Received: 26 September 2023 | Revised: 13 January 2024 | Accepted: 31 January 2024

DOI: 10.1002/eng2.12866

RESEARCH ARTICLE

Engineering Reports Open Access WILEY

Diagnosis of glaucoma using multi-scale attention block in convolution neural network and data augmentation techniques

Hamid Reza Khajeha[1] | **Mansoor Fateh**[1] | **Vahid Abolghasemi**[2]

[1]Faculty of Computer Engineering, Shahrood University of Technology, Shahrud, Iran

[2]School of Computer Science and Electronic Engineering, University of Essex, Colchester, UK

Correspondence
Mansoor Fateh, Faculty of Computer Engineering, Shahrood University of Technology, Iran.
Email: mansoor_fateh@shahroodut.ac.ir

Abstract

Glaucoma is defined as an eye disease leading to vision loss due to the optic nerve damage. It is often asymptomatic, thus, timely diagnosis and treatment is crucial. In this article, we propose a novel approach for diagnosing glaucoma using deep neural networks, trained on fundus images. Our proposed approach involves several key steps, including data sampling, pre-processing, and classification. To address the data imbalance issue, we employ a combination of suitable image augmentation techniques and Multi-Scale Attention Block (MAS Block) architecture in our deep neural network model. The MAS Block is a specific architecture design for CNNs that allows multiple convolutional filters of various sizes to capture features at several scales in parallel. This will prevent the over-fitting problem and increases the detection accuracy. Through extensive experiments with the ACRIMA dataset, we demonstrate that our proposed approach achieves high accuracy in diagnosing glaucoma. Notably, we recorded the highest accuracy (97.18%) among previous studies. The results from this study reveal the potential of our approach to improve early detection of glaucoma and offer more effective treatment strategies for doctors and clinicians in the future. Timely diagnosis plays a crucial role in managing glaucoma since it is often asymptomatic. Our proposed method utilizing deep neural networks shows promise in enhancing diagnostic accuracy and aiding healthcare professionals in making informed decisions.

KEYWORDS

ACRIMA, data augmentation, deep neural networks, glaucoma, MAS block

1 | INTRODUCTION

Glaucoma damages to the vision area causing vision loss.[1] Sixty five million people worldwide are affected by glaucoma.[2] Glaucoma can lead to serious damage or even complete loss of optic nerve fiber.[3] The front of eye is filled with a clear liquid called aqueous humor which is continuously produced and expelled from the eye to maintain constant pressure. If the fluid is not properly secreted out of the eye, the pressure rises and leads to glaucoma. If the eye pressure (called

intraocular pressure (IOP))[4] increases from its normal value, it can lead to blood flow loss to the optic nerve. Popular conventional examinations for the diagnosis of glaucoma are IOP evaluation, visual field test, evaluation of the head of optic nerve, and gonioscopy. However, these techniques are not accurate enough for the diagnosis of glaucoma, especially at early stages. As a result, visual impairment can even occur without a noticeable increase in the IOP and hence glaucoma is often known to be asymptomatic, thus, early diagnosis of glaucoma is essential.

To further observe the eye structure, Figure 1 shows the image of the fundus, the inner view of the eye, including the retina, optic nerve, macula, and blood vessels. The head of optic nerve is where the axons of ganglion cells leave the eye and form the optic disc. The optic disc is divided into two bright and central areas called the outer rim and the cup, respectively. The optic cup is placed in the center of the optic disc (Figure 1A) and is a white cup-like area, and a peripheral part is called the retinal nerve edge.[5]

The abnormal cup size compared to the optical disc is a feature of glaucomatous eyes (Figure 1B). There are different image processing approaches to cup and disc splitting to diagnose glaucoma in colored images of fundus. Some works focus only on the optical cup or optical disc split,[6,7] while others focus on calculating the ratio of the cup diameter to disc (CDR),[8] which is used as glaucoma indicator. However, measuring the cup to disc ratio requires a great deal of effort to obtain an optical disc and segment the appropriate cup disc.

