provides a user-friendly interface to test a fundus image for glaucoma conditions [26]. The developed model was trained and tested on a computing system equipped with an Intel Core i5 processor with 16GB of RAM. The use of a Solid-State Drive (SSD) contributed to faster data access and retrieval. The operating system employed was Windows 10, offering a stable and widely used environment for deep learning development. TensorFlow, a leading deep learning framework, was chosen as the core software tool for implementing and executing the model. The processor architecture utilized was x64, leveraging the capabilities of a 64-bit system. Moreover, the application was deployed using Streamlit, providing an interactive and user-friendly interface. GitHub served as the version control repository, ensuring collaboration
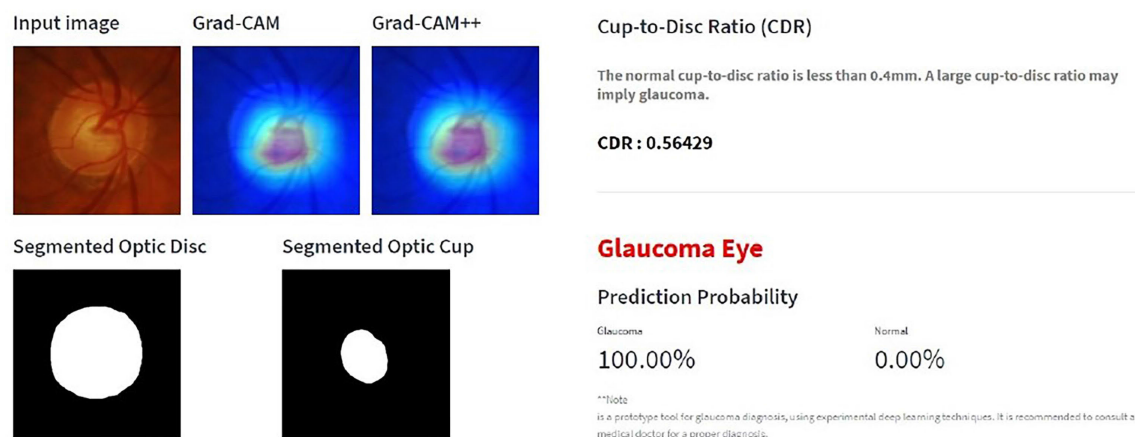
FIGURE 9. GUI of the GlaucoCare web application.

and code management. For the classification process, key libraries such as NumPy, Pandas, and Scikit-learn were extensively utilized, offering essential functionalities for data manipulation and deep learning tasks. Python was chosen as the primary programming language for its versatility and extensive ecosystem of libraries. The model training phase took place on Google Colab, leveraging its cloud-based infrastructure and GPU support for efficient computation and faster training times.

Users can either drag and drop the fundus image or using browse files and select the image. Once the user uploads a fundus image, the system generates the corresponding heatmaps together with the segmented optic disc (OD) and optic cup (OC) masks as shown in Figure 9. Additionally, the system calculates the cup-to-disk ratio and indicates the predicted class with the probability. Considering the outputs with the tool support, medical practitioners can get an idea about the position of OD and OC, the image regions that impacted for making the classification, and details related to the decision-making process. Accordingly, this application enables the trustworthiness of the underline DL model with explainable techniques. The model size is optimized at 271 MB with an inference time of 0.27 seconds. Moreover, the web application generates outputs, achieving a response time of 45 seconds, which is deployed on the Streamlit fee version platform. This includes both the processing time and the network latency of the web service.

E. VALIDATION OF RESULTS

We conducted a study among experts in the medical domain to validate the results obtained from the proposed approach [60]. We used 20 eye fundus images that could belong to a person infected with glaucoma or healthy and obtained the domain experts' knowledge to identify the category of each fundus image. Additionally, we recorded the reasons for their decision for each image and their view on having a computational support tool to assist the Glaucoma diagnosis process. We selected fundus images with 10 normal eye conditions and 10 glaucoma conditions. This

survey is done among an eye surgeon with 5+ years of experience, 5 ophthalmologists with 3-5 years of experience, and 6 medical interns who have 1-year experience with eye fundus image analysis. These medical interns are practicing under senior ophthalmologists. The main objective of this is to check the accuracy and the time taken to make the decisions.

We conducted a comprehensive validation of our model using randomly selected samples from five fundus image datasets. Figure 10 provides an overview of the performance of our model for each individual sample image, showcasing its efficacy in diverse scenarios. The heatmaps produced through the application of Grad-CAM and Grad-CAM++ methods emphasize significant areas within the input image that were taken into account during the prediction process. According to the heatmap, it generates the class (Normal or Glaucoma) along with the CDR value. Typically, the Cup-to-Disc Ratio (CDR) ranges between 0.1 and 0.9mm. However, if the CDR exceeds 0.5mm, it is identified as a potential case of glaucoma [61].

We check the classification accuracy of the same set of images from the automated model that we have proposed. The accuracy of the classification results provided by different user categories and the presented model is shown in Figure 11. Accordingly, the average classification accuracy shown by an eye surgeon, optometrists, and medical students is 100%, 88%, and 89.16%, respectively.

