Machine Intelligence Research
www.mi-research.net

19(6), December 2022, 563-580
DOI: 10.1007/s11633-022-1354-z

# Glaucoma Detection with Retinal Fundus Images Using Segmentation and Classification

Thisara Shyamalee  Dulani Meedeniya

Department of Computer Science and Engineering, University of Moratuwa, Katubedda 10400, Sri Lanka

**Abstract:** Glaucoma is a prevalent cause of blindness worldwide. If not treated promptly, it can cause vision and quality of life to deteriorate. According to statistics, glaucoma affects approximately 65 million individuals globally. Fundus image segmentation depends on the optic disc (OD) and optic cup (OC). This paper proposes a computational model to segment and classify retinal fundus images for glaucoma detection. Different data augmentation techniques were applied to prevent overfitting while employing several data pre-processing approaches to improve the image quality and achieve high accuracy. The segmentation models are based on an attention U-Net with three separate convolutional neural networks (CNNs) backbones: Inception-v3, visual geometry group 19 (VGG19), and residual neural network 50 (ResNet50). The classification models also employ a modified version of the above three CNN architectures. Using the RIM-ONE dataset, the attention U-Net with the ResNet50 model as the encoder backbone, achieved the best accuracy of 99.58% in segmenting OD. The Inception-v3 model had the highest accuracy of 98.79% for glaucoma classification among the evaluated segmentation, followed by the modified classification architectures.

**Keywords:** Attention U-Net, segmentation, classification, Inception-v3, visual geometry group 19 (VGG19), residual neural network 50 (ResNet50), glaucoma, fundus images.

**Citation:** T. Shyamalee, D. Meedeniya. Glaucoma detection with retinal fundus images using segmentation and classification. *Machine Intelligence Research*, vol.19, no.6, pp.563–580, 2022. http://doi.org/10.1007/s11633-022-1354-z

## 1 Introduction

Glaucoma is the second most prevalent eye disease globally, and it can potentially lead to lifelong blindness. World Glaucoma Association estimated that around 80 million individuals have glaucoma worldwide by 2022. By 2040, the number of cases is expected to have risen to 111.8 million[1]. Despite the absence of signs in the initial stages, the disease affects the shape of the optic nerve head, harms nerve cells, consumes nerve fibers, and finally leads to irreversible blindness. Glaucoma is known as the "thief of sight" because it causes permanent vision loss. It is critical to discover the condition early to assist individuals in treating it and preventing it from progressing[2].

Over the last two decades, a range of techniques for medical image segmentation have been developed[3–5]. Image segmentation aims for pixel-level classification, which breaks down an image into subgroups[6]. Deep learning (DL) approaches have made substantial progress in a variety of medical image analysis applications in recent years[7–10]. Accordingly, several studies have addressed glaucoma analysis using convolutional neural networks (CNNs)[11–13]. DL approaches based on fully convolutional networks (FCNs), such as U-Net, have been explored in a series of studies using fundus images for various of image segmentation tasks, with DL models employed to extract features for classification[14].

U-Net was introduced for segmenting biomedical images with small datasets[15]. It is a convolutional model and consists of two types of segmentation models, namely, traditional U-Net and attention U-Net. The traditional U-Net architecture has an encoder path for extracting features and a decoder path for image restoration. The Attention U-Net architecture is based on the same concept as the traditional U-Net architecture. Attention U-Net focuses on different forms and sizes of target structures on its own. On the other hand, batch normalization (BN) is not a feature of both paths, resulting in model overfitting[16, 17]. As a result of these limitations, we developed a method for segmenting fundus images combining attention U-Net with CNN architectures, which subsequently be utilized for classification to produce promising glaucoma detection results.

This paper proposes a computational model to segment and classify retinal fundus images for glaucoma detection. We aim to enhance the performance of optic disc segmentation (ODS) and optic cup segmentation (OCS), and provide a better framework for glaucoma identification with promising results. We begin by designing and

Research Article
Manuscript received April 5, 2022; accepted June 21, 2022; published online November 7, 2022
Recommended by Associate Editor Jun-Zhou Huang

developing an attention U-Net with several backbones. Then we utilize three modified CNN architectures as the backbones, that is, Inception-v3, visual geometry group 19 (VGG19) and residual neural network 50 (ResNet50).

In recent years, CNN based models such as Inception-v3, VGG19, and ResNet50 have produced impressive outcomes in medical image analysis[18, 19]. We fine-tuned these models for the segmentation process in this study. The Inception-v3 model[20], has shown high-performance gain on CNNs. This model utilizes computing resources efficiently with a minimal increase in computational load. Additionally, it can extract features from data at varying scales with different convolutional filter sizes. The VGG19 model[21] was built as a deep CNN with 19 layers. It outperformed baselines on many tasks and is still one of the most commonly used vision models. This arrangement of convolution and max pool layers follows consistently throughout the whole architecture. In the end, it has two fully connected layers followed by a softmax for output. On the other hand, ResNet50[22], is a fast-performing model. It can train many layers without increasing the training error percentage. Additionally, this model can handle the vanishing gradient problem using identity mapping.

This study proposes an attention U-Net architecture by merging transfer learning with fundus images. We evaluate the models from different perspectives such that segmentation only and segmentation followed by classification using Inception-v3 architecture, as our previous results have shown better accuracy with the modified Inception-v3[12]. Joining attention U-Net with three different CNN architectures and highlighting the optimum models is the novelty of this study. The conclusions of this study will aid researchers and developers in the same sector.

In this paper, Sections 2–5 explore the related works, detail the design of the attention U-Net model for ODS and OCS, provide the outcomes, and wrap up the paper and discussion, respectively.

## 2 Background and related studies

### 2.1 U-Net architecture

The traditional U-Net architecture is based on a CNN approach that has expanded with minor CNN architecture changes. Generally, U-Net localizes and detects boundaries, as it classifies each pixel. The U-Net contains two essential parts: a contraction path and an expanding path. As demonstrated in Fig. 1, the contracting path (encoder) on the left is the general convolutional procedure for extracting the factors in the image. The contracting path includes two convolutions, including rectified linear unit (ReLU) activation and max pooling for downsampling. In the downsampling step, feature chan-

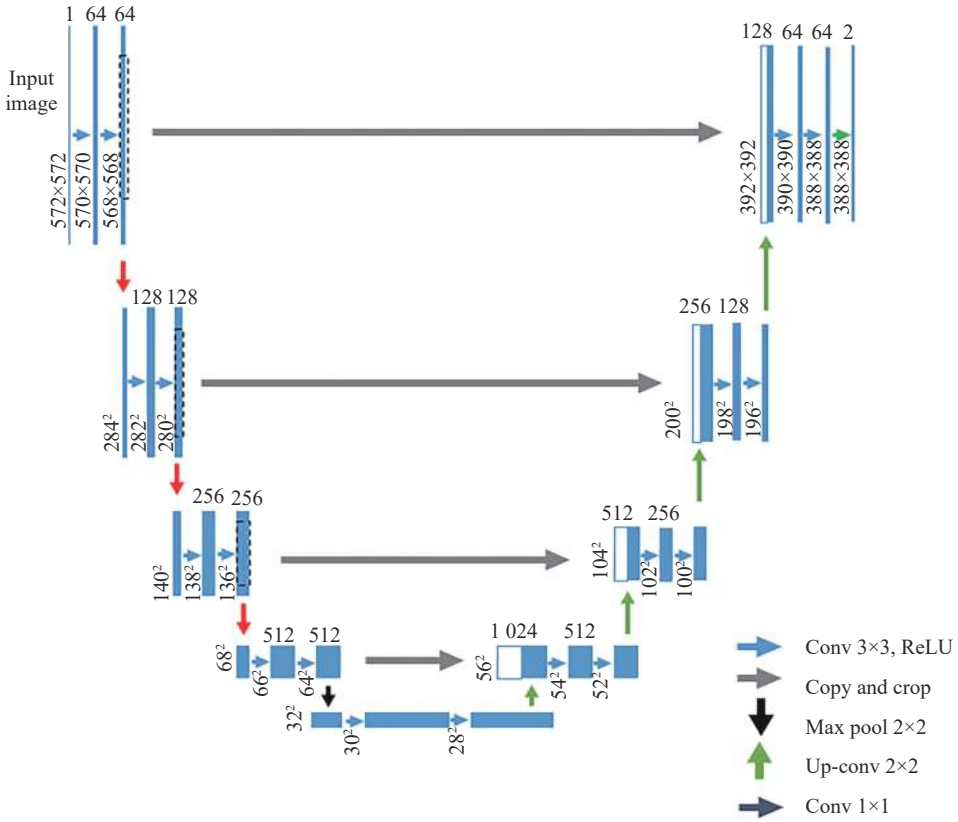



Fig. 1 Traditional U-Net architecture

nels will be doubled. The expanding path (decoder) on the right comprises transposed 2D convolutional layers (fully connected layers) that allow for localization. As a result, the feature channel size will be decreased to 50% of the initial size in the upsampling procedure[15]. Figs. 2 and 3, represent the contraction path and the expansive path, respectively. During the expanded path, the image will be enlarged to its original size.

Generally, the U-Net architecture offers promising results for images segmentation[4, 7, 14]. Considering the segmentation methods, semantic image segmentation aims to assign a label to anything with a class of an image representing a pixel. This task is known as a dense prediction since it performs estimation for every pixel in the image. Additionally, the targeted output is not merely labels and bounding box parameters in semantic segmentation. The result is a high-resolution image classified into a different class with each pixel. Skip connection brings redundant low-level features from the encoder path to the decoder path, and it is derived from the first layers and includes more spatial information, however feature information in the initial layers is limited[23]. Therefore, it is a considerable disadvantage of this U-Net.

However, the challenge of the traditional U-Net architecture is that both encoder and decoder parts consist of plain convolution, and it can be performed less efficiently for some complicated medical imaging tasks. As the solution to this issue, researchers have used variants of U-Nets as follows.

1) Replace the plain convolutional backbone with another model such as deep recurrent residual convolution, DenseNet, and inception block. For instance, the recurrent units unfold to a predefined time step making the network deeper at each layer and increasing the field of view in the lower layers of the neural network. Thus, these blocks aid in extracting crucial features such as precise low-level boundaries for segmentation. Hence, accurate feature extraction increases the performance.

2) Modify the skip connections between two paths, employing a dense skip pathway that decreases the semantic gap between two paths. Generally, dense skip connections deduct the semantic gap between encoder and decoder features before concatenation and forward different scale information to the decoder, aggregating different scale features to enhance the segmentation accuracy. Thus, these pathways accumulate features coming from multiple scales and apply concatenation. Additionally, extracting the advanced features in the segmentation tasks, recovering the lost fine details of foreground objects, and ensemble multi-depth output to ensure better accuracy is the other solution for that issue. Hence, embedding multi-depth models and an ensemble of outputs taken from varying depths improves the performance.