Classification of the medical images using deep learning models has become very popular in different applications. However, accurate analysis of medical images using deep learning models requires access to large datasets. In medical applications, the distribution of collected samples in the dataset is normally unbalanced due to the unbalanced prevalence of each disease class among patients. Therefore, the diagnosis of diseases with insufficent and unbalanced datasets using deep learning can be very challenging. In other words, it is difficult to obtain sufficient information for the training phase of a neural network in some disease. In such circumstances, a reliable method of data augmentation will counteract the effects of unbalanced data and improves the overall performance by preventing over-fitting.[9] This has motivated us to conduct our research on this topic.

We proposed a novel model using a combination of suitable MAS Block and image augmentation techniques. The MAS Block consists of multiple parallel convolutional layers. This block has different filter sizes, followed by some concatenation and pooling operations. By using filters of different sizes, the MAS Block enables the network to learn features at varying levels. The pooling operations help reduce spatial dimensions, while the concatenation operation combines the outputs of different convolutional layers into a single tensor.

This architecture allows the network to capture both fine-grained details and broader contextual information, leading to better representation learning and improved performance in numerous applications including image classification and object detection.[10]

We also employed data augmentation to increase the size of the glaucoma dataset and hence achieving a reliable diagnosis performance. To this end, we propose a new combination of data augmentation techniques that significantly increases the classification accuracy with normalized parameters. This new combination includes cropping, rotation, flip and zoom data.

Through our extensive experiments, it is shown that the obtained diagnosis accuracy when using the proposed approach is higher than that of pre-trained networks. In the proposed method, as opposed to traditional approaches, there is no need to use handcrafted features or measure the geometric optic nerve head structures such as cup to disc ratio,

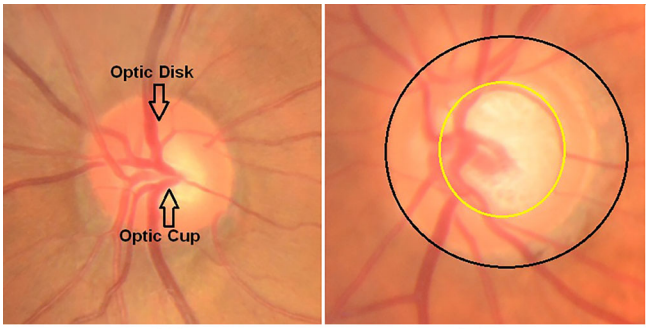

FIGURE 1 The main structures of the head of optic nerve that can be seen in of the fundus. It consists of (A) a healthy optical disk and (B) an optical disc glaucomatous.

which normally comes with error. Instead, the disease can be diagnosed using the features automatically extracted from raw images in the dataset.

In the next section, the related works for glaucoma diagnosis is provided. Then, the proposed method will be described in Section 3. Section 4 is devoted to presentation of the performance evaluation of the proposed method. Finally, we will present the discussion and conclusion in Sections 5 and 6, respectively.

2 | RELATED WORKS

To date, many algorithms for segmenting different layers of the retina have been proposed to diagnose glaucoma. Dividing the retinal layers helps ophthalmologists diagnose eye diseases accurately. However, accurate extraction of the retinal layer is very important for calculating various diagnostic parameters such as cup to disc ratio. Current techniques for diagnosing glaucoma based on retinal layer extraction fall into three categories. The first group is related to cup and disc area detection algorithms, and CDR index determination. The second group introduces deep learning algorithms based on CNN for the automatic extraction of features in retinal layers. The third group is deep learning algorithms based on transfer learning using pre-trained networks.[11]

Diagnosis of glaucoma needs the calculation of three indicators: optic disc, optic cup, and CDR. The CDR is often calculated as the ratio of the vertical diameters of the cup and the disc. To find the exact size of the CDR, one usually needs to calculate the diameter of optical disc and optic cup exactly. In addition, three-dimensional images may be used for this purpose. However, the determination of the CDR for glaucoma diagnosis is not applicable for all patients. Yuji Hatanaka et al. have proposed a method for diagnosing glaucoma using the CDR in two-dimensional retinal images. This method has two steps. First, fundus images are captured using the retinal fundus camera, and the optic disc area is extracted. In the Second step, the CDR index is determined.