It can be observed that the classification by the eye surgeon gives 100% accurate results and it complies with the results obtained from the proposed automated DL model, for the selected images. Thus, we can state that our model performs well and can be used in clinical practice as a tool that gives a second opinion. Considering the responses of medical students, the results are satisfactory, based on their level of experience. However, they can use this tool to verify their decisions. However, considering the responses of the ophthalmologist, their classifications are not provided with high accuracy. Therefore, this application can be mainly used by people in the field with less experience, as a learning

| Dataset	Input Image	Grad-CAM	Grad-CAM++	Segmented OD	Segmented OC	Original Class	Predicted Class	CDR
REFUGE						Normal	Normal	0.40067
REFUGE						Glaucoma	Glaucoma	0.53947
ORIGA						Normal	Normal	0.43293
ORIGA						Glaucoma	Glaucoma	0.57471
DRISHTI						Normal	Normal	0.44966
DRISHTI						Glaucoma	Glaucoma	0.59864
ACRIMA						Normal	Normal	0.49512
ACRIMA						Glaucoma	Glaucoma	0.56429

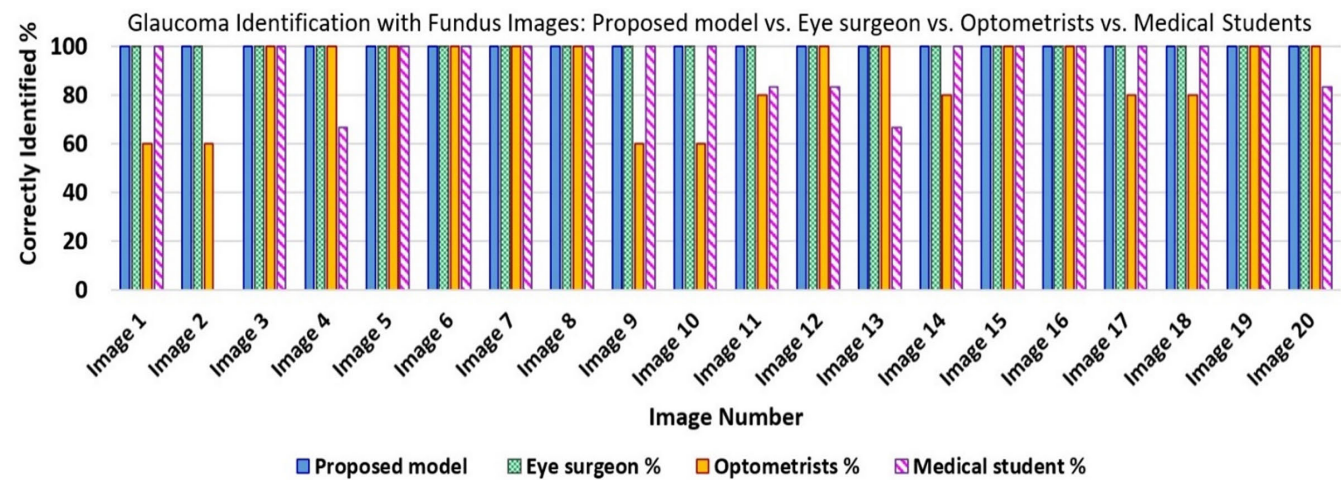**FIGURE 10.** Set of images from different datasets and the corresponding outcomes.

FIGURE 11. Proposed model vs. Overall users' classification.

support tool that gives a second opinion for glaucoma detection using fundus images.

F. SYSTEM USABILITY STUDY

The usability of the application based on the developed segmentation, classification, and explainability models was evaluated using a system usability study (SUS) [62], [63]. The SUS method provides a fast, cost-effective, and accurate way to assess the usability of a system. The survey consists of ten questionnaires on both positive and negative aspects of

usability and the responses are based on a 5-point Likert scale, where 1 represents "strongly disagree" and 5 represents "strongly agree" [64]. The odd and even numbered questions correspond to positive and negative impact, respectively. The overall SUS score was calculated using the corresponding number score for each response, as in Algorithm 2. The final SUS score is calculated by taking the average of the SUS scores of individual user responses. Generally, the average SUS score below 51 is considered "Awful," between 51 and 66 is considered "Poor," 67 is considered "Okay," between
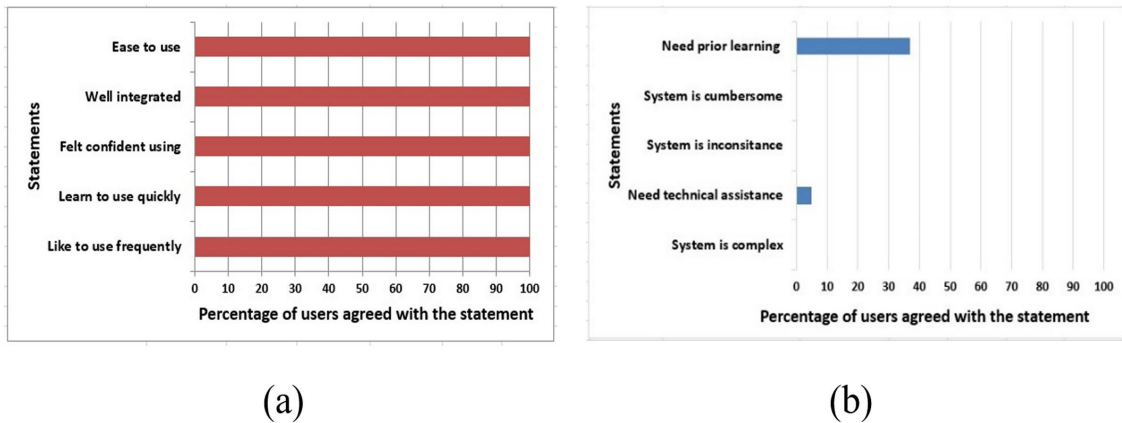
FIGURE 12. The percentage of (a) positive responses and (b) responses.