Although the increase in the network depth causes improved performance, the U-Net architecture endures a vanishing gradient problem. Such issues can be addressed using different activation functions such as ReLU, exponential linear units (ELU), and applying batch normalization between the layers. However, applying more deep layers increases the complexity and consumes more time for the training process. Identity mapping and skip connection in each considered model can be considered as solutions to that issue. Moreover, combining DL models with the original U-Net architecture leads to improved performance. For instance, residual units aid in training, whereas recurrent and residual convolutional layers improve feature representation for segmentation duties. It permits the construction of more efficient U-Net architectures for medical image segmentation with the same range of network parameters and increased performance[24, 25]. Furthermore, in related studies, a study with a dense-inception block has been used to combine an Inception-ResNet module and a U-Net architecture for feature extraction. In addition, a new convolution structure with a residual module that improves feature quality has been replaced with an inception module, to gain more semantic information by expanding the network width[26]. Thus,

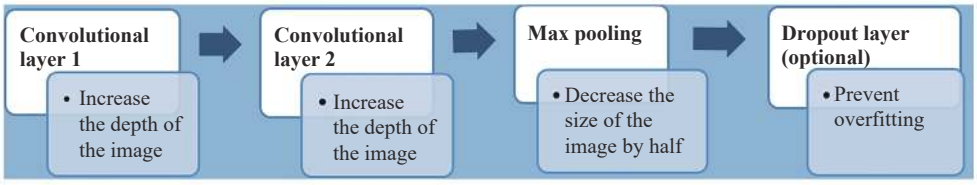



Fig. 2 Contraction path

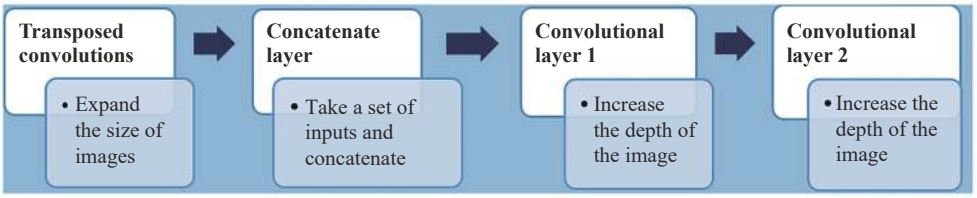



Fig. 3 Expansive path

U-Net is commonly employed in the segmentation of medical images[27, 28].

## 2.2 Related studies

Several studies have addressed fundus image segmentation and classification for glaucoma identification[7, 14].

For instance, an optic disc (OD) and optic cup (OC) segmentation based study is addressed by Tabassum et al.[29] They used a cup disc encoder-decoder network (CDED-Net) with three datasets namely, DRISHTI-GS, RIM-ONE and REFUGE. Eight convolutional layers with four blocks made up this modified U-Net. In the encoder route, each block contains a batch normalization layer. The path in the decoder is made up of four convolutional blocks. DRISHTI-GS has shown 99.71% and 99.66% accuracy for OC segmentation and OD segmentation, respectively.

In another study, Jin et al.[30] designed an encoder-decoder network structure called the aggregation channel attention network architecture. The aim of using an encoder is to gradually shrink the feature map′s spatial size while capturing complex semantic features. The decoder recovers the object′s features and spatial dimensions while retaining additional spatial data. They employed the DenseNet block to obtain more significant advanced semantic features and performed an aggregation channel attention upsampling (ACAU) module to retain more spatial information in the decoding path. The DenseNet block comprises a dense block for varying feature resolutions, and the transition comprises 4 DenseNet blocks. The decoder path is made up of four aggregate channel attention upsampling modules that keep the encoder′s high-level features while preserving the spatial resolution of the feature map. The OD segmentation using the aggregation channel attention upsampling network (ACAU-Net) has shown a lower overlapping error of 0.046 9.

Moreover, Civit-Masot et al.[31] developed a cloud-based U-Net architecture to perform cup and disc segmentation. The recommended design consisted of two convolutional layers. Both encoder and decoder pathways employ the ReLU activation function and dropout layers. To produce networks with acceptable performance, they merged several publicly accessible datasets, including RIM-ONE V3, DRISHTI, and DRIONS. This method performs well on both segmentation types obtaining a 95% of Dice coefficient (DC).

In addition, an attention U-Net with pre-trained models for ODS and OCS is presented by Zhao et al.[16] An attention gate is placed between the U-Net encoder and the decoder to focus on the target area. After network training using the DRIONS-DB dataset to obtain the encoder′s weights, the DRISHTI-GS dataset was used to train it to further alter the weights. As a result, the DRISHTI-GS dataset obtained the best results for the DC, Jaccard coefficient (JC), sensitivity, specificity, and accuracy of 96.38%, 93.01%, 94.88%, 99.93%, and 99.75%, respectively, for OD segmentation.

Furthermore, Sevastopolsky[32] suggested a modified U-Net combining deep learning for OD and OC segmentation on fundus images. They have looked at two options for pre-processing OD and OC segmentation. First, the altered U-Net contains the fewest convolution filters, resulting in fewer parameters and a quicker training time. This suggested method performs well on OD segmentation, obtaining DC, intersection-over-union (IoU) of 95% and 89%, respectively, for the RIM-ONE V3 dataset. However, considering the results of optic cup segmentation, this method is not ideal for cup segmentation.

# 3 System design and implementation

## 3.1 Process view

The high-level view of the proposed methodology is represented in Fig. 4. Our approach mainly consists of data pre-processing, augmentation, attention U-Net architecture-based segmentation, DL-based classification, and model evaluations. Different pre-processing and feature extraction techniques have been used in the domain of medical image analysis[33–35]. The following subsections describe the whole process in detail. The pipeline shown in yellow, which occupies only the classification process to obtain the output, is based on our previous work[12].

Our segmentation and classification process is based on three models: Inception-v3, VGG19, and ResNet50. We have selected these models due to the in-built features in those architectures. For instance, Inception-v3 uses the inception module to aid in the extraction of more detailed features from the supplied image. Furthermore, the VGG19 architecture consists of small kernels and allows us to learn more complex features.

The reason behind selecting the ResNet50 architecture is the vanishing gradient problem. It includes skip connections that serve as gradient superhighways, allowing the gradient to flow freely. It allows gradients to spread to deeper layers before becoming attenuated to tiny or zero levels. Another issue with training deeper networks is that the optimization is done on a large parameter space, resulting in naively adding layers and increasing training error. Residual networks enable the training of deep networks by creating the network using residual model modules.

## 3.2 Segmentation process

### 3.2.1 Dataset details

We employed two publicly available datasets for the segmentation process, RIM-ONE[36] and ACRIMA[37]. The RIM-ONE dataset has three versions, and the first and

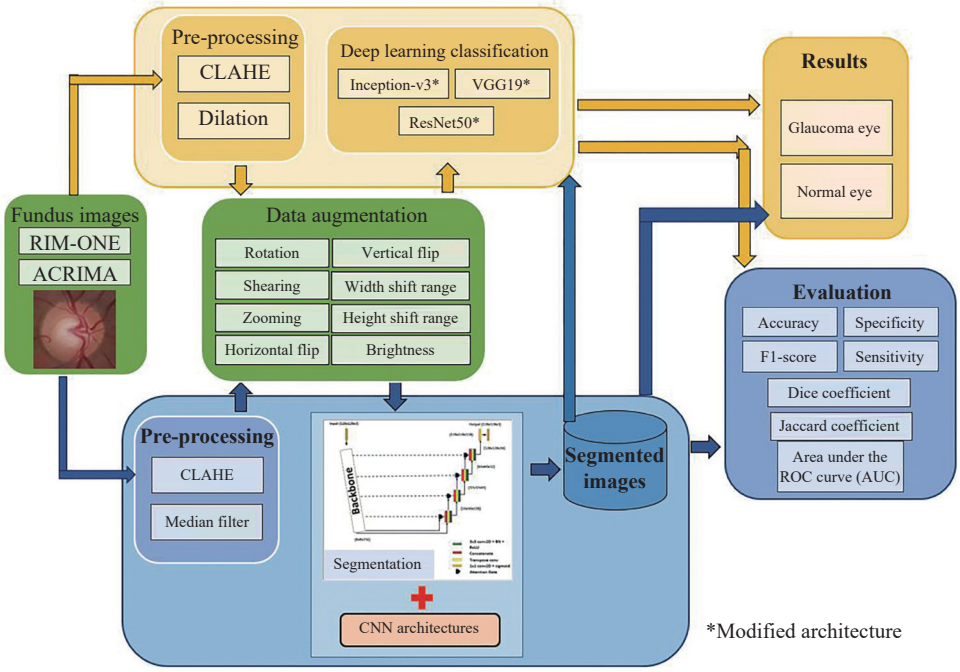



Fig. 4 Overall view of the model

second versions include 169 and 455 monosporic fundus images, respectively. The third version contains 318 stereoscopic images categorized into four categories: normal, moderate, early, and deep. We treated all three classes as glaucoma since the deep, early, and moderate classes are all associated with glaucoma. Thus, the RIM-ONE dataset consists of 942 fundus images with 399 glaucomatous and 543 normal subjects. The ACRIMA dataset consists of 396 glaucomatous and 309 normal cases.

Each fundus image must have a mask for implementing the attention U-Net architecture. However, optic disc masks are available only in the RIM-ONE version 1 dataset, and optic disc and cups masks are consistent in RIM-ONE version 3. Therefore, we manually constructed masks of the remaining dataset images to fulfil the segmentation procedure[38]. We split the images into a 70:15:15 ratio as the training, testing and validation sets during the U-Net segmentation process. Due to a shortage of OD and OC ground truth images, the ACRIMA dataset can only be used for classification techniques[37]. As a result, no segmentation experiments have been conducted using the ACRIMA dataset[4, 38, 39].

### 3.2.2 Pre-processing and augmentation techniques

We used two pre-processing techniques, contrast limited adaptive histogram equalization (CLAHE) and a median filter, for the segmentation process. CLAHE is a well-known pre-processing technique that improves image quality and contrast[40]. On the other hand, since noise destroys the quality of the images, the median filter technique is used to reduce the noise while keeping the image's edge features[39].

Augmentation techniques to handle the data imbalance issue, data overfitting, and enhance the number of images available in the dataset. We employed six augmentation techniques: rotation range of 10 degrees, zooming range of 0.1, shearing range of 0.2, horizontal flip, height shift range, and width shift range of 10% for both datasets. Moreover, two augmentation techniques were employed for both datasets, namely, a brightness range of 1 and a contrast range of 2, to address the data imbalance problem. After splitting both sets into a 70:15:15 ratio, the images of the training, testing and validation sets of the RIM-ONE dataset consist of 4 512 976 and 970 fundus images. The ACRIMA dataset contains 3 193 724 and 718 images from the training, testing and validation sets, respectively. Table 1 shows the data distribution of the training and testing sets after the pre-processing process.

### 3.2.3 Proposed attention U-Net architecture

Studies with segmentation have focused on fully convolutional networks (FCN). Based on the FCN architecture, the U-Net architecture is adequate for binary image segmentation, making it suitable for biomedical image segmentation[16]. The attention U-Net architecture is similar to the standard U-Net architecture, however, it has been significantly adjusted to obtain the significant outcomes. Other than the encoder and decoder paths, each level of the attention U-Net contains an attention gate (AG) at the skip connection.

In the proposed model, we substituted the original encoder of the standard U-Net with pre-trained networks, namely VGG19, Inception-v3, and ResNet50 as backbones independently, in the contraction path, to discover the best segmentation performances. The view of the proposed attention U-Net is shown in Fig. 5, where the backbone is replaced using the pre-trained Inception-v3,

Table 1 Data distribution

| Dataset | Training set 70% | | Testing set 15% | | Validation set 15% | |
|---|---|---|---|---|---|---|
| | Glaucoma | Normal | Glaucoma | Normal | Glaucoma | Normal |
| RIM-ONE | 2 232 | 2 280 | 480 | 496 | 480 | 490 |
| ACRIMA | 1 590 | 1 603 | 372 | 352 | 366 | 352 |

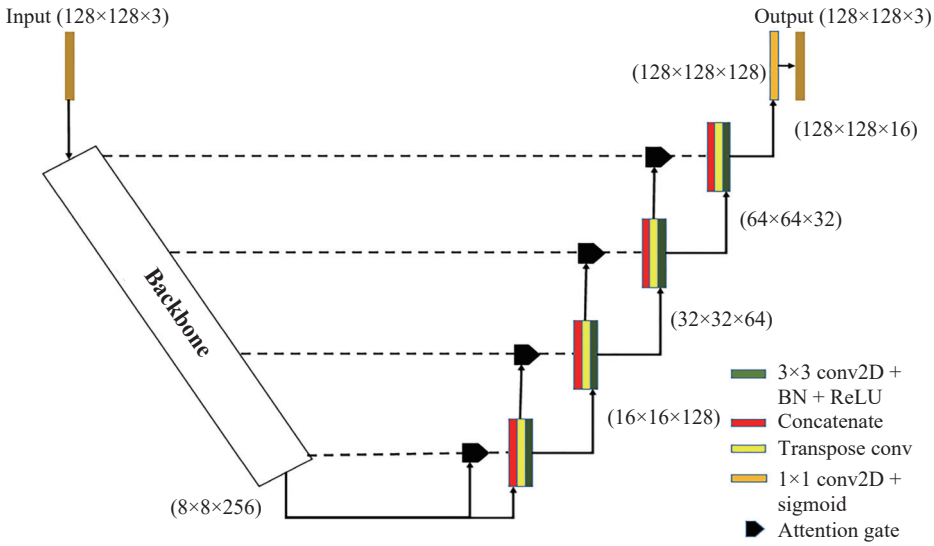



Fig. 5 Proposed U-Net architecture

VGG19, and ResNet50 models. Convolutional, upsampling, and concatenation are used in all three networks.