Manual segmentation of the optical disc and optic cup images needs much time. Hence, scores of methods are provided to automatically split the optical disc and optic cup. Optical disc splitting methods are usually categorized into pattern-based, modular-based, and pixel-based classification methods. In most pattern-based methods,[12,13] the optical disc is approximated in a circle or ellipse, and therefore the Hough transform is used. In Reference 12, for example, the edge detection was performed by Sobel method, and then the approximate margin of the optical disc was determined by Hough transform. To perform optical disc segmentation, Aquino et al.[14] simultaneously used morphological and edge detection methods along with the circular Hough transform. Regarding model-based deformable methods, Lowell and colleagues[14] found the optic disc first. Afterwards, the optical disc was divided using a contour deformation model, which used a universal elliptical model and a local deformation model. To reduce the changes in the optical disc area, Joshi et al.[15] used strong multidimensional feature spaces extracted from points of interest. In the case of pixel-based classification methods, in Reference 16 proposed a method for pixel classification to split the optical disc and optic cup. However, to train the classification module, selecting pixels and extracting properties from a large number of pixels is challenging. Cheng et al.[17] reduced the number of pixels, optical disc, and optic cup using the superpixel strategy.[18] This method is used to increase the performance of optical disc splitting in superpixel classifications. The optical cup segmentation is more challenging than optical disc segmentation due to the blurring of optical cup boundaries. Table 1 summarizes the properties of the related works in the literature.

Retinal fundus images usually use two types of image surface information consist of paleness, and curvature of the arteries to determine the boundary of the optic cup.[19] In Reference 20, the researchers used fundus images to the intensity of the optic cup extraction threshold. Wong et al.[8] combined a variational level set and thresholding to divide the optical cup. All of these methods classified the optic cup using pallor information. Because the inlet vessels to the cup often bend, blood vessels were useful to detect the optic cup.[15,21] Many other methods such as superpixel were also used to divide the optic cup. Y. Xu et al.[22] described superpixel classification as a low-grade superpixel representation problem.

Another way to diagnose glaucoma is to use convolution neural networks (CNN).[23] These networks are focused on new algorithms for automatic feature extraction.[24] Bock et al. suggested a data-driven approach.[25] Their method is not based on the accurate measurement of the CDR index. Instead, they used "special images" idea to obtain features that are Support Vector Machine (SVM)-classified. Evaluation of the algorithm on 575 from the Erlangen Glaucoma Registry (EGR) indicated an accuracy of 88%. However, these images are not publically available, and the method cannot be compared with other methods.

Researchers have used CNN in areas such as artificial intelligence and machine vision. However, their importance was not noticed until the introduction of ImageNet in 2012, whose primary purpose was to estimate the content of real

TABLE 1 Summary of the previous studies considering glaucoma prediction, screening and/or diagnosis.

Authors	Diagnosis	Screening	Type of dataset	Dataset	Classifier/ Detection	Advantages and disadvantages
(Muramatsu et al. 2009)[38]	Yes	No	Images	Fundus image	Determination of cup-to-disc ratio	Identifying strong edges within the depth structure (+) Use of a small dataset (−)
(Hatanaka et al., 2011)[39]	Yes	No	Images	Fundus image	Determination of cup-to-disc ratio	Achieving a similarity rate of 85% relative to the disc areas (+) Frequently failing to erase the blood vessels in the nasal side (−)
(Joshi et al., 2011)[15]	Yes	No	Images	Fundus image	Determination of cup-to-disc ratio	Robust OD segmentation using active contour (+) The obscurity in 2-D information (−)
(Mookiah et al., 2012)[40]	Yes	Yes	Images	Fundus images	SVM	Designing a combined feature extraction approach (HOS-DWT) for digital fundus images (+) lack of a solid guideline leading to diverse evaluation scores among ophthalmologists (−)
(Pathan et al., 2021)[41]	Yes	Yes	Images	Fundus images	ANN, SVM, AdaBoost	Designing an efficient dynamic classifier selection method (+) Only applicable to a limited type of structural changes in blood vessel occurring due to glaucoma. (−)
(P Elangovan et al., 2022)[42]	Yes	No	Images	Fundus images	CNN	Detecting glaucoma from colored fundus images using an ensemble of deep CNN (+) High computational complexity, lack of interpretability, and potential limitations due to data availability and quality (−)
(P Elangovan et al., 2021)[43]	Yes	No	Images	Fundus images	CNN	Analyzing colored fundus images using a CNN, with improved accuracy and efficiency (+) Relying on a large image dataset for training stage (−)
(MK Nath et al., 2021)[44]	Yes	No	Images	Retinal images	Segmentation of cup	Analyzing the information content through wavelet-transformed signals, for understanding the condition's characteristics (+) High complexity of interpreting the differential entropy values within the wavelet sub-bands (+) Limited use in practical applications (−)
(P Elangovan et al., 2020)[45]	Yes	No	Images	Fundus images	Segmented optic disc and optic cup	Using simple statistical parameters for glaucoma detection from color fundus images (+) Dependency on the availability of high-quality and diverse datasets for training (−)
(M Madhusudhan et al., 2011)[46]	Yes	No	Images	Fundus images	Cup disc ratio evaluation	Improved accuracy and efficiency in glaucoma detection through the use of image processing techniques (+) Limited effectiveness on low quality images (−)
(T Shyamalee et al., 2022)[47]	Yes	No	Images	Fundus images	CNN	Employing a CNN model for accurate and efficient identification of glaucoma (+) Requiring a large dataset and high computational resources (−)