68 and 80.2 is considered "Good," and above 80.3 is considered "Excellent".

Algorithm 2 SUS Score
___
for user 1 to n:
   get $(x_i) \leftarrow \sum_{i=1}^{5}$ *(score of odd_num_question - 1)*;
   get $(y_i) \leftarrow \sum_{i=1}^{5}$ *(5 - score of even_num_question)*;
   $SUS\_score_i \leftarrow$ (get(x) + get(y)) * 2.5
   $Average\_SUS \leftarrow \sum SUS\_score_i$ /n
___

This study considered a sample of 22 users of which 16 are optometrists who have experience in fundus image analysis for 1 to 5 years. In addition, 6 medical interns who are practicing under the supervision of senior ophthalmologists and have 6 months to 1-year experience participated in this study. The ages of the sample range from 25 to 34. The proposed system has shown an average SUS score of 90.9%, indicating excellent usability.

Figure 12(a) indicates the percentage of each positive response from the 22 participants. Overall, 100% showed the system is easy to use, well integrated, felt confident when using, can learn to use quickly, and likes to use. Because the majority of the participants are ophthalmologists, they have extensive experience diagnosing glaucoma using various tests and machines. Therefore, these results are reasonable. Figure 12(b) depicts each negative response from the survey results. According to participant responses, 5% and 37% have found it difficult to use the system due to a lack of technical knowledge and less prior learning. However, this is addressable by refereeing to the provided user manual. The usability of the system can be enhanced further by having training sessions to assist users to understand the system's functionalities.

V. DISCUSSION

We proposed a glaucoma identification framework with explainability that increases the trustworthiness of the application. We have shown that segmentation followed by classification provides better performance compared to applying those techniques alone for glaucoma identification using fundus images [5], [17]. When implementing the segmentation segmentation process, we performed fine-tuning on the pre-trained ResNet50 by initially modifying the last two layers, specifically the dense and SoftMax layers. Subsequently, we replaced the fully connected layer with another densely connected layer featuring 256 units, integrated into the same pre-trained network [5]. This computational model, employing advanced deep learning techniques, serves as a valuable support tool for efficiently identifying glaucoma eye conditions, especially for non-experts such as medical students within the medical field. Additionally, this tool can be employed as a secondary opinion, contributing to and enhancing the clinical decision-making process. The implementation of this model has the potential to automate the diagnostic process, mitigating time-consuming aspects and reducing the susceptibility to human error. Thus, our approach is proven to be usable by conducting the results validation and usability measures with the domain experts' knowledge.

Moreover, Grad-CAM and Grad-CAM++ generate heatmaps highlighting regions of the input image that significantly contribute to the model's decision-prediction process. In the context of Glaucoma severity classification, there is often an emphasis on features related to the ONH. The presence of red color shading, as visualized in the heatmaps, may indicate regions associated with more severe cases of Glaucoma. The heatmap's color intensity, ranging from red (most important) to blue (least important), provides a qualitative indication of feature importance, with red indicating high significance.

Our tool generates predictions that encompass factors beyond heatmaps, incorporating critical features such as the Cup-to-Disc Ratio (CDR). The two main components of the optic nerve head, which are evaluated in glaucoma diagnosis, are the optic disc and the optic cup. We have considered those two components in this study. In glaucoma, changes to the optic disc, such as optic disc cupping, can be indicative of damage to the optic nerve. Cupping refers to the enlargement of the central area of the optic disc, known as the optic cup. The cup-to-disc ratio (CDR), which is the

ratio of the size of the optic cup to the size of the optic disc, is commonly used in glaucoma assessment. An increased CDR may suggest glaucomatous damage. According to the CDR value, it performs the stage of the glaucoma disease. If CDR > 0.5 mm then this is called case of glaucoma. If CDR is more than 0.5 and less than 0.55 then its moderate glaucoma. If cup to disc ratio size < 0.5 mm then its normal eye condition. Specifically, the assessment of the CDR plays a pivotal role in our prediction model. By considering both factors including heatmaps and CDR, our web app test the severity of the glaucoma.

The majority of existing studies have neglected the significance of addressing various forms of noise present in images. Our model is well-trained and it designed to provide meaningful predictions by filter out such noises. We have applied the median filter technique to reduce noise in an image while preserving its edges and fine details. Median filters are particularly effective in reducing salt-and-pepper noise, which often appears as isolated white and black pixels in images. By applying a median filter, the extreme pixel values caused by noise are replaced with more representative values from the surrounding region, leading to a cleaner input for the model. Applying a median filter can help in the extraction of more relevant and consistent features, providing the model with a cleaner and more representative input for analysis.

In this work, we have introduced novel contributions that address key gaps in existing research methodologies and significantly advance the field of glaucoma detection.