Each convolution layer is followed by a batch normalization layer and a ReLU activation layer. Furthermore, the ReLU activation function is given in (1), where $x$ represents the input to a neuron.

$$ReLU(x) = \max(0, 1). \tag{1}$$

At each level of the attention U-Net architecture, an attention gate allows two inputs: gating signal ($g$) and skip connection ($x$). The spatial information is fulfilled by a gating signal that originates from the next lower layers. Skip connections are consistent with good feature representation. As a result, the sum of the $x$ and $g$ signals retains good feature and spatial information. Fig. 6 represents the attention gate, where $X$ is the feature map of the encoder path. The output of the attention gate ($X^1$) is defined in (2), and the attention coefficient vector $\alpha$ is stated in (3), where $\sigma_1$ and $\sigma_2$ coordinate with ReLU and sigmoid activation function, respectively. $W_x$ and $W_g$ define the linear transformation, $b_g$ and $b_\theta$ demonstrate the bias. We obtained the output of the attention gate by taking the dot product of the feature map of the encoder path and the attention coefficient vector.

$$X^1 = X \cdot \alpha \tag{2}$$

$$\alpha = \sigma_2(W_\theta^{\mathrm{T}}(\sigma_1(W_x^{\mathrm{T}}X + W_g^{\mathrm{T}}g + b_g)) + b_\theta). \tag{3}$$

The first CNN design utilizes an inception-v3 encoder as the U-Net encoder. It is made up of convolutional layers and max-pooling layers. Inception modules were utilized to replace each convolutional layer in the original U-Net, including 3×3 convolutions, 1×1 convolutions, and 3×3 max-pooling. The VGG19 encoder was used in the second attention U-Net architecture and it made use of convolutional and max-pooling layers. The filter size will be doubled following the max-pooling layer. ResNet50 is the final CNN architecture, and it consists of residual blocks with skip connections and skip connections that handle the vanishing gradient problem. The ResNet50 architecture comprises residual blocks.

### 3.3 Classification process

We fed the segmented results through a classification process utilizing three separate modified CNN architectures, namely Inception-v3*, VGG19*, and ResNet50*, from our previous work[12]. The primary objective of this classification step is to obtain highly accurate results. Thus, we used the segmented images obtained from the ResNet50 backbone, which gives the highest accuracy of the segmentation process, as the input images for the classification process. We split the segmented image set into a 70:15:15 ratio. Overall, there are 2 722 training images, 583 testing images and 583 validation images in the image set.

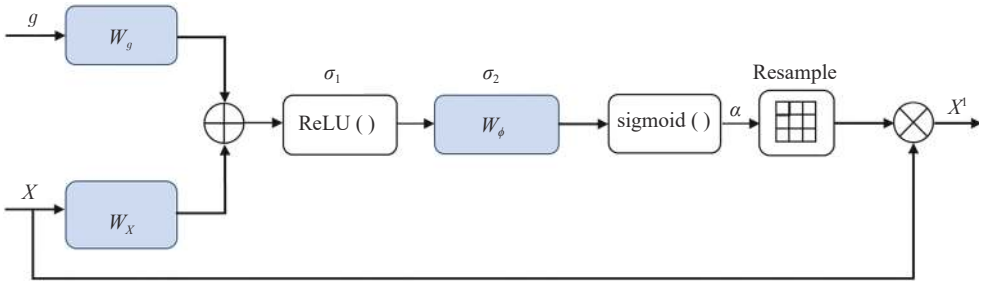



Fig. 6 Illustration of the AG gate

The modified Inception-v3* model has an additional three layers instead of the top dense layers. For instance, the global average pooling (GAP) layer, was used to minimize the parameters. A dense layer with 512 units and a softmax layer were added. Two neurons were used in the softmax layer for the normal and glaucoma classes. Finally, we applied a dropout layer with 0.7 rates to decrease the model overfitting. Furthermore, the vanishing gradient problem was addressed by the auxiliary classifier of the Inception-v3 model. They have utilized a high learning rate with the Adam optimizer to train the model. The modified Inception-v3 architecture is shown in Fig. 7.

Moreover, for the modified VGG19*, which consists of 19 layers, we added additional convolution layers as the final three blocks as shown in Fig. 8. The last three newly added layers are inserted after the GAP layer: a dropout layer with a 0.5 rate, a dense layer (256 units) with a ReLU activation function, and finally, softmax layer with two outputs.

Finally, Fig. 9, shows the modified ResNet50* architecture used for the classification process. It incorporates a few skip connections to combat the vanishing gradient problem that occurs when a large number of layers are used, allowing it to cling to data from two layers. Initially, we proceeded by changing the final two dense and softmax layers. The fully connected layer is then replaced by a 256-unit fully connected dense layer in the same pre-trained networks.

## 4 Results and analysis

### 4.1 Evaluation metrics

The image segmentation and classification were evalu-

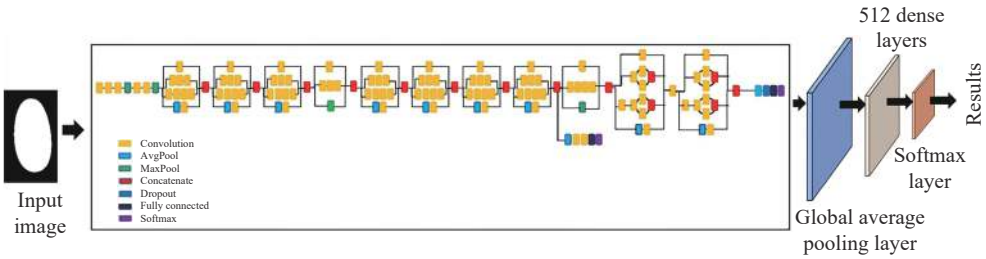



Fig. 7 Modified Inception-v3* architecture

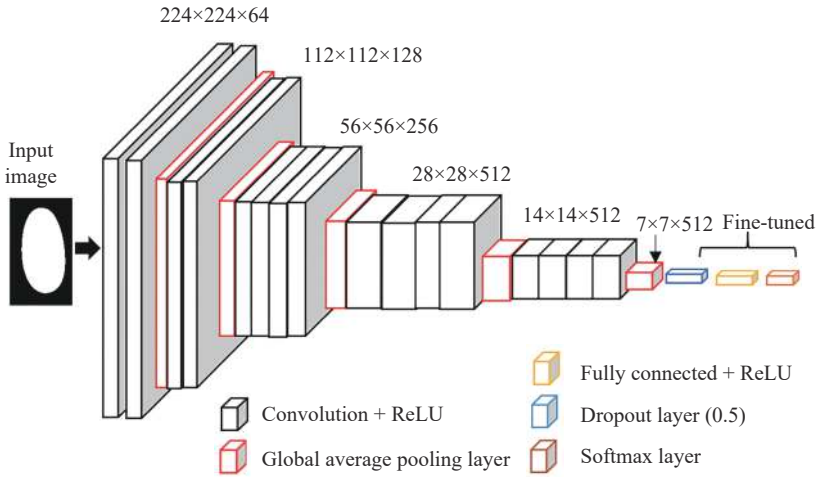



Fig. 8 Modified VGG19* architecture

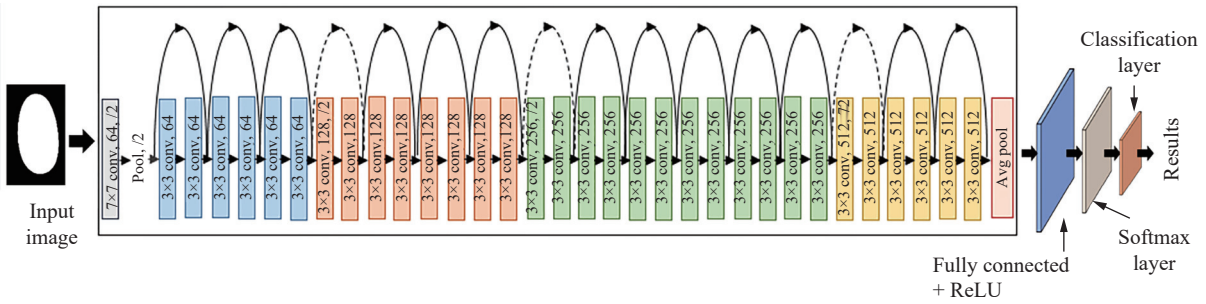



Fig. 9 Modified ResNet50 architecture

ated using the widely used performance metrics DC, JC, accuracy, sensitivity, and specificity, as given in (4)–(8), respectively. The DC is calculated by dividing the number of active pixels at the intersection of the actual and predicted masks by the total number of active pixels in both masks. Furthermore, sensitivity is critical in medical field-based research since it reflects the percentage of diseased individuals that the model is capable of detecting[41].

In this study, the terms true positive, false positive, false negative and false positive denoted by TP, FP, FN and FP, respectively, are described as follows.

**True positive (TP).** Predicted as positive and the actual outcome is also true. The model identifies the OD/OC correctly, and the ground truth mask is also the same.

**True negative (TN).** Predicted as unfavourable, but the actual outcome is true. The model identifies the OD/OC incorrectly, and the ground truth mask does not correctly identify the OD/OC.

**False positive (FP).** Predicted as positive, but the actual outcome is not valid. The model identifies the OD/OC correctly, but the ground truth mask does not correctly identify the OD/OC.

**False negative (FN).** Predicted as unfavourable, but the actual outcome is true. The model identifies the OD/OC incorrectly, but the ground truth mask is correctly identified as the OD/OC.

$$\text{Dice coefficient (DC)} = \frac{2 \times TP}{(2 \times TP) + FP + FN} \tag{4}$$

$$\text{Jaccard coefficient (JC)} = \frac{TP}{TP + FP + FN} \tag{5}$$

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \tag{6}$$

$$Sensitivity = \frac{TP}{TP + FN} \tag{7}$$

$$Specificity = \frac{TN}{TN + FN}. \tag{8}$$

Cross-entropy loss (CE) is a widely used loss function in classification problems. It displays the difference between two probability distributions as in (9), where $b$ is the binary indication (0 or 1) that represents the sample's class, and a signifies the sample's predicted probability between 0 and 1. In some image segmentation problems, the background comprises a significant portion while the object consists of a small portion. Thus, the loss function should be utilized properly[42, 43].

$$CE = CE_{a,b} = -(b \log(a) + (1 - b) \log(1 - a)). \tag{9}$$

Furthermore, the confusion matrix displays the number of correct and incorrect predictions for each class. The confusion matrix is organized in a tabular format, with the four values divided into two categories: actual and predicted.