images for automatic detection using the ImageNet dataset.[26] Their success was achieved using Graphical Processing Units (GPU) activation functions such as ReLU, data augmentation methods, and dropout technique.[27] The main strength of CNN architecture is its ability to automatically extract features.[28,29]

Proper training of CNNs needs labeled data. This is a limiting factor to address in assessing glaucoma and computational resources. Although data augmentation is valuable to artificially increase a dataset's size, in the absence of sufficient data for CNN training, transfer learning is also effective. However, there are some studies that use base deep learning models in this regard. Chen et al.[30] proposed a six-layer CNN architecture including (1) four convolutional and (2) two fully connected layers for the automatic classification of images of fundus glaucomatous. Another study by Alghamdi et al.[31] used four public and four private datasets to diagnose visual disc abnormalities. They introduced a new approach by using two CNN networks: first, a CNN is trained and classifies the visual disc area, and another to categorize the visual disc area into normal, suspicious, and abnormal categories.

As mentioned above, transfer learning could be a useful approach to deal with small training data in glaucoma diagnosis. To exploit this strategy, one must obtain a fixed feature extractor by considering the CNN convolution layers as an extractor and then transfer these features to a linear classifier for training.[32] In Reference 5, the analysis was presented by five different CNN-trained architectures with ImageNet, which is used as a glaucoma classifier. They were precisely tuned and tested using unique public datasets. The high accuracy and sensitivity of their analysis showed that the CNN architecture trained by ImageNet was a powerful option for the automatic glaucoma detection algorithm. Carneiro et al.,[33] have shown that CNN models already trained in natural images, such as ImageNet, are very practical in medical image applications although images' appearance is significantly different. In another study by Tajbakhsh et al.,[34] the researchers demonstrated the performance of CNN already trained for four medical imaging programs. CNN-trained networks with ImageNet can extract features so that where a CNN applies to an image, a specific hidden layer extracts the features. Finally, the extracted features are classified using SVM, decision tree, k-nearest neighbor (KNN), or Naive Bayes classification. For example, Barr et al.,[35] used pre-trained CNN networks as feature extractors to identify chest pathology.

In Reference 36, is introduced Rethinking the Inception Architecture for Computer Vision. This work proposed improvements to the original Inception module design by introducing several modifications such as dimension reductions with 3×3 convolutions, factorizing larger convolutions into smaller ones, and utilizing label smoothing regularization.

Christian Szegedy et al,[37] introduced a modified version of an Inception module called a 'grid reduction unit' to improve efficiency while maintaining accuracy. It explored different variants of convolutional layers and demonstrated improved performance on image recognition tasks.

Medical image analysis has been widely used to diagnose diseases such as breast cancer, prostate cancer, and so forth. Recently, due to great success of artificial intelligence techniques, deep learning analysis for medical images has been widely studied. However, obtaining sufficient medical information is difficult, and the small amount of training data prevents the improvement of deep learning performance. To solve these issues, various studies have been conducted and data enhancement is one of the mitigation strategies. For example, Ronneberger et al.,[48] added modification, rotation, and deformation in microscopic images to the model during grid training.