- Integrated five distinct datasets, significantly enhancing the generalizability and scalability of our approach.
- Developed a real-time web application, named GlaucoCare, as a practical implementation of our research findings, ensuring accessibility and usability. Also, it serves as a support tool specifically designed for non-experts within the medical domain, as evidenced by the positive outcomes from conducted surveys.
- Unlike many existing studies that primarily focus on DL-based classification models, our methodology adopts a generic approach involving segmentation followed by a classification process.
- The segmentation followed by classification approach in our methodology results in high-performance outcomes, demonstrating the effectiveness of our proposed method.
- Our methodology includes an explainability mechanism, enhancing the trustworthiness of the model by providing insights into its decision-making process.
- The predictive capacity of GlaucoCare extends beyond the use of explainable techniques, encompassing critical functionalities such as Cup-to-Disc Ratio (CDR) calculation.

Each of these novelties contributes to the unique and impactful aspects of our research, providing a comprehensive solution in the domain of glaucoma detection and diagnosis.

Additionally, Table 6 provides a comparison of the results with the existing studies. It is observable that, compared to the other related studies that have used classification and explainability techniques, our approach provides high accuracy values. Mainly, our model shows the best results with the RIM-ONE dataset. Further, we have considered five different fundus image datasets, without relying on one dataset, to show the generalizability of the proposed framework.

Many of the existing studies in the field have not explicitly addressed key metrics such as model size, inference time or response time, thereby complicating direct comparisons of computational complexity with existing research. In our proposed Dl model, the inference time, which is the time taken to predict the outcome of unseen data, is 0.27 seconds. This time depends on the network complexity and other hyperparameters such as the number of layers and the number of neurons in each layer. By having a small inference time, we can show that the model complexity and the size are low in the proposed model. In addition, our web application distinguishes itself by demonstrating a notably low response time of 45 seconds, that is deployed in the Streamlit free version platform. This indicates the time takes for the web service to respond to a request including the networking latency. This emphasis on response time highlights the efficiency of our model in real-world applications.

When considering the limitations of this study, it is only tailored for the detection of glaucoma disease. Further, this model supports only for detect the fundus image has glaucoma or not. However, it does not extend its capabilities to delineate the progression of the disease.

Future extensions of this work hold substantial potential for advancing the scope and impact of our research. The glaucoma classification process can be further improved with advanced and evolving DL techniques such as contrastive learning and vision transformers. Contrastive Learning helps to improve classification performance by learning the common attributes between classes and the attributes that differ among classes. Vision transformers use multi-head self-attention and process the image by splitting it into a set of positional embedding patches. Moreover, linguistic explanations can be incorporated with the output visualizations to support the transparency of the explainability. Additionally, the proposed application can be extended to support multi-modal and multi-variate data types to support the reliability and usability of the model. This will enable a complete framework to diagnose glaucoma in clinical practices.

An expansion into the identification of additional ocular conditions, including cataracts and diabetic retinopathy, would contribute to a more comprehensive and clinically relevant diagnostic framework. Moreover, Introducing refined evaluation criteria, such as the adoption of k-fold cross-validation for classification tasks, promises to enhance the model's robustness and generalizability across diverse datasets. Furthermore, incorporating quantitative measures to assess model explainability would represent a significant

TABLE 6. Comparison of glaucoma classification with explainability.

Study	Dataset	Classification	Explainability	Accuracy %	Precision %	Recall %	F1-score %	Sensitivity %	Specificity %	AUC %
[1]	Refuge	Xception	Grad-CAM	97	-	-	-	-	-	-
[15]	Private Dataset	ResNet-152	Grad-CAM	93.5	-	-	-	92.9	92.9	-
[38]	ORIGA	CNN	CAM	93.5	-	97.7	95.7	97.7	92.6	95.1
[41]	Drishti, RIM-ONE, RFUGE, ACRIMA	CNN (CoG-NET)	Activation Maps	95.3	-	95	-	95	99	99
[29]	NTUH	ResNet50	CAM & LIME	90.80	-	92.10	90.30	92.10	89.30	91.20
[33]	DRISHTI-GS1 RIM-ONE ACRIMA Private Refuge	CNN	N/A	90.32 90.51 96.70 88.89 89.17	90.91 87.10 97.46 100 46.51	- - - - -	93.02 89.26 97.05 90.91 50.63	95.24 89.74 96.64 83.33 55.57	80.00 91.53 96.67 100 92.90	85.24 90.63 96.71 91.67 80.40
[65]	RIM-ONE ORIGA G1020 DRISHTI-GS1	ResNet-50	N/A	96.15 92.59 98.48 97.03	- - - -	- - - -	97.00 95.00 98.00 97.00	97.85 98.39 99.30 93.75	92.38 79.26 96.52 98.55	94.20 93.00 97.00 96.00
This Work	RIM-ONE ACRIMA ORIGA REFUGE DRISHTI	InceptionV3 + ResNet50-based U-Net segmentation	Grad-CAM, Grad-CAM++	98.97 97.24 91.17 91.11 80.95	99.41 98.92 95.45 93.75 91.66	99.42 97.87 91.30 93.75 83.33	99.40 98.04 93.32 93.74 87.29	99.42 97.87 91.30 93.75 83.33	95.59 93.33 99.90 84.61 72.22	99.00 98 90 93 83

stride towards establishing a transparent and interpretable diagnostic system.