## 4.2 Classification results

This section describes the performance metrics obtained from the classification of the fundus images without segmentation. These results are based on our previous work[12], and are included as a summary together with evaluation extensions for the completeness of the paper. The modified architectures were used for the considered Inception-v3*, VGG19* and ResNet50*. We employed hyperparameter tuning to find the most effective training settings. For the internal layers, the ReLU activation function is employed, and each model training batch size is set to eight. The ADAM optimizer is used to train the Inception-v3* model with a learning rate of 0.000 1. The VGG19* and ResNet50* models are trained using stochastic gradient descent (SGD) optimizer updates with 0.9 momentum and a learning rate of 0.001. Since the goal is binary classification (glaucoma and healthy), we used binary cross-entropy loss for each model. We trained the data for 200 epochs during this comparative study. Table 2 summarizes the obtained results. The highest classification accuracy of 98.52% is given using the modified Inception-v3 model for the ACRIMA dataset. The highest sensitivity value of 99.5% was also given by the modified inception-v3 model.

Table 2 Performance of CNN architectures for classification

| Performance metrics | Inception-v3* | | VGG19* | | ResNet50* | |
|---|---|---|---|---|---|---|
| | RIM-ONE | ACRIMA | RIM-ONE | ACRIMA | RIM-ONE | ACRIMA |
| Accuracy | 96.56% | 98.52% | 94.95% | 92.64% | 95.49% | 95.58% |
| Precision | 97.22% | 98.84% | 95.37% | 93.45% | 96.50% | 96.49% |
| Recall | 98.31% | 99.50% | 97.98% | 97.08% | 97.59% | 98.21% |
| F1-score | 97.76% | 99.17% | 96.65% | 95.23% | 97.04% | 97.34% |
| Sensitivity | 98.31% | 99.50% | 97.98% | 97.08% | 97.59% | 98.21% |
| Specificity | 90.90% | 90.90% | 86.13% | 78.78% | 88.88% | 83.33% |
| Loss | 0.060 1 | 0.041 4 | 0.079 2 | 0.075 6 | 0.068 9 | 0.059 8 |
| AUC | 0.98 | 0.99 | 0.95 | 0.94 | 0.97 | 0.96 |

## 4.3 Segmentation results

We used several metrics to measure the performance of the fundus image segmentation process. These results are an extension of our previous work[17]. As mentioned in the methodology, the three CNN architectures are trained for 150 epochs with a 0.001 learning rate. The Inception-v3 utilized an ADAM optimizer, and the other remaining architectures employed SGD. Figs. 10–13 represent the training accuracy, validation accuracy, training loss, and validation loss of OD and OC segmentation for the RIM-ONE and ACRIMA datasets. Considering a given figure, the left (a), middle (b) and right (c) figures show the results obtained using the attention U-Net with Inception-v3, VGG19, and ResNet50 architectures, respectively. Thus, we have shown the acquired findings for both datasets individually for each architecture.

Accordingly, Figs. 10(a)–10(c) show the OD segmentation of Inception-v3, VGG19, and ResNet50 on the RIM-ONE dataset. Fig. 10(a) shows the accuracy of 98.91% for training and the accuracy of 98.65% for validation at the end of 150 epochs. In Fig. 10(b), VGG19 yielded 99.53%

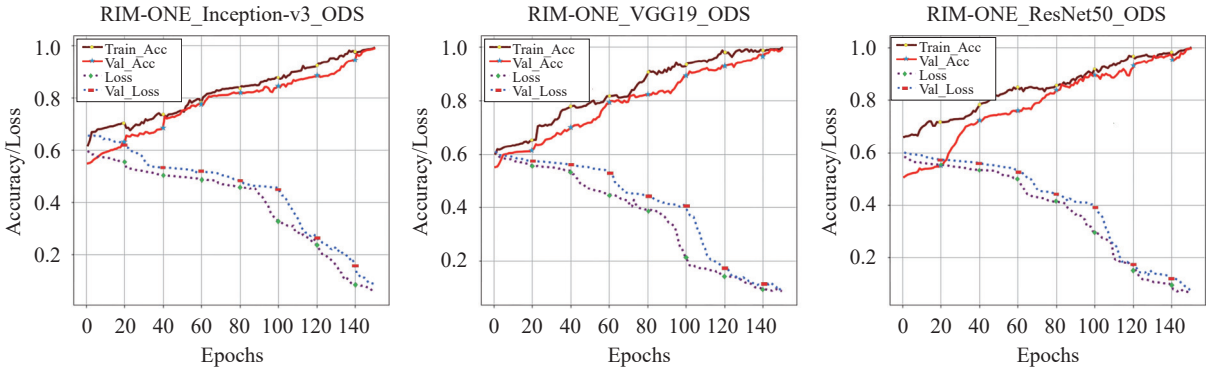



Fig. 10 Optic disc segmentation of (a) Inception-v3, (b) VGG19, (c) ResNet50 on RIM-ONE

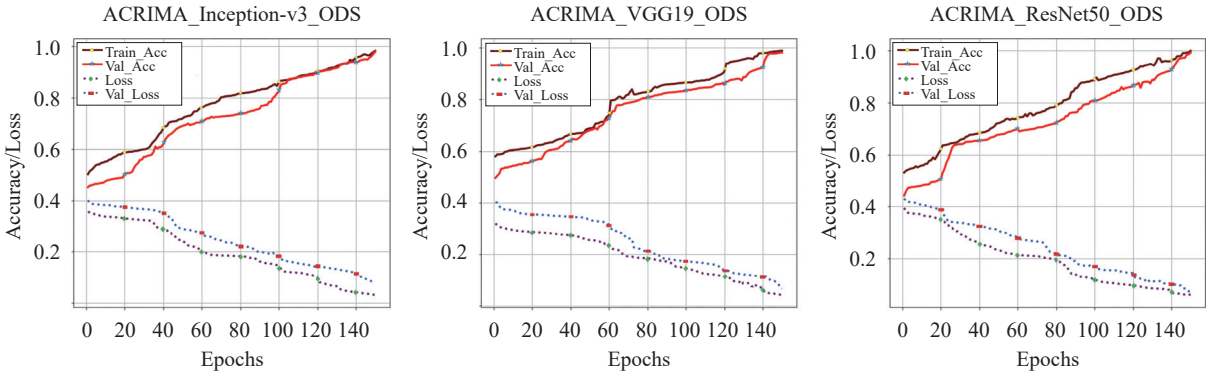



Fig. 11 Optic disc segmentation of (a) Inception-v3, (b) VGG19, (c) ResNet50 on ACRIMA

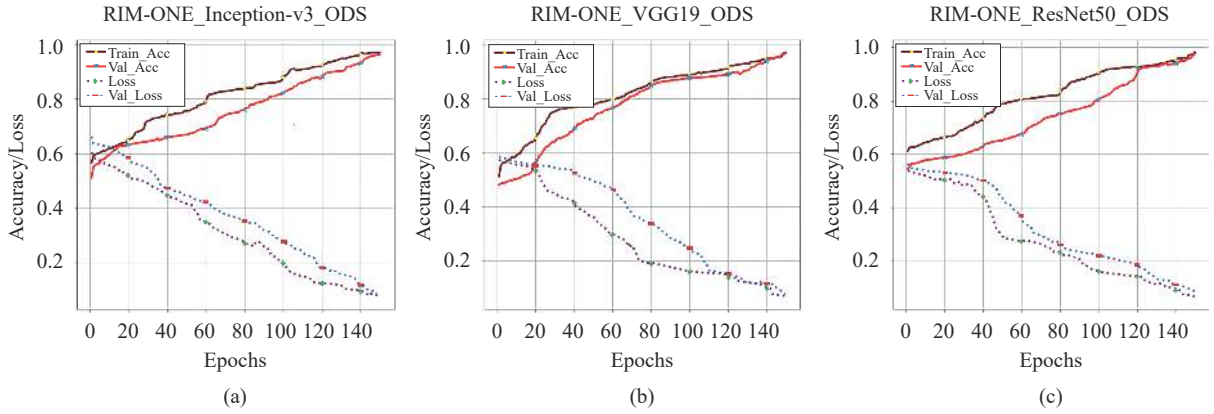



Fig. 12 Optic cup segmentation of (a) Inception-v3, (b) VGG19, (c) ResNet50 on RIM-ONE

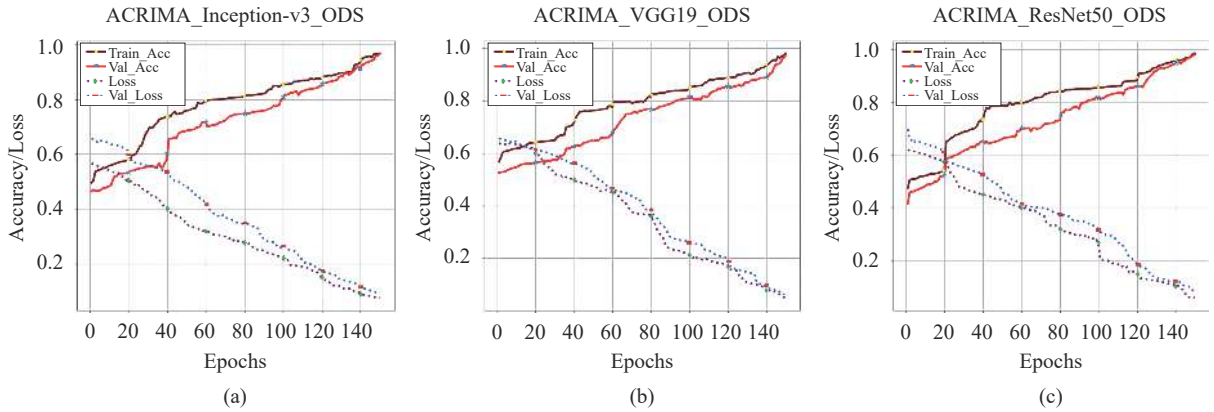



Fig. 13 Optic cup segmentation of (a) Inception-v3, (b) VGG19, (c) ResNet50 on ACRIMA

training accuracy and 99.07% validation accuracy. The training and validation accuracy given by the ResNet50 architecture were 99.89% and 99.58%, respectively. According to the training and testing results, these three models perform well on the training and validation data. Therefore, no clear signs of model overfitting were observed.

Similarly, Figs. 11(a)–11(c) show the OD segmentation of Inception-v3, VGG19, and ResNet50, respectively, on the ACRIMA dataset. Fig. 11(a) shows the accuracy of 98.26% and 97.91% for training and validation, respectively, for 150 epochs. In Fig. 11(b), VGG19 yielded 98.81% and 98.18% training and validation accuracy, respectively. Furthermore, the ResNet50 model showed training and validation accuracy of 99.84% and 99.26%, respectively. Since the models performed well on the training and validation data, there were no obvious indicators of overfitting.

Moreover, Figs. 12(a)–12(c) show the accuracy and loss in the training and testing stages of the OC segmentation of Inception-v3, VGG19, and ResNet50, respectively, on RIM-ONE. The Inception-v3 model gave the accuracy of 97.05% and 96.39% for training and validation, respectively, for 150 epochs. The training and validation accuracy yielded by the VGG19 architecture for OC segmentation was 97.12% and 96.70%, respectively. The ResNet50 model showed 97.90% and 97.42% for training and validation, respectively. Accordingly, the models perform well on the training and validation sets and overfitting is not observable.

Furthermore, Figs. 13(a)–13(c) present the results of the training and testing stages of the OC segmentation of Inception-v3, VGG19, and ResNet50, respectively, on ACRIMA. The observed training and validation accuracy for each model is as follows. Inception-v3: 96.84% and 96.65%, VGG19: 98.05% and 97.35%, ResNet50: 98.41% and 98.05%, respectively. Thus, the models perform well on both training and validation data.