Utilization of data augmentation could be different from one application to another. For example, in learning to recognize human speech, a little noise is added to the voice signal so that the network can learn to recognize words or emotions even with opaque sounds. In image processing applications, techniques such as rotating and changing the brightness and cropping the image are used to detect the presence of an object in the image or to find the exact location of several objects in an image. In image data augmentation, the modifications on images are applied in way so that the statistical distributions and main properties of the images are preserved. The modified versions of images are generated by image processing operations such as shifts, reversals, zooms. The new images should be probable and believable. For example, the rotating 180° is not suitable for producing new images of dogs or cats because dogs and cats do not move on the ceiling or in the air. Therefore, the data augmentation techniques should be carefully selected according to the content and type of data.

Recent deep learning methods such as convolution neural networks are capable of learning features which are independent of the location of the object of interest in the image. However, enhancing and adding data by changing rotation, light, and color can further aid in the learning of independent features in the image. Image data augmentation is usually applied to training images and not to test images.[49]

Fundus imaging is an effective and inexpensive tool for screening of retinal diseases. At the same time, deep learning algorithms can perform similar to or even better than human experts (e.g., physicians) in some image classifications. One key tool is the use of data enhancement techniques to improve the performance of such models. Commonly used augmentation methods for eye fundus images include rotation, shearing, and flipping.

3 | PROPOSED METHOD

In this section, we describe different steps of our approach for processing glaucoma images. The main part of this model is a deep convolutional neural network for classifying input glaucoma images.

The CNN consists of several layers. The first layer is a Conv2D layer with 256 filters, each with a kernel size of (3, 3), and ReLU activation function. It takes input images of size (256, 256, 3). This layer applies convolutional filters to extract features from the input images. The second layer is Batch Normalization, which normalizes the outputs of the previous layer to ensure stable training. Third is the MaxPooling2D layer with a pool size of (2, 2), which reduces the spatial dimensions of the previous layer's output by selecting the maximum value in each pooling window. This helps in capturing the most important features and reducing the computational complexity. The fourth layer is a Dropout layer with a dropout rate of 0.4, which randomly sets a fraction of input units to 0 at each update during training, helping to prevent over-fitting. The next few layers follow a similar pattern as the previous layers, with Conv2D, Batch Normalization, MaxPooling2D, and Dropout layers applied successively.

The proposed network is a convolutional neural network (CNN) with three MAS Blocks that classify images into predefined classes. Each MAS block consists of three filters: 1×1, 3×3, and 5×5, which are located in the first, second, and third layers, respectively.

3.1 | Data augmentation techniques for fundus images

Classification of medical images plays a key role in patients' treatment and diagnosis. This process can be performed through either manual inspection or using artificial intelligence. In the former, a specialist visually inspects images which is a highly error-prone and time-consuming process. On the other hand, extracting and selecting suitable image features using classical machine learning techniques is very challenging. Deep neural network is currently an emerging machine learning method that has shown its potential for various classification tasks. In addition, it has produced promising results in the classification of medical images.

Furthermore, image data augmentation is a technique that artificially increases the size of the data without new imaging and only by creating manipulated versions of existing images.[50] Training deep neural networks with large-scale data can lead to more accurate networks. Hence, image data augmentation process can increase the size and diversity of the input images to allow learning detailed properties by the CNN and achieve better understanding of each image class leading to a so-called generalized model. In this study, we tried using the GAN (Generative Adversarial Network) method for image generation. But, due to the noisy nature of the generated images, the proposed model did not train well.

3.2 | Using MAS block to extract features of fundus images

In the following, we introduce the Multi-Scale Attention Block (MAS Block), a pivotal component that offers significant advantages in the domain of fundus image feature extraction.[51] By incorporating convolutions with different kernel sizes, an MAS Block can gather contextual information from neighboring pixels in a fundus image. This is particularly useful for understanding relationships between objects or regions within the image and capturing global context that aids in distinguishing normal from abnormal areas. Fundus images typically have high-resolution input dimensions that may lead to increased computational complexity and memory requirements in deep learning models.

The MAS Block is a building block used in deep learning models, particularly in CNNs. It was introduced by the Google Research team in their paper "Going Deeper with Convolutions" for image classification tasks.[37] The MAS Block aims to capture multi-scale features within an image by applying multiple different-sized filters simultaneously and concatenating their outputs. This allows the network to learn both local and global patterns effectively.