A crucial avenue for future exploration involves acquiring longitudinal data that captures the progression of disease stages, enabling a deeper understanding of the model's efficacy in varying clinical contexts. Further, While real-time data collection from medical institutions posed a challenge in the present study, future endeavors can focus on extending experiments to validate the model's performance under real-time conditions, fostering practical applicability and translational potential.

VI. CONCLUSION

This study presented a deep learning-based framework with segmentation, classification, and explainability to identify glaucoma conditions using different fundus image datasets. We deployed the proposed approach as a web application and showed the validity of the output and the usability of the system. The primary goal of this research was to apply explainable techniques to understand the reasons for the prediction using the underline classification model. This helps to increase the trustworthiness of the system and confidence in using such medical diagnosis support tools. We have considered five different fundus image datasets namely RIM-ONE, ACRIMA, ORIGA, REFUGE, and DRISHTI. With a comparative study of different CNN models, we used ResNet50-based attention U-Net for the segmentation and a modified architecture of the Inception V3 model for the classification. Both Grad-CAM and Grad-CAM++ techniques were used to generate heat maps for visual explanations. Thus, a computation model with deep learning techniques can be used as a support tool to identify glaucoma eye conditions efficiently.

REFERENCES

[1] A. Neto, J. Camara, and A. Cunha, "Evaluations of deep learning approaches for glaucoma screening using retinal images from mobile device," *Sensors*, vol. 22, no. 4, p. 1449, Feb. 2022, doi: 10.3390/s22041449.

[2] J. D. Stein, A. P. Khawaja, and J. S. Weizer, "Glaucoma in adults–screening, diagnosis, and management," *JAMA*, vol. 325, no. 2, pp. 164–174, Jan. 2021, doi: 10.1001/jama.2020.21899.

[3] World Health Organization, "World report on vision," 2019.

[4] A. K. Schuster, C. Erb, E. M. Hoffmann, T. Dietlein, and N. Pfeiffer, "The diagnosis and treatment of glaucoma," *Deutsches Ärzteblatt Int.*, vol. 117, no. 13, pp. 225–234, Mar. 2020, doi: 10.3238/arztebl.2020.0225.

[5] T. Shyamalee and D. Meedeniya, "Attention U-Net for glaucoma identification using fundus image segmentation," in *Proc. Int. Conf. Decis. Aid Sci. Appl. (DASA)*, Mar. 2022, pp. 6–10, doi: 10.1109/dasa54658.2022.9765303.

[6] L. J. Coan, B. M. Williams, V. Krishna Adithya, S. Upadhyaya, A. Alkafri, S. Czanner, R. Venkatesh, C. E. Willoughby, S. Kavitha, and G. Czanner, "Automatic detection of glaucoma via fundus imaging and artificial intelligence: A review," *Surv. Ophthalmol.*, vol. 68, no. 1, pp. 17–41, Jan. 2023, doi: 10.1016/j.survophthal.2022.08.005.

[7] N. Wijethilake, D. Meedeniya, C. Chitraranjan, I. Perera, M. Islam, and H. Ren, "Glioma survival analysis empowered with data engineering—A survey," *IEEE Access*, vol. 9, pp. 43168–43191, 2021, doi: 10.1109/access.2021.3065965.

[8] D. Meedeniya, H. Kumarasinghe, S. Kolonne, C. Fernando, I. D. L. T. Díez, and G. Marques, "Chest X-ray analysis empowered with deep learning: A systematic review," *Appl. Soft Comput.*, vol. 126, Sep. 2022, Art. no. 109319, doi: 10.1016/j.asoc.2022.109319.

[9] (2015). *Drishti-GS Dataset*. [Online]. Available: https://cvit.iiit.ac.in/projects/mip/drishti-gs/mip-dataset2/Home.php

[10] (2019). *Glaucoma Fundus Imaging Datasets Kaggle*. [Online]. Available: https://www.kaggle.com/datasets/arnavjain1/glaucoma-datasets?select=REFUGE

[11] (2010). *Glaucoma Fundus Imaging Datasets, Kaggle*. [Online]. Available: https://www.kaggle.com/datasets/arnavjain1/glaucoma-datasets?resource=download&select=ORIGA

[12] (Jul. 2021). *Medical Image Analysis Group*. [Online]. Available: https://medimrg.webs.ull.es/research/downloads/

[13] L. Gamage, U. Isuranga, S. De Silva, and D. Meedeniya, "Melanoma skin cancer classification with explainability," in *Proc. 3rd Int. Conf. Adv. Res. Comput. (ICARC)*. Belihuloya, Sri Lanka: IEEE, Feb. 2023, pp. 30–35.

[14] P. Harasymowycz, C. Birt, P. Gooi, L. Heckler, C. Hutnik, D. Jinapriya, L. Shuba, D. Yan, and R. Day, "Medical management of glaucoma in the 21st century from a Canadian perspective," *J. Ophthalmol.*, vol. 2016, pp. 1–22, Nov. 2016, doi: 10.1155/2016/6509809.

[15] M. Kim, J. C. Han, S. H. Hyun, O. Janssens, S. Van Hoecke, C. Kee, and W. De Neve, "Medinoid: Computer-aided diagnosis and localization of glaucoma using deep learning," *Appl. Sci.*, vol. 9, no. 15, p. 3064, Jul. 2019, doi: 10.3390/app9153064.