Considering the model evaluation techniques, sensitivity and specificity measures the probability of a positive and negative diagnostic test, respectively[44]. As a result, sensitivity is particularly essential in medical image analysis, since it indicates the percentage of participants who have the disease that the model properly recognizes. We have illustrated each CNN architecture's DC, JC, accuracy (Acc), specificity (Spe), and sensitivity (Sen) for OD and OC segmentation. Tables 3 and 4 show the ODS and OCS results, respectively. The metrics are shown for both the RIM-ONE and ACRIMA datasets, and for each model Inception-v3, VGG19, and ResNet50. Accordingly, ResNet50 on the RIM-ONE dataset obtains 99.58% and 97.42% accuracy for the OD and OC segmentation with

Table 3 ODS results using attention U-Net with CNN backbones

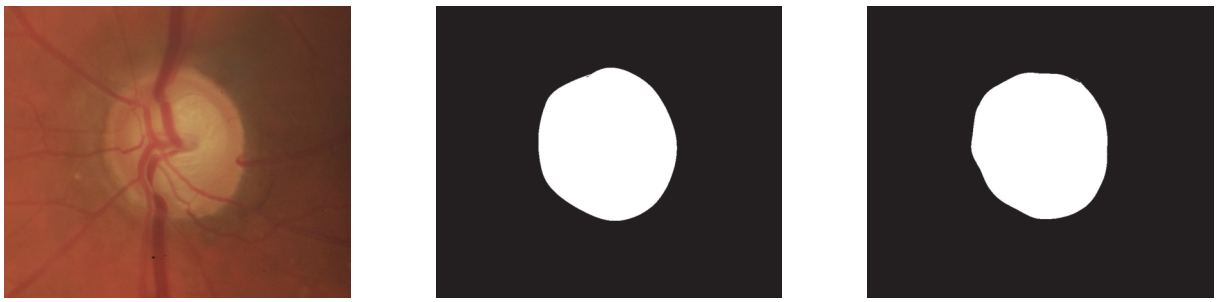
| Performance metrics | Inception-v3* | | VGG19* | | ResNet50* | |
|---|---|---|---|---|---|---|
| | RIM-ONE | ACRIMA | RIM-ONE | ACRIMA | RIM-ONE | ACRIMA |
| DC | 99.19% | 98.85% | 99.44% | 99.00% | 99.75% | 99.54% |
| JC | 98.41% | 97.74% | 98.89% | 98.03% | 99.51% | 99.09% |
| Acc | 98.65% | 97.91% | 99.07% | 98.18% | **99.58%** | 99.26% |
| Sen | 99.01% | 98.63% | 99.38% | 98.78% | 99.75% | 99.54% |
| Spe | 96.81% | 90.00% | 97.46% | 91.66% | 98.69% | 94.73% |

Table 4 OCS results using attention U-Net with CNN backbones

| Performance metrics | Inception-v3* | | VGG19* | | ResNet50* | |
|---|---|---|---|---|---|---|
| | RIM-ONE | ACRIMA | RIM-ONE | ACRIMA | RIM-ONE | ACRIMA |
| DC | 97.80% | 98.16% | 97.99% | 98.55% | 98.43% | 98.93% |
| JC | 95.69% | 96.40% | 96.06% | 97.14% | 96.91% | 97.89% |
| Acc | 96.39% | 96.65% | 96.70% | 97.35% | 97.42% | **98.05%** |
| Sen | 97.49% | 98.01% | 97.74% | 98.32% | 98.37% | 98.78% |
| Spe | 91.27% | 82.25% | 91.81% | 86.66% | 92.98% | 89.83% |

the minor test loss. Furthermore, ResNet50 on the ACRIMA dataset achieves 99.26% and 98.05% accuracy for OD and OC segmentation with less testing loss, respectively.

Furthermore, Fig. 14 shows the ground truth image sample and the predicted image generated from the trained model. Figs. 15 and 16 show the confusion matrices for OD segmentation of each architecture. Figs. 17 and 18 depict the confusion matrices for OC segmentation of each architecture based on the RIM-ONE and ACRIMA datasets, respectively. Tables 3 and 4 state the prediction results of segmentation that separates OC



(a) (b) (c)

Fig. 14 Original fundus image, ground truth image from the dataset, and the predicted image generated from the trained model: (a) Fundus image; (b) Ground truth image; (c) Predicted image.

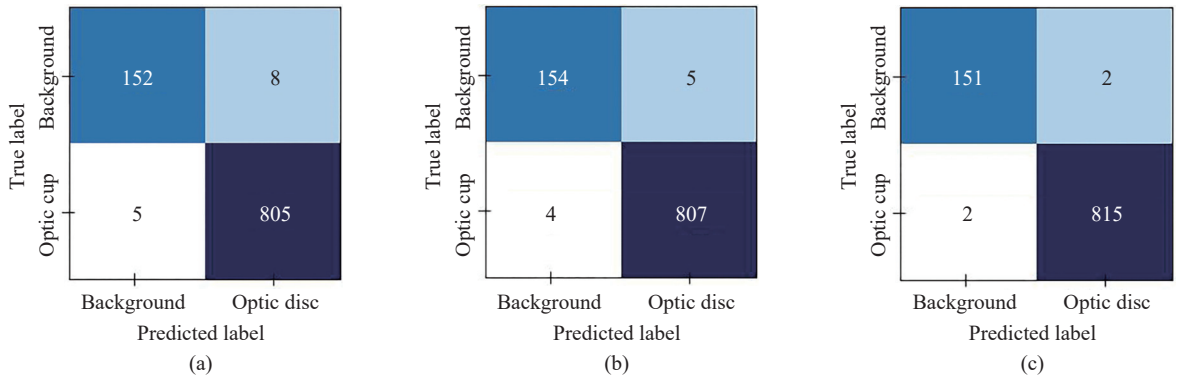



Fig. 15 Confusion matrix of (a) Inception-v3, (b) VGG19, (c) ResNet50 for OD segmentation on RIM-ONE

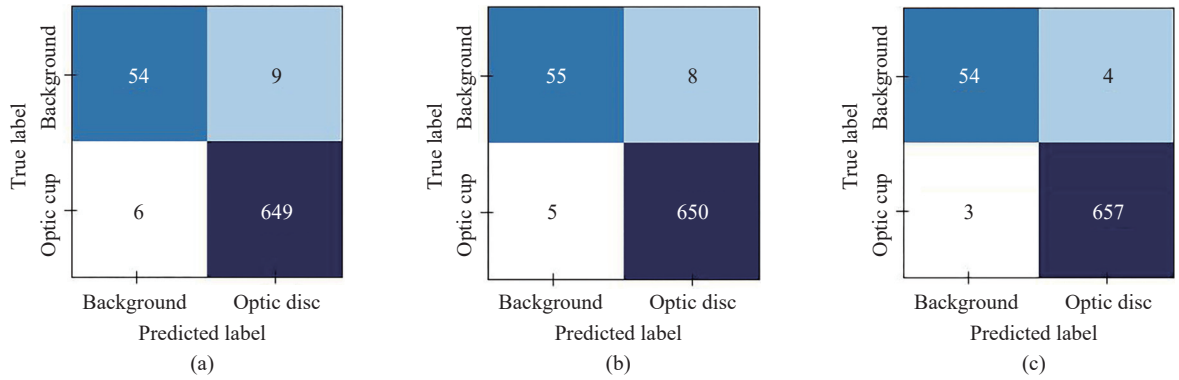



Fig. 16 Confusion matrix of (a) Inception-v3, (b) VGG19, (c) ResNet50 for OD segmentation on ACRIMA

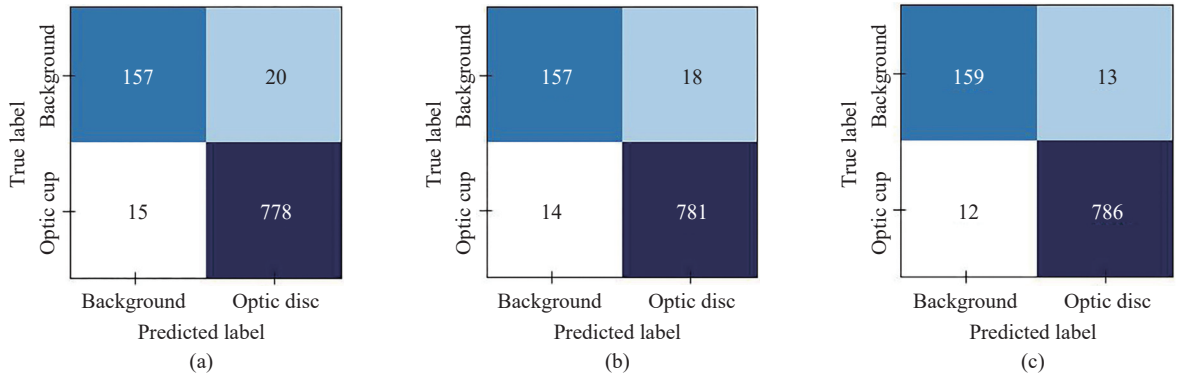



Fig. 17 Confusion matrix of (a) Inception-v3, (b) VGG19, (c) ResNet50 for OC segmentation on RIM-ONE

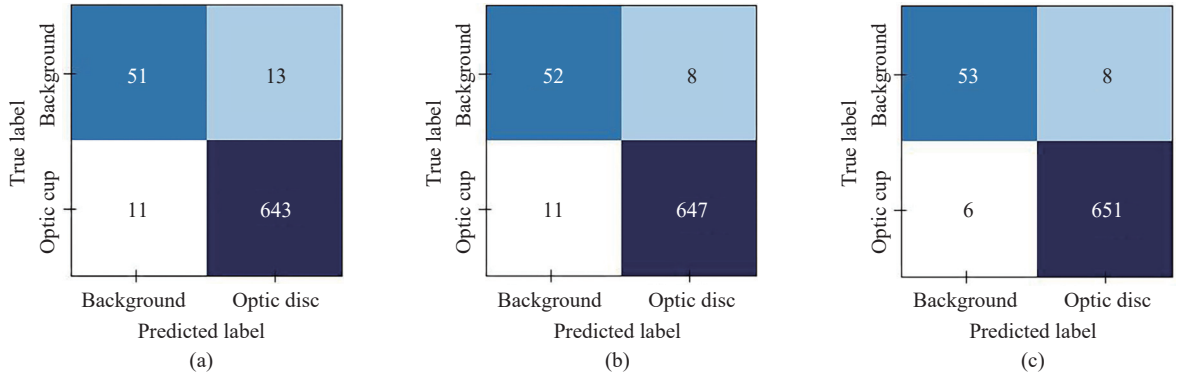



Fig. 18 Confusion matrix of (a) Inception-v3, (b) VGG19, (c) ResNet50 for OC segmentation on ACRIMA

and OD from the background of fundus images. In the confusion matrix, the top left and bottom right sections indicate true positives (TPs) and true negatives (TNs), respectively. The top right and bottom left sections show false positives (FPs) and false negatives (FNs), respectively. Accordingly, all the models produced good results, as they obtained higher values for TP and TN.

## 4.4 Segmentation followed by classification results

As the main contribution of this paper, we have obtained the performance metric for the complete glaucoma classification pipeline with segmentation. As described in the methodology, we used the segmented images obtained from ResNet50-based attention U-Net, which gave the highest accuracy in the segmentation process. Then, we classified those segmented images using the modified Inception-v3*, VGG19*, and ResNet50*. Fig. 19 shows the classification results of the three models for the RIM-ONE dataset.