An MAS Block typically consists of a series of parallel branches, each performing different convolutions on the input feature maps. These convolutions can include 1×1, 3×3, or 5×5 filters as well as pooling operations. By combining these different filter sizes together, the model can extract features at various spatial resolutions.

Overall, the MAS Block has been widely adopted in many state-of-the-art CNN architectures due to its ability to capture diverse scale information efficiently, enabling better performance on image recognition tasks.

In addition to its role in capturing contextual information, the MAS Block operates as a complicated mechanism for adaptive feature extraction. The incorporation of convolutions with different kernel sizes enables the MAS Block to dynamically adjust its receptive field, facilitating the extraction of features at multiple scales within a fundus image. This adaptability is particularly crucial in the analysis of complex structures and subtle variations present in medical images, allowing the model to recognize detailed patterns, aiding in precise diagnosis. As a fundamental building block in deep learning models, the MAS Block not only enhances feature extraction but also plays a key role in optimizing the model's ability to generalize across diverse fundus images, contributing to its robust performance in real-world scenarios.

The multi-scale attention block allows the model to capture both fine-grained details and broader contextual information simultaneously. By attending to relevant regions at different scales, the model can effectively handle variations in object sizes, complex scenes, and context-dependent classification scenarios. This network provides a new structure for the diagnosis of glaucoma. As explained in the introduction, previous research on the diagnosis of the disease has been based on pre-trained networks or the calculation of the CDR index.

A multi-scale attention block for image classification is a building block or module that combines the concepts of multi-scale processing and attention mechanisms in CNNs. It aims to improve the performance of image classification models by capturing contextual information at different scales or resolutions. The multi-scale attention block consists of the following components:

I. Multi-Scale Feature Extraction: The input image is processed by a set of convolutional layers to extract features at different scales or levels. These convolutional layers capture information at various receptive field sizes, allowing the model to capture both local details and global context.
II. Attention Mechanism: Within each scale, an attention mechanism is applied to compute the importance or relevance of different spatial positions or regions. This is typically achieved through self-attention, where each position is compared to all other positions within the same scale, and attention weights are assigned accordingly. The attention weights reflect the significance of each position in relation to others within the same scale.
III. Attention Fusion: The attention weights obtained from each scale are combined or fused to create a unified representation that incorporates contextual information from multiple scales. This fusion enables the model to have a comprehensive understanding of the image, considering both local and global relationships.
IV. Integration and Classification: The fused representation is then further processed by subsequent layers, such as fully connected layers or a classifier, to perform the final image classification task.

In addition to its primary function in integration and classification, the multi-scale attention block plays a key role in enhancing the model's robustness. The adaptability of the attention mechanism ensures that the model can effectively handle variations, making it resilient to noise, diverse image qualities, and potential variations in fundus images. This adaptability is paramount in real-world applications, where the conditions of capturing images may vary. The multi-scale attention block's capacity to capture both local and global context allows the model to generalize well across different datasets, ensuring consistent performance across a spectrum of glaucoma cases.

Because of high computational demand, most deep learning applications require high powered computing resources. As the number of layers in a deep neural network increases, the number of required calculations will significantly increase, and thus using GPUs would be necessary. As the number of layers of deep neural networks increases, more training data would be needed compared to classical artificial neural networks. The more the training data, the more accurate the learning for deep multilayer networks can be obtained. Therefore, increasing the amount of data alongside using GPUs improve the performance of deep neural networks.

There are several software libraries for designing deep learning structures. In this article, a CNN image classifier was created using Tensorflow. Figure 2 depicts the detailed architecture of our proposed model.

4 | EXPERIMENTAL RESULTS

In this study, we used an HP Proliant DL380 G10 and an NVIDIA Quadro RTX 5000 16 GB graphics card to implement and evaluate the proposed method. We also used Python version 3.9 and CUDA Toolkit 11.3 for the graphics card to run the proposed model. In the following section, the details of setting the parameters and results of each module are provided along with the performance criteria of the whole system.