[16] S. Wickramanayake, S. Rasnayaka, M. Gamage, D. Meedeniya, and I. Perera, "Explainable artificial intelligence for enhanced living environments: A study on user perspective," (Advances in Computers), Amsterdam, NL, Europe: Elsevier, 2023, doi: 10.1016/bs.adcom.2023.10.002.

[17] T. Shyamalee and D. Meedeniya, "CNN based fundus images classification for glaucoma identification," in *Proc. 2nd Int. Conf. Adv. Res. Comput. (ICARC)*, Feb. 2022, pp. 200–205, doi: 10.1109/icarc54489.2022.9754171.

[18] T. Shyamalee and D. Meedeniya, "Glaucoma detection with retinal fundus images using segmentation and classification," *Mach. Intell. Res.*, vol. 19, no. 6, pp. 563–580, Dec. 2022, doi: 10.1007/s11633-022-1354-z.

[19] F. Fumero, S. Alayon, J. L. Sanchez, J. Sigut, and M. Gonzalez-Hernandez, "RIM-ONE: An open retinal image database for optic nerve evaluation," in *Proc. 24th Int. Symp. Comput.-Based Med. Syst. (CBMS)*, Jun. 2011, pp. 1–6, doi: 10.1109/CBMS.2011.5999143.

[20] A. Diaz-Pinto, S. Morales, V. Naranjo, T. Köhler, J. M. Mossi, and A. Navea, "CNNs for automatic glaucoma assessment using fundus images: An extensive validation," *Biomed. Eng. OnLine*, vol. 18, no. 1, pp. 1–19, Dec. 2019, doi: 10.1186/s12938-019-0649-y.

[21] Z. Zhang, F. S. Yin, J. Liu, W. K. Wong, N. M. Tan, B. H. Lee, J. Cheng, and T. Y. Wong, "ORIGA^{-light}: An online retinal fundus image database for glaucoma analysis and research," in *Proc. Annu. Int. Conf. IEEE Eng. Med. Biol.*, Sep. 2010, pp. 3065–3068, doi: 10.1109/iembs.2010.5626137.

[22] 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.

[23] J. Sivaswamy, S. R. Krishnadas, G. Datt 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, doi: 10.1109/ISBI.2014.6867807.

[24] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, "Grad-CAM: Visual explanations from deep networks via gradient-based localization," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 618–626, doi: 10.1109/ICCV.2017.74.

[25] A. Chattopadhay, A. Sarkar, P. Howlader, and V. N. Balasubramanian, "Grad-CAM++: Generalized gradient-based visual explanations for deep convolutional networks," in *Proc. IEEE Winter Conf. Appl. Comput. Vis. (WACV)*, Mar. 2018, pp. 839–847, doi: 10.1109/WACV.2018.00097.

[26] (Aug. 15, 2023). *Glaucoma Analysis Tool Glaucocare*. [Online]. Available: https://glaucocare.streamlit.app/

[27] (Aug. 15, 2023). *Glaucoma classification Bio-Health Informatics*. [Online]. Available: https://sites.google.com/cse.mrt.ac.lk/biohealth/home/imaging/glaucoma

[28] B. Zhou, A. Khosla, A. Lapedriza, A. Oliva, and A. Torralba, "Learning deep features for discriminative localization," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 2921–2929, doi: 10.1109/CVPR.2016.319.

[29] J.-M. Guo, Y.-T. Hsiao, W.-W. Hsu, S. Seshathiri, J.-D. Lee, Y.-M. Luo, and P. Liu, "A study of the interpretability of fundus analysis with deep learning-based approaches for glaucoma assessment," *Electronics*, vol. 12, no. 9, p. 2013, Apr. 2023, doi: 10.3390/electronics12092013.

[30] S. Dasanayaka, V. Shantha, S. Silva, D. Meedeniya, and T. Ambegoda, "Interpretable machine learning for brain tumour analysis using MRI and whole slide images," *Softw. Impacts*, vol. 13, Aug. 2022, Art. no. 100340, doi: 10.1016/j.simpa.2022.100340.

[31] S. Dasanayaka, S. Silva, V. Shantha, D. Meedeniya, and T. Ambegoda, "Interpretable machine learning for brain tumor analysis using MRI," in *Proc. 2nd Int. Conf. Adv. Res. Comput. (ICARC)*, Belihuloya, Sri Lanka, Feb. 2022, pp. 212–217, doi: 10.1109/ICARC54489.2022.9754131.

[32] V. V. N. S. Kumar, G. Harinath Reddy, and M. N. GiriPrasad, "A novel glaucoma detection model using Unet++-based segmentation and ResNet with GRU-based optimized deep learning," *Biomed. Signal Process. Control*, vol. 86, Sep. 2023, Art. no. 105069, doi: 10.1016/j.bspc.2023.105069.

[33] K. Sonti and R. Dhuli, "A new convolution neural network model "KR-NET" for retinal fundus glaucoma classification," *Optik*, vol. 283, Jul. 2023, Art. no. 170861, doi: 10.1016/j.ijleo.2023.170861.

[34] A. Raza, S. Adnan, M. Ishaq, H. S. Kim, R. A. Naqvi, and S.-W. Lee, "Assisting glaucoma screening process using feature excitation and information aggregation techniques in retinal fundus images," *Mathematics*, vol. 11, no. 2, p. 257, Jan. 2023, doi: 10.3390/math11020257.