The results are obtained using the same batch sizes with different optimizers for each model on 150 epochs. Fig. 19(a) shows the classification results of the Inception-v3 architecture. It obtained the accuracy of 99.42% for training and the accuracy of 98.79% for testing at the end of 200 epochs with less test loss. In Fig. 19(b), VGG19 yielded 95.74% training accuracy and 94.85% testing ac-

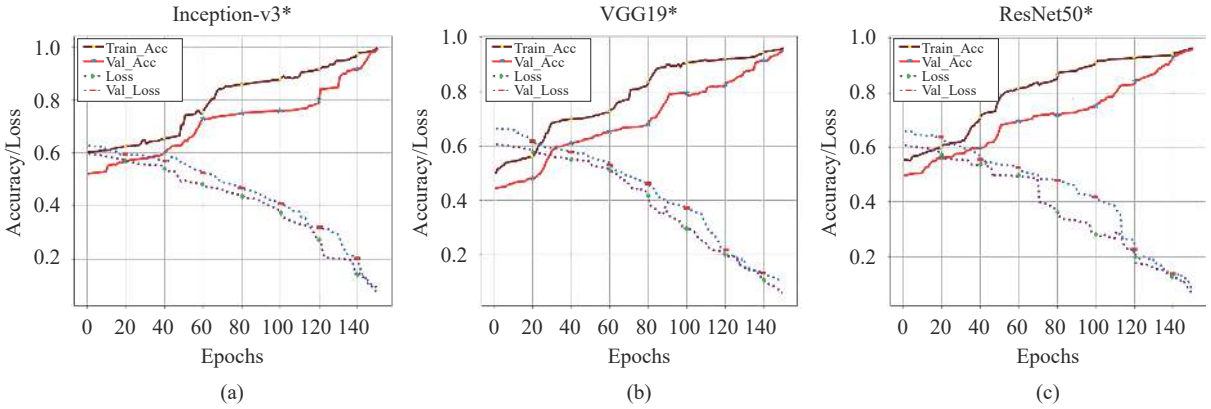



Fig. 19 Classification results of (a) Inception-v3, (b) VGG19, (c) ResNet50 models for the segmented images on the ACRIMA dataset

curacy. The training and testing accuracy given by the ResNet50 architecture, was 96.24% and 95.71%, respectively, as shown in Fig. 19(c).

However, as shown in Fig. 19, the testing accuracy is lower than the training accuracy in the initial epochs, and the testing loss is larger than the training loss. The testing loss evaluates how well the model fits new data, whereas the training loss evaluates how well the model fits the training data. The model has learned restricted characteristics during the early epochs, and as a result, initially, the performance is low. Since the training and testing loss and accuracy curves are shown to overlap at the end of the 150 epochs, it eventually indicates no significant variance or overfitting.

Moreover, the confusion matrix of the Inception-v3*, VGG19*, and ResNet50* models for the classification of the segmented images is shown in Fig. 20. Generally, a summary of the prediction outcome of a classification problem is shown in the confusion matrix. As can be observed, the model produces more accurate predictions.

Furthermore, the three model's receiver operating characteristic (ROC) curves are plotted in Fig. 21. This indicates a binary classification's diagnostic ability for changing discrimination thresholds or the likelihood of picking a positive class over a negative class. From the ROC curve, it can be observed that the curves are closer to the top-left corner. This means that with a very accurate test, the classifiers performed better. AUC, which represents the diagnostic predicting performance of the model, measures a classifier's ability to discriminate between classes.

Moreover, Table 5 represents the performances shown by each architecture on the RIM-ONE dataset. Based on each CNN architecture's performance, Inception-v3 on RIM-ONE achieves 98.79% accuracy and the least test loss.

## 5 Discussions

### 5.1 Study contributions

We presented an approach to segment the OD and OC using attention U-Net with fundus images. We have proposed three pre-trained CNN architectures to use U-Net's encoder part as the attention. Furthermore, we have employed different data pre-processing and data augmentation techniques to overcome the model overfitting and data imbalance issues. Furthermore, we extended this proposed approach for classification. We utilized segmented images generated from the highest accuracy model in the segmentation process. We obtained the accuracy of 98.58% utilizing the modified pre-trained Inception-v3 model as the best result. The main contributions of this study are as follows.

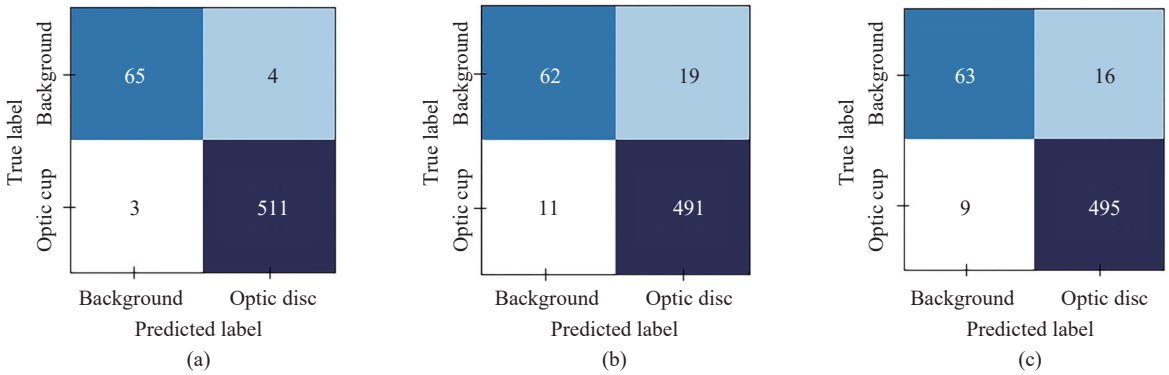



Fig. 20 Confusion matrix of (a) Inception-v3*, (b) VGG19*, (c) ResNet50* for the classification with segmentation

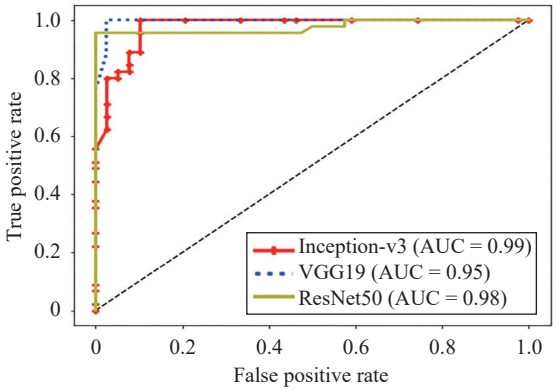



Fig. 21 ROC curve

Table 5 Performance of CNN architectures for the segmented images

| Performance | Segmentation followed by CNN architecture | | |
|---|---|---|---|
| | Inception-v3 | VGG19 | ResNet50 |
| Accuracy | 98.79% | 94.85% | 95.71% |
| F1-score | 99.31% | 97.03% | 97.53% |
| Precision | 99.41% | 97.80% | 98.21% |
| Recall | 99.22% | 96.27% | 96.86% |
| Specificity | 95.58% | 84.93% | 87.50% |
| Sensitivity | 99.22% | 96.27% | 96.86% |
| Loss | 0.071 1 | 0.100 7 | 0.081 3 |
| AUC | 0.99 | 0.95 | 0.98 |

1) This study explored literature related to retinal fundus image segmentation and classification for glaucoma detection.

2) We applied data pre-processing techniques to enhance the quality of the images and remove unwanted noise.

3) Our pre-processing approach addressed data overfitting and increased the algorithm′s generalization using the augmentation technique.

4) We designed and developed an attention U-Net-based fundus image segmentation method employing three CNN architectures, namely Inception-v3, VGG19, and ResNet50.

5) We classified the segmented images and analyze the accuracy of the classification results.

6) Our proposed approach is evaluated with different metrics and showed good performance in both segmentation and classification of fundus images.

## 5.2 Solution assessment

We have compared our work with existing studies as follows. Table 6 represents the comparison of the classification studies, Table 7 states the results comparison for the segmentation process and Table 8 compares the results with the studies that have addressed both segmentation and classification for glaucoma identification.

In Table 6, we selected the related studies that have used the same datasets namely RIM-ONE and ACRIMA, and the same DL models namely Inception-v3, VGG19 and ResNet50, as per our proposed approach for consistent comparison. As observed, utilizing the Inception-v3 model with the ACRIMA dataset, we were able to obtain the best accuracy of 98.52%.

Table 8 compares previous studies, where performance of optic cup segmentation (OCS) and optic disc segmentation (ODS) was calculated with performance matrices such as DC, JC, accuracy, sensitivity, specificity and TL referring to transfer learning with pre-trained weights. The main contribution of this research is that the OD and OC segmentation employed attention U-Net with different CNN architectures for two datasets. In addition, we performed data augmentation techniques to

Table 6 Comparison with existing studies: Classification only

| Study | Dataset | Model | Modified layers | Accuracy |
|---|---|---|---|---|
| [45] | RIM-ONE | Inception-v3 | A global average pooling layer followed by a dense layer having 128 ReLU activated neurons, a dropout of 0.7, and finally, a softmax layer was added | 92.51% |
| [36] | ACRIMA | VGG19 | Converted FC layers to GAP layer and added a softmax layer | 90.69% |
| | RIM-ONE | Inception-v3 | | 90.00% |
| | | ResNet50 | | 90.29% |
| [38] | ACRIMA | Inception-v3 | Added GAP layers and softmax layers | 93.87% |
| | RIM-ONE | ResNet50 | | 95.75% |
| | | Inception-v3 | | 71.05% |
| | | ResNet50 | | 92.11% |
| [46] | ACRIMA | Inception-v3 | Added GAP layers, dropout and softmax layers | 98.5% |
| | RIM-ONE | Inception-v3 | | 92.2% |
| Proposed work | ACRIMA | Inception-v3 | Added GAP layers, dense layers, dropout layers and softmax layers | 98.52% |
| | RIM-ONE | VGG19 | | 92.64% |
| | | ResNet50 | | 95.58% |
| | | Inception-v3 | | 96.56% |
| | | VGG19 | | 94.95% |
| | | ResNet50 | | 95.49% |

Table 7 Comparison with existing studies: Segmentation only

| Study | Dataset | Model | Type | DC (%) | JC (%) | Acc (%) | Sen (%) | Spe (%) |
|---|---|---|---|---|---|---|---|---|
| [16] | DRISHTI-GS | U-Net with | ODS | 92.64 | 86.3 | 99.48 | 95.51 | 99.62 |
| | RIM-ONE | transfer learning | OCS | 79.01 | 65.30 | 99.11 | 86.59 | 99.36 |
| | | | ODS (TL) | 96.38 | 93.01 | 99.75 | 94.88 | 99.93 |
| | | | OCS (TL) | 87.93 | 78.46 | 99.53 | 87.65 | 99.77 |
| | | | ODS (TL) | 94.01 | 88.70 | 99.61 | 92.36 | 99.86 |
| | | | OCS (TL) | 83.97 | 72.37 | 99.69 | 81.33 | 99.87 |
| [29] | DRISHTI-GS | Cup disc encoder- | OCS | 92.4 | 86.32 | 99.71 | 95.67 | 99.81 |
| | RIM-ONE | decoder network | ODS | 95.97 | 91.83 | 99.66 | 97.54 | 99.73 |
| | | (CDED-Net) | OCS | 86.22 | 75.32 | 99.61 | 95.17 | 99.81 |
| | | | ODS | 95.82 | 91.01 | 99.56 | 97.34 | 99.73 |
| [32] | RIM-ONE | U-Net with VGG16 | ODS | 94 | N/A | N/A | N/A | N/A |
| | DRISHTI-GS | architecture | OCS | 94 | N/A | N/A | N/A | N/A |
| [31] | RIM-ONE | Generalized U-Net | ODS | N/A | N/A | 94 | N/A | N/A |
| | DRISHTI-GS | | OCS | N/A | N/A | 94 | N/A | N/A |
| Proposed work | RIM-ONE | Attention U-Net with ResNet50 | ODS | 99.75 | 99.51 | 99.58 | 99.75 | 98.69 |
| | | | OCS | 98.43 | 96.91 | 97.42 | 98.37 | 92.98 |
| | ACRIMA | | ODS | 99.54 | 99.09 | 99.26 | 99.54 | 94.73 |
| | | | OCS | 98.93 | 97.89 | 98.05 | 98.78 | 89.83 |

Table 8 Comparison with existing studies: Segmentation followed by the classification

| Study | Dataset | Model | Acc (%) | Sen (%) | Spe (%) | AUC |
|---|---|---|---|---|---|---|
| [14] | ORIGA | U-Net with VGG19 | 95.54 | N/A | N/A | N/A |
| [47] | ORIGA, SCES | U-Net with Deep Neural network (DNN) | 86 | N/A | N/A | 98 |
| [48] | ORIGA | M-Net | N/A | N/A | N/A | 89 |
| [49] | REFUGE | U-Net | 90 | 85 | 91 | N/A |
| Proposed work | Attention U-Net with ResNet50 based segmented images from RIM-ONE | Inception-v3 | 98.79% | 99.22% | 95.58% | 99% |
| | | VGG19 | 94.85% | 96.27% | 84.93% | 95% |
| | | ResNet50 | 95.71% | 96.86% | 87.50% | 98% |

address the class imbalance problem and avoid model overfitting. Accordingly, the proposed attention U-Net-based OD segmentation of ResNet50 as the backbone of the RIM-ONE dataset obtained the highest accuracy of 99.58% among the other architectures.