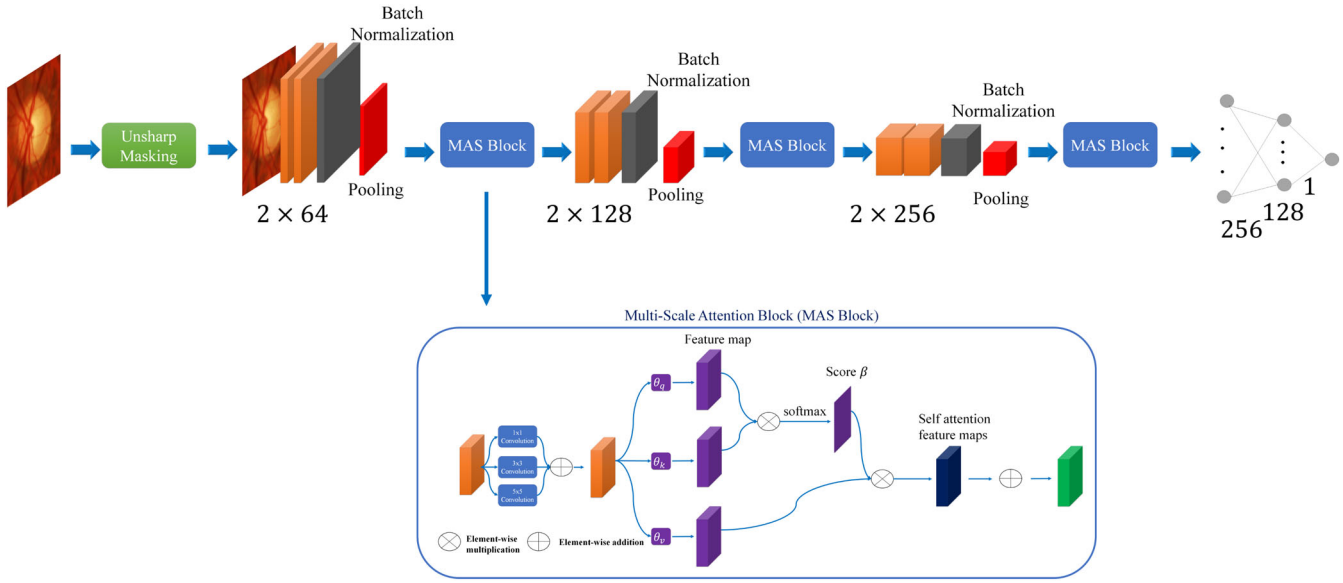

FIGURE 2 Proposed convolutional neural network architecture.

TABLE 2 Number of images in public datasets labeled with glaucoma.

Dataset	Glaucoma	Normal	Total
HRF[52]	27	18	45
Drishti-GS1[6]	70	31	101
sjchoi86[53]	101	300	401
RIM-ONE[54]	194	261	455
ACRIMA[5]	396	309	705

TABLE 3 Number of images of training ACRIMA, RIM-ONE and Sjchoi86 datasets after data augmentation.

Dataset	Glaucoma	Normal	Total
ACRIMA	2219	1729	3948
RIM-ONE	1085	1463	2548
sjchoi86	576	1680	2256

4.1 | Dataset

To train a CNN model, a large collection of retina images is required. Five public datasets are available for glaucoma. The specifications of these datasets are listed in Table 2.

ACRIMA[5] clinical dataset is the largest public dataset for glaucoma diagnosis. This dataset contains 705 labeled images, including 396 glaucoma and 309 normal images. Patients provided the images with prior consent and by the ethical principles and standards set out in the 1964 Helsinki Declaration. During the examination, all patients were selected by specialists based on clinical criteria and findings. Some images have been discarded due to noise and poor contrast. They were taken with a Topcon TRC retina camera and an IMAGE net imaging system with a 35° field of view. All ACRIMA dataset images are annotated by two glaucoma experts who have 8 years of experience. No other clinical information was considered when labeling the images.

To train a CNN network, a large collection of retina images is required. Therefore, according to Table 2, we have used ACRIMA, RIM-ONE, and Sjchoi86 separate datasets to train the model. We almost included 10% of the data in these datasets as a test set and examined the results of various experiments on this data set. Since the number of images in this dataset is not enough to train the model, we have used the data augmentation technique in this study to increase the number of images. Table 3 shows the number of glaucoma and non-glaucoma images of training data after performing

data augmentation on the ACRIMA, RIM-ONE, and Sjchoi86 datasets. As one of the most popular techniques to evaluate the classification performance, K-fold cross-validation was used with $K = 10$.