[35] G. D'Souza, P. C. Siddalingaswamy, and M. A. Pandya, "AlterNet-K: A small and compact model for the detection of glaucoma," *Biomed. Eng. Lett.*, vol. 14, no. 1, pp. 23–33, Jan. 2024, doi: 10.1007/s13534-023-00307-6.

[36] W. Liao, B. Zou, R. Zhao, Y. Chen, Z. He, and M. Zhou, "Clinical interpretable deep learning model for glaucoma diagnosis," *IEEE J. Biomed. Health Informat.*, vol. 24, no. 5, pp. 1405–1412, May 2020, doi: 10.1109/jbhi.2019.2949075.

[37] D. Das, D. Ranjan Nayak, and R. Bilas Pachori, "CA-Net: A novel cascaded attention-based network for multistage glaucoma classification using fundus images," *IEEE Trans. Instrum. Meas.*, vol. 72, pp. 1–10, 2023, doi: 10.1109/tim.2023.3322499.

[38] O. Deperlioglu, U. Kose, D. Gupta, A. Khanna, F. Giampaolo, and G. Fortino, "Explainable framework for glaucoma diagnosis by image processing and convolutional neural network synergy: Analysis with doctor evaluation," *Future Gener. Comput. Syst.*, vol. 129, pp. 152–169, Apr. 2022, doi: 10.1016/j.future.2021.11.018.

[39] H. Tian, S. Lu, Y. Sun, and H. Li, "GC-Net: Global and class attention blocks for automated glaucoma classification," in *Proc. IEEE 17th Conf. Ind. Electron. Appl. (ICIEA)*, Dec. 2022, pp. 498–503, doi: 10.1109/iciea54703.2022.10005946.

[40] V. K. Velpula and L. D. Sharma, "Multi-stage glaucoma classification using pre-trained convolutional neural networks and voting-based classifier fusion," *Frontiers Physiol.*, vol. 14, Jun. 2023, Art. no. 1175881, doi: 10.3389/fphys.2023.1175881.

[41] M. Juneja, S. Thakur, A. Uniyal, A. Wani, N. Thakur, and P. Jindal, "Deep learning-based classification network for glaucoma in retinal images," *Comput. Electr. Eng.*, vol. 101, Jul. 2022, Art. no. 108009, doi: 10.1016/j.compeleceng.2022.108009.

[42] S. Sreng, N. Maneerat, K. Hamamoto, and K. Y. Win, "Deep learning for optic disc segmentation and glaucoma diagnosis on retinal images," *Appl. Sci.*, vol. 10, no. 14, p. 4916, Jul. 2020, doi: 10.3390/app10144916.

[43] Aziz-Ur-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. Biosciences Eng.*, vol. 18, no. 5, pp. 5321–5346, 2021, doi: 10.3934/mbe.2021270.

[44] A. Singh, S. Sengupta, and V. Lakshminarayanan, "Glaucoma diagnosis using transfer learning methods," *Appl. Mach. Learn.*, vol. 11139, pp. 223–232, Sep. 2019.

[45] J. J. Shanthamalar and R. G. Ramani, "A novel approach for glaucoma disease identification through optic nerve head feature extraction and random tree classification," *Int. J. Comput. Digit. Syst.*, vol. 10, no. 1, pp. 675–688, May 2021, doi: 10.12785/ijcds/100164.

[46] S. Liu, S. L. Graham, A. Schulz, M. Kalloniatis, B. Zangerl, W. Cai, Y. Gao, B. Chua, H. Arvind, J. Grigg, D. Chu, A. Klistorner, and Y. You, "A deep learning-based algorithm identifies glaucomatous discs using monoscopic fundus photographs," *Ophthalmol. Glaucoma*, vol. 1, no. 1, pp. 15–22, Jul. 2018, doi: 10.1016/j.ogla.2018.04.002.

[47] N. Gupta, H. Garg, and R. Agarwal, "A robust framework for glaucoma detection using CLAHE and EfficientNet," *Vis. Comput.*, vol. 38, no. 7, pp. 2315–2328, Jul. 2022, doi: 10.1007/s00371-021-02114-5.

[48] N. E. Benzebouchi, N. Azizi, and S. E. Bouziane, "Glaucoma diagnosis using cooperative convolutional neural network," *Int. J. Adv. Electron. Comput. Sci. (IJAECS)*, vol. 5, no. 1, pp. 38–43, Jan. 2018.

[49] A. Pal, M. R. Moorthy, and A. Shahina, "G-EyeNet: A convolutional autoencoding classifier framework for the detection of glaucoma from retinal fundus images," in *Proc. 25th IEEE Int. Conf. Image Process. (ICIP)*, Oct. 2018, pp. 2775–2779, doi: 10.1109/ICIP.2018.8451029.

[50] 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.

[51] (Mar. 2019). *CNNs for Automatic Glaucoma Assessment Using Fundus Images: An Extensive Validation*. [Online]. Available: https://figshare.com/s/c2d31f850af14c5b5232

[52] M. N. Bajwa, M. I. Malik, S. A. Siddiqui, A. Dengel, F. Shafait, W. Neumeier, and S. Ahmed, "Two-stage framework for optic disc localization and glaucoma classification in retinal fundus images using deep learning," *BMC Med. Informat. Decis. Making*, vol. 19, no. 1, pp. 1–6, Dec. 2019, doi: 10.1186/s12911-019-0842-8.