Among the latest studies on OD and OC segmentation, many studies have accounted for the standard U-Net architecture. However, the highest accuracy of 99.75% is shown using [16], on the DRISHTI-GS dataset for OD segmentation. Then, Tabassum[29] obtained 99.71% accuracy for OC segmentation and 99.66% accuracy for OD segmentation using the DRISHTI-GS dataset. Furthermore, Sevastopolsky[32] achieved 94% accuracy for both OD and OC segmentation. Moreover, considering the related studies shown in Table 6, few authors have used the DRISHTI dataset, which consists of 101 images. However, in this study, we considered RIM-ONE and ACRIMA datasets with 942 and 705, respectively.

Moreover, Table 8 shows a comparison of our work with the related studies that have used segmentation followed by classification, to identify glaucoma using fundus images. Classification performances were evaluated using accuracy, sensitivity, specificity and AUC. Among the considered deep learning models, the proposed Inception-v3 based classification method obtained the highest accuracy at 98.79%.

Among the latest studies on segmentation followed by classification, the highest accuracy of 95.54% is shown by [14], using U-Net with the VGG19 model on the ORIGA dataset. Moreover, Li et al.[47] obtained 86% accuracy for glaucoma classification by employing U-Net with deep neural network (DNN). Fu et al.[48] showed an AUC of 89% for M-Net based glaucoma classification on the ORIGA dataset. Furthermore, Orlando et al.[49] achieved 90% accuracy for segmentation followed by glaucoma classification using the REFUGE dataset. Moreover, considering the related studies shown in Table 8, most authors have used the ORIGA and REFUGE datasets. The ORIGA dataset consists of a total of 650 fundus images from 149 normal subjects and 501 glaucoma subjects. The REFUGE dataset consists of a total of 1 200 fundus images.

Consider the overall glaucoma identification studies with segmentation and/or classification, stated in Tables 6–8. The highest accuracy of 98.5% for classification only is shown by Rehman et al.[46] using a modified Inception-v3 with the ACRIMA dataset. For the same dataset and a modified Inception-v3, our approach shows 98.52% accuracy. For the segmentation followed by classification, Sudhan et al.[14], showed an accuracy of 95.54% for the ORI-

GA dataset using U-Net and VGG19. However, our approach has shown an accuracy of 98.79% for Inception-v3 based classification of the segmented images retrieved from our attention U-Net with ResNet50 based model for the RIM-ONE dataset.

Although our study shows a slight increase in accuracy compared to existing studies, considering the global glaucoma affected population, this model can be beneficial for identifying glaucoma patients. For instance, 4.13 million people had moderate to severe visual impairment in 2020, while 3.6 million people were blind owing to glaucoma. Thus, even a slight increase in accuracy in the proposed computational model can assist in starting early treatments and save many patients with the correct identification of glaucoma conditions.

### 5.3 Current challenges and future research directions

Different types of challenges are associated with fundus image processing using deep learning methods in clinical practice[50]. Considering the proposed study, the main limitation of the segmentation process by employing deep learning or convolutional neural networks is the requirement of large training data to increase OD and OC segmentation performance. Another significant limitation encountered during the model training was time and resource consumption during the training process.

This study currently focused on classifying glaucoma and normal subjects using different models based on deep learning. The effectiveness and efficiency of the approach can be further improved by designing optimization techniques and fine-tuning the parameters. Moreover, incorporating explainability is a trending area in medical image analysis[51]. The proposed solution can be extended to a real-world usable tool with an interpretation of the predictions and the explanations of the process, in a way that can be understood by humans. In addition, there are different types of glaucoma such as open-angle, angle-closure, normal tension, and secondary; thus a multiclass classification approach can be used to identify the glaucoma category. Additionally, glaucoma has stages namely, early, moderate, and deep, where we can apply different analyses to the optic nerve of the fundus images. This will help to identify the severity of the glaucoma condition using retinal fundus images. Therefore, the presented approach may be used as a support tool for glaucoma detection with real data in clinical practice.

## 6 Conclusions

Glaucoma is a potentially blinding condition that, if left untreated, can result in visual loss over time. Therefore, early treatment is significant to prevent vision loss. Glaucoma treatment planning relies on images such as retinal fundus and OCT scans acquired from the optic nerve. Computational models can be used as a second opinion in the glaucoma diagnosis process to address the issue due to misdiagnosis and provide a better solution to classify fundus images. This paper proposed a segmentation and classification based approach to support glaucoma identification using retinal fundus images.

Using attention U-Net with different CNN based backbones, this study offered a segmentation approach for OD and OC segmentation of fundus images. We have provided three pre-trained architectures, namely, Inception, VGG and ResNet, for the encoder part of attention U-Net. We have applied different data augmentation techniques and pre-processing techniques for each architecture. On the RIM-ONE dataset, attention to U-Net-based OD segmentation of ResNet50 as the backbone achieved an accuracy of 99.58% among other architectures. Furthermore, this study performed the classification process using segmented images utilizing three modified CNN architectures. The Inception-v3 model provided the highest glaucoma classification accuracy of 98.79% among the other CNN architectures.

## References

[1] World Health Organization. World report on vision, [Online], Available: https://apps.who.int/iris/bitstream/handle/10665/328717/9789241516570-eng.pdf?sequence=18&isAllowd=y, 2019.

[2] L. Storgaard, T. L. Tran, J. C. Freiberg, A. S. Hauser, M. Kolko. Glaucoma clinical research: Trends in treatment strategies and drug development. *Frontiers in Medicine*, vol. 8, Article number 733080, 2021. DOI: 10.3389/fmed.2021.733080.

[3] I. B. Senkyire, Z. Liu. Supervised and semi-supervised methods for abdominal organ segmentation: A review. *International Journal of Automation and Computing*, vol. 18, no. 6, pp. 887–914, 2021. DOI: 10.1007/s11633-021-1313-0.

[4] I. Rubasinghe, D. Meedeniya. Ultrasound nerve segmentation using deep probabilistic programming. *Journal of ICT Research and Applications*, vol. 13, no. 3, pp. 241–256, 2019. DOI: 10.5614/itbj.ict.res.appl.2019.13.3.5.

[5] Y. Prajna, M. K. Nath. Efficient blood vessel segmentation from color fundus image using deep neural network. *Journal of Intelligent & Fuzzy Systems*, vol. 42, no. 4, pp. 3477–3489, 2022. DOI: 10.3233/jifs-211479.

[6] J. Kim, L. Tran, E. Y. Chew, S. Antani. Optic disc and cup segmentation for glaucoma characterization using deep learning. In *Proceedings of the 32nd International Symposium on Computer-Based Medical Systems*, IEEE, Córdoba, Spain, pp. 489–494, 2019. DOI: 10.1109/cbms.2019.00100.

[7] R. Mahum, S. Ur Rehman, O. D. Okon, A. Alabrah, T. Meraj, H. T. Rauf. A novel hybrid approach based on deep CNN to detect glaucoma using fundus imaging. *Electronics*, vol. 11, no. 1, Article number 26, 2022. DOI: 10.3390/electronics11010026.

[8] G. Ariyarathne, S. De Silva, S. Dayarathna, D. Meedeniya, S. Jayarathne. ADHD identification using convolutional neural network with seed-based approach for fMRI data. In *Proceedings of the 9th International Conference on Software and Computer Applications*, ACM, Langkawi,

Malaysia, pp. 31–35, 2020. DOI: 10.1145/3384544.3384552.

[9] C. Z. Wu, J. Sun, J. Wang, L. F. Xu, S. Zhan. Encoding-decoding network with pyramid self-attention module for retinal vessel segmentation. *International Journal of Automation and Computing*, vol. 18, no. 6, pp. 973–980, 2021. DOI: 10.1007/s11633-020-1277-0.

[10] B. B. Jia, M. L. Zhang. Multi-dimensional classification via selective feature augmentation. *Machine Intelligence Research*, vol. 19, no. 1, pp. 38–51, 2022. DOI: 10.1007/s11633-022-1316-5.

[11] P. Elangovan, D. Vijayalakshmi, M. K. Nath. Detection of glaucoma from fundus image using pre-trained Densenet201 model. *Indian Journal of Radio & Space Physics*, vol. 50, no. 1, pp. 33–39, 2021.

[12] T. Shyamalee, D. Meedeniya. CNN based fundus images classification for glaucoma identification. In *Proceedings of the 2nd International Conference on Advanced Research in Computing*, IEEE, Belihuloya, Sri Lanka, pp. 200–205, 2022. DOI: 10.1109/ICARC54489.2022.9754171.

[13] P. Elangovan, M. K. Nath. Glaucoma assessment from color fundus images using convolutional neural network. *International Journal of Imaging Systems and Technology*, vol. 31, no. 2, pp. 955–971, 2021. DOI: 10.1002/ima.22494.

[14] M. B. Sudhan, M. Sinthuja, S. Pravinth Raja, J. Amutharaj, G. Charlyn Pushpa Latha, S. Sheeba Rachel, T. Anitha, T. Rajendran, Y. Asrat Waji. Segmentation and classification of glaucoma using U-net with deep learning model. *Journal of Healthcare Engineering*, vol. 2022, Article number 1601354, 2022. DOI: 10.1155/2022/1601354.

[15] O. Ronneberger, P. Fischer, T. Brox. U-Net: Convolutional networks for biomedical image segmentation. In *Proceedings of the 18th International Conference on Medical image Computing and Computer-assisted Intervention*, Springer, Munich, Germany, pp. 234–241, 2015. DOI: 10.1007/978-3-319-24574-4_28.

[16] X. Y. Zhao, S. Y. Wang, J. Zhao, H. C. Wei, M. X. Xiao, N. Ta. Application of an attention U-Net incorporating transfer learning for optic disc and cup segmentation. *Signal, Image and Video Processing*, vol. 15, no. 5, pp. 913–921, 2021. DOI: 10.1007/s11760-020-01815-z.

[17] T. Shyamalee, D. Meedeniya. Attention U-Net for glaucoma identification using fundus image segmentation. In *Proceedings of International Conference on Decision Aid Sciences and Applications*, IEEE, Chiangrai, Thailand, pp. 6–10, 2022. DOI: 10.1109/dasa54658.2022.9765303.

[18] L. Herath, D. Meedeniya, M. A. J. C. Marasingha, V. Weerasinghe. Autism spectrum disorder diagnosis support model using Inception-v3. In *International Research Conference on Smart Computing and Systems Engineering*, IEEE, Colombo, Sri Lanka, pp. 1–7, 2021. DOI: 10.1109/SCSE53661.2021.9568314.

[19] C. Fernando, S. Kolonne, H. Kumarasinghe, D. Meedeniya. Chest radiographs classification using multi-model deep learning: A comparative study. In *Proceedings of the 2nd International Conference on Advanced Research in Computing*, IEEE, Belihuloya, Sri Lanka, pp. 165–170, 2022. DOI: 10.1109/ICARC54489.2022.9753811.

[20] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, Z. Wojna. Rethinking the inception architecture for computer vision. In *Proceedings of Conference on Computer Vision and Pattern Recognition*, Las Vegas, USA, pp. 2818–2826, 2016. DOI: 10.1109/cvpr.2016.308.