[53] D. Vijayalakshmi and M. K. Nath, "A novel multilevel framework based contrast enhancement for uniform and non-uniform background images using a suitable histogram equalization," *Digit. Signal Process.*, vol. 127, Jul. 2022, Art. no. 103532, doi: 10.1016/j.dsp.2022.103532.

[54] 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.

[55] K. A. Thakoor, X. Li, E. Tsamis, P. Sajda, and D. C. Hood, "Enhancing the accuracy of glaucoma detection from OCT probability maps using convolutional neural networks," in *Proc. 41st Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Jul. 2019, pp. 2036–2040, doi: 10.1109/EMBC.2019.8856899.

[56] D. Meedeniya, *Deep Learning: A Beginners' Guide*. Boca Raton, FL, USA: CRC Press LLC, 2023. [Online]. Available: www.routledge.com/9781032473246

[57] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Proc. 18th Int. Conf.*, Springer International Publishing, Oct. 2015, pp. 234–241, doi: 10.1007/978-3-319-24574-4_28.

[58] Y.-C. Ko, S.-Y. Wey, W.-T. Chen, Y.-F. Chang, M.-J. Chen, S.-H. Chiou, C. J.-L. Liu, and C.-Y. Lee, "Deep learning assisted detection of glaucomatous optic neuropathy and potential designs for a generalizable model," *PLoS ONE*, vol. 15, no. 5, May 2020, Art. no. e0233079, doi: 10.1371/journal.pone.0233079.

[59] L. Baskaran, S. Al'Aref, G. Maliakal, B. Lee, Z. Xu, J. Choi, S.-E. Lee, J. Sung, F. Lin, S. Dunham, B. Mosadegh, Y.-J. Kim, I. Gottlieb, B. Lee, E. Chun, F. Cademartiri, E. Maffei, H. Marques, S. Shin, and L. Shaw, "Automatic segmentation of multiple cardiovascular structures from cardiac computed tomography angiography images using deep learning," *PLoS ONE*, vol. 15, no. 5, 2020, Art. no. e0232573, doi: 10.1371/journal.pone.0232573.

[60] (2022). *Survey for Glaucoma Identification in Eye Fundus Images Google Docs*. [Online]. Available: https://forms.gle/ESAdW2fbqfm2yH7K9

[61] J. Patil and S. Chaudhari, "Determination of glaucoma grade with cup to disc ratio," *Acta Sci. Ophthalmol.*, vol. 4, no. 3, pp. 44–53, 2021.

[62] J. Brooke, "SUS: A retrospective," *J. Usability Stud.*, vol. 8, no. 2, pp. 29–40, Feb. 2013.

[63] (Oct 4, 2022). *Survey of System Usability Scale of Glaucocare Web Application Googleform*. [Online]. Available: https://forms.gle/AuYC6divDXLnUnEo7

[64] J. R. Lewis, "Item benchmarks for the system usability scale," *J. Usability Stud.*, vol. 13, no. 3, pp. 158–167, 2018.

[65] A. Shoukat, S. Akbar, S. A. Hassan, S. Iqbal, A. Mehmood, and Q. M. Ilyas, "Automatic diagnosis of glaucoma from retinal images using deep learning approach," *Diagnostics*, vol. 13, no. 10, p. 1738, May 2023, doi: 10.3390/diagnostics13101738.

THISARA SHYAMALEE is currently pursuing the M.Sc. (by Research) degree with the Department of Computer Science and Engineering, University of Moratuwa, Sri Lanka. Her research interests include machine learning, deep learning, and computer interaction.

DULANI MEEDENIYA (Senior Member, IEEE) received the Ph.D. degree in computer science from the University of St Andrews, U.K. She is currently a Professor in computer science and engineering with the University of Moratuwa, Sri Lanka. She is also the Director of the Bio-Health Informatics Group in her department and engages in much collaborative research. She is the coauthor of more than 100 publications in indexed journals, peer-reviewed conferences, and international book chapters. She has received several awards and grants for her contribution to research. She serves as a reviewer, a program committee, and an editorial team member in many international conferences and journals. Her main research interests include software modeling and design, bio-health informatics, deep learning, and technology-enhanced learning. She is a fellow of HEA, U.K., and MIET; a member of ACM; and a Chartered Engineer Registered at EC, U.K.

GILBERT LIM (Member, IEEE) received the Ph.D. degree in computer science from the Department of Computer Science, National University of Singapore. He is currently a Postdoctoral Researcher in the fields of computer vision and machine learning, with an emphasis on their application to medical imaging. His current work at SingHealth carries on from my time at iLab and involves the characterization and diagnosis of diseases, such as diabetic retinopathy and glaucoma through the analysis of retinal fundus photographs and related modalities. His Ph.D. thesis titled "Automated Methods for Retinopathy and Glaucoma Screening."

MIHIPALI KARUNARATHNE received the M.B.B.S. degree from the University of Sri Jayewardenepura. She is currently a Postgraduate Trainee in ophthalmology with National Eye Hospital, Colombo, Sri Lanka. She is also a member of the College of Ophthalmologists, Sri Lanka. Her main research interests include inherited retinal degenerative diseases, glaucoma, and diabetic retinopathy.

● ● ●