[21] K. Simonyan, A. Zisserman. Very deep convolutional networks for large-scale image recognition. In *Proceedings of the 3rd International Conference on Learning Representations*, San Diego, USA, 2015.

[22] K. M. He, X. Y. Zhang, S. Q. Ren, J. Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Las Vegas, USA, pp. 770–778, 2016. DOI: 10.1109/CVPR.2016.90.

[23] T. Iqbal, H. Ali. Generative adversarial network for medical images (MI-GAN). *Journal of Medical Systems*, vol. 42, no. 11, Article number 231, 2018. DOI: 10.1007/s10916-018-1072-9.

[24] S. J. Wang, L. Q. Yu, X. Yang, C. W. Fu, P. A. Heng. Patch-based output space adversarial learning for joint optic disc and cup segmentation. *IEEE Transactions on Medical Imaging*, vol. 38, no. 11, pp. 2485–2495, 2019. DOI: 10.1109/tmi.2019.2899910.

[25] M. Z. Alom, C. Yakopcic, M. Hasan, T. M. Taha, V. K. Asari. Recurrent residual U-Net for medical image segmentation. *Journal of Medical Imaging*, vol. 6, no. 1, Article number 014006, 2019. DOI: 10.1117/1.jmi.6.1.014006.

[26] A. You, J. K. Kim, I. H. Ryu, T. K. Yoo. Application of generative adversarial networks (GAN) for ophthalmology image domains: A survey. *Eye and Vision*, vol. 9, no. 1, Article number 6, 2022. DOI: 10.1186/s40662-022-00277-3.

[27] G. Litjens, T. Kooi, B. E. Bejnordi, A. A. A. Setio, F. Ciompi, M. Ghafoorian, J. A. W. M. van der Laak, B. van Ginneken, C. I. Sánchez. A survey on deep learning in medical image analysis. *Medical Image Analysis*, vol. 42, pp. 60–88, 2017. DOI: 10.1016/j.media.2017.07.005.

[28] H. Z. Fu, J. Cheng, Y. W. Xu, D. W. K. Wong, J. Liu, X. C. Cao. Joint optic disc and cup segmentation based on multi-label deep network and polar transformation. *IEEE Transactions on Medical Imaging*, vol. 37, no. 7, pp. 1597–1605, 2018. DOI: 10.1109/tmi.2018.2791488.

[29] M. Tabassum, T. M. Khan, M. Arsalan, S. S. Naqvi, M. Ahmed, H. A. Madni, J. Mirza. CDED-Net: Joint segmentation of optic disc and optic cup for glaucoma screening. *IEEE Access*, vol. 8, pp. 102733–102747, 2020. DOI: 10.1109/access.2020.2998635.

[30] B. X. Jin, P. P. Liu, P. Wang, L. D. Shi, J. Zhao. Optic disc segmentation using attention-based U-Net and the improved cross-entropy convolutional neural network. *Entropy*, vol. 22, no. 8, Article number 844, 2020. DOI: 10.3390/e22080844.

[31] J. Civit-Masot, F. Luna-Perej ón, S. Vicente-Díaz, J. M. R. Corral, A. Civit. TPU cloud-based generalized U-Net for eye fundus image segmentation. *IEEE Access*, vol. 7, pp. 142379–142387, 2019. DOI: 10.1109/access.2019.2944692.

[32] A. Sevastopolsky. Optic disc and cup segmentation methods for glaucoma detection with modification of U-Net convolutional neural network. *Pattern Recognition and Image Analysis*, vol. 27, no. 3, pp. 618–624, 2017. DOI: 10.1134/s1054661817030269.

[33] S. De Silva, S. U. Dayarathna, G. Ariyarathne, D. Meedeniya, S. Jayarathna. fMRI feature extraction model for ADHD classification using convolutional neural network. *International Journal of E-Health and Medical Communications*, vol. 12, no. 1, Article number 6, 2021. DOI: 10.4018/IJEHMC.2021010106.

[34] D. Haputhanthri, G. Brihadiswaran, S. Gunathilaka, D. Meedeniya, S. Jayarathna, M. Jaime, C. Harshaw. Integration of facial thermography in EEG-based classification of ASD. *International Journal of Automation and Computing*, vol. 17, no. 6, pp. 837–854, 2020. DOI: 10.1007/s11633-020-1231-6.

[35] I. D. Rubasinghe, D. A. Meedeniya. Automated neuroscience decision support framework. *Deep Learning Techniques for Biomedical and Health Informatics*, B. Agarwal, V. E. Balas, L. C. Jain, R. C. Poonia, Manisha, Eds., Amsterdam, Netherlands: Elsevier Academic Press, pp. 305–326, 2020. DOI: 10.1016/B978-0-12-819061-6.00013-6.

[36] A. Diaz-Pinto, S. Morales, V. Naranjo, T. Köhler, J. M. Mossi, A. Navea. CNNs for automatic glaucoma assessment using fundus images: An extensive validation. *BioMedical Engineering OnLine*, vol. 18, no. 1, Article number 29, 2019. DOI: 10.1186/s12938-019-0649-y.

[37] K. A. Thakoor, X. H. Li, E. Tsamis, P. Sajda, D. C. Hood. Enhancing the accuracy of glaucoma detection from OCT probability maps using convolutional neural networks. In *Proceedings of the 41st Annual International Conference of IEEE Engineering in Medicine and Biology Society*, Berlin, Germany, pp. 2036–2040, 2019. DOI: 10.1109/embc.2019.8856899.

[38] S. Sreng, N. Maneerat, K. Hamamoto, K. Y. Win. Deep learning for optic disc segmentation and glaucoma diagnosis on retinal images. *Applied Sciences*, vol. 10, no. 14, Article number 4916, 2020. DOI: 10.3390/app10144916.

[39] M. N. Bajwa, M. I. Malik, S. A. Siddiqui, A. Dengel, F. Shafait, W. Neumeier, S. Ahmed. Two-stage framework for optic disc localization and glaucoma classification in retinal fundus images using deep learning. *BMC Medical Informatics and Decision Making*, vol. 19, no. 1, Article number 136, 2019. DOI: 10.1186/s12911-019-0842-8.

[40] H. N. Veena, A. Muruganandham, T. Senthil Kumaran. A novel optic disc and optic cup segmentation technique to diagnose glaucoma using deep learning convolutional neural network over retinal fundus images. *Journal of King Saud University — Computer and Information Sciences*, to be published. DOI: 10.1016/j.jksuci.2021.02.003.

[41] S. De Silva, S. Dayarathna, G. Ariyarathne, D. Meedeniya, S. Jayarathna, A. M. P. Michalek. Computational decision support system for ADHD identification. *International Journal of Automation and Computing*, vol. 18, no. 2, pp. 233–255, 2021. DOI: 10.1007/s11633-020-1252-1.

[42] Z. A. Zhang, C. D. Wu, S. Coleman, D. Kerr. DENSE- INception U-net for medical image segmentation. *Computer Methods and Programs in Biomedicine*, vol. 192, Article number 105395, 2020. DOI: 10.1016/j.cmpb.2020.105395.

[43] Y. F. Xing, L. Zhong, X. Zhong. An encoder-decoder network based FCN architecture for semantic segmentation. *Wireless Communications and Mobile Computing*, vol. 2020, Article number 8861886, 2020. DOI: 10.1155/2020/8861886.

[44] M. J. Ankenbrand, L. Shainberg, M. Hock, D. Lohr, L. M. Schreiber. Sensitivity analysis for interpretation of machine learning based segmentation models in cardiac MRI. *BMC Medical Imaging*, vol. 21, no. 1, Article number 27, 2021. DOI: 10.1186/s12880-021-00551-1.

[45] A. Singh, S. Sengupta, V. Lakshminarayanan. Glaucoma diagnosis using transfer learning methods. In *Proceedings of SPIE 11139, Applications of Machine Learning*, San Diego, USA, pp. 1–11, 2019. DOI: 10.1117/12.2529429.

[46] Aziz-ur-Rehman, I. A. Taj, M. Sajid, K. S. Karimov. An ensemble framework based on Deep CNNs architecture for glaucoma classification using fundus photography. *Mathematical Biosciences and Engineering*, vol. 18, no. 5, pp. 5321–5346, 2021. DOI: 10.3934/mbe.2021270.

[47] S. Y. Li, Z. Li, L. M. Guo, G. B. Bian. Glaucoma Detection: Joint segmentation and classification framework via deep ensemble network. In *Proceedings of the 5th International Conference on Advanced Robotics and Mechatronics*, IEEE, Shenzhen, China, pp. 678–685, 2020. DOI: 10.1109/icarm49381.2020.9195312.

[48] H. Z. Fu, J. Cheng, Y. W. Xu, C. Q. Zhang, D. W. K. Wong, J. Liu, X. C. Cao. Disc-aware ensemble network for glaucoma screening from fundus image. *IEEE Transactions on Medical Imaging*, vol. 37, no. 11, pp. 2493–2501, 2018. DOI: 10.1109/tmi.2018.2837012.

[49] J. I. Orlando, H. Z. Fu, B. Breda, K. van Keer, D. R. Bathula, A. Diaz-Pinto, R. G. Fang, P. A. Heng, J. Kim, J. H. Lee, J. Lee, X. X. Li, P. Liu, S. Lu, B. Murugesan, V. Naranjo, S. S. R. Phaye, S. M. Shankaranarayana, A. Sikka, J. Son, A. van den Hengel, S. J. Wang, J. Y. Wu, Z. F. Wu, G. H. Xu, Y. L. Xu, P. S. Yin, F. Li, X. L. Zhang, Y. W. Xu, H. Bogunovic. REFUGE challenge: A unified framework for evaluating automated methods for glaucoma assessment from fundus photographs. *Medical Image Analysis*, vol. 59, Article number 101570, 2018. DOI: 10.1016/j.media.2019.101570.

[50] G. Lim, W. Hsu, M. L. Lee, D. S. W. Ting, T. Y. Wong. Technical and clinical challenges of AI in retinal image analysis. *Computational Retinal Image Analysis: Tools, Applications and Perspectives*, E. Trucco, T. MacGillivray, Y. W. Xu, Eds., Amsterdam, Netherlands: Elsevier Academic Press, pp. 445–466, 2019. DOI: 10.1016/b978-0-08-102816-2.00022-8.

[51] S. Dasanayaka, S. Silva, V. Shantha, D. Meedeniya, T. Ambegoda. Interpretable machine learning for brain tumor analysis using MRI. In *Proceedings of the 2nd International Conference on Advanced Research in Computing*, IEEE, Belihuloya, Sri Lanka, pp. 212–217, 2022. DOI: 10.1109/ICARC54489.2022.9754131.

**Thisara Shyamalee** received the B.Sc. degree in information technology from Sri Lanka Institute of Information Technology (SLIIT), Sri Lanka in 2019. She is currently a master student in computer science (with a major component of research) at Department of Computer Science and Engineering, University of Moratuwa, Sri Lanka.

Her research interests include machine learning, deep learning, and computer interaction.

E-mail: thisaras@cse.mrt.ac.lk

ORCID iD: 0000-0001-7533-779X

**Dulani Meedeniya** received the Ph.D. degree in computer science from University of St Andrews, UK in 2013. She is a professor in computer science and engineering at the University of Moratuwa, Sri Lanka. She is the director of the Bio-Health Informatics Group in her department and engages in many collaborative projects. She is a co-author of 100+ publications in indexed journals, peer-reviewed conferences and international book chapters. She has received several awards and grants for her contribution in research. She serves as a reviewer, program committee and editorial team member in many international conferences and journals. She is a fellow of HEA (UK), MIET, senior member IEEE, member ACM and a chartered engineer registered at EC (UK).

Her research interests include software modelling and design, bio-health informatics, deep learning and technology-enhanced learning.

E-mail: dulanim@cse.mrt.ac.lk (Corresponding author)

ORCID iD: 0000-0002-4520-3819