4.2 | Tuning hyperparameters

4.2.1 | Tuning optimization functions

To improve the accuracy and performance, we have used three optimizer functions in the proposed model. Two major performance metrics including loss function and are used to measure the performance of the model after each iteration According to Table 4, the highest accuracy has been obtained for the Adam optimizer.

The Loss function is equivalent to the mean square error (MSE) which can be calculated via Equation (1) where n is the number of data points, Y is the vector of observed, and $\hat{Y}$ is the predicted values:

$$\mathrm{MSE} = \frac{1}{n}\sum_{i=1}^{n}\left(Y_i - \hat{Y}_i\right)^2 \tag{1}$$

In other words, the MSE is the mean $(\frac{1}{n}\sum_{i=1}^{n})$ of the squares of the errors $\left(Y_i - \hat{Y}_i\right)^2$. This is an easily computable quantity for a particular sample.

4.2.2 | Tuning nonlinear functions

In the proposed model, we use three nonlinear functions Elu, ReLU, and Tanh. According to Table 5, the highest accuracy was obtained for the Elu function.

The proposed model includes three stages of preprocessing, training, and model evaluation. In the first step, initial preprocessing consists of normalization, and unsharp masking (USM) on all training, validation, and testing data for each image in the entire dataset. We used the unsharp masking technique to enhance the quality of fundus images and reduce the natural noise. Using this technique, some image features such as edges, boundaries, or contrast are improved. In the second stage, we trained the proposed neural network through 200 iterations. Finally, in the third step, which is model evaluation, we evaluate the trained model and checked the accuracy and loss obtained on the experimental data.

In this study, the proposed model was evaluated on the ACRIMA, RIM-ONE and Sjchoi86 datasets. By reviewing and comparing the results obtained from the model, the lack of data in the training phase results in over-fitting the data of the experimental set (Table 6). According to Table 6, the accuracy of the model in the training phase is very good, but it is low in the testing phase, and it indicates the over-fitting phenomenon.

TABLE 4 Comparison of optimizer functions in the proposed model.

Optimizer functions	Test accuracy (%)	Test loss
Adam	97.18	12.2
Adamx	87.32	26.4
SGD	78.87	40.2

TABLE 5 Comparison of nonlinear functions in the proposed model.

Activation function	Test accuracy (%)	Test loss
Elu	97.18	12.2
ReLU	90.14	22.9
Tanh	87.32	26.2

TABLE 6 Accuracy obtained in ACRIMA, RIM-ONE and Sjchoi86 datasets before data augmentation technique.

Dataset	Training accuracy (%)	Test accuracy (%)
ACRIMA	99.63	67.6
Sjchoi86	83.4	23.6
RIM-ONE	85.6	21.1

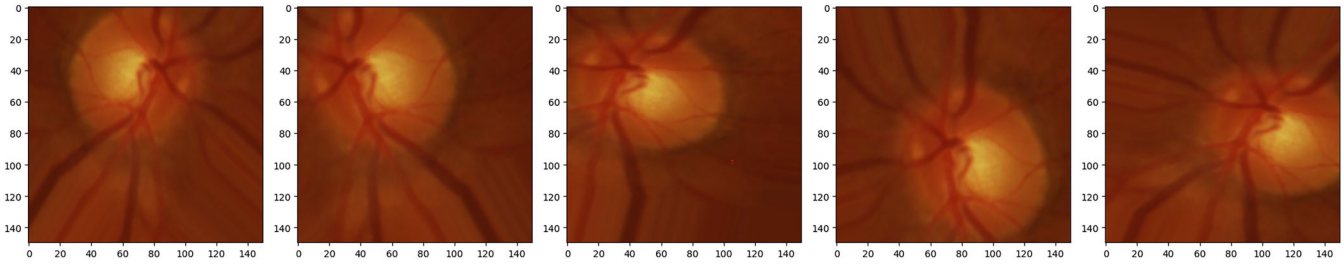

FIGURE 3 An example of data augmentation on a sample of the fundus Image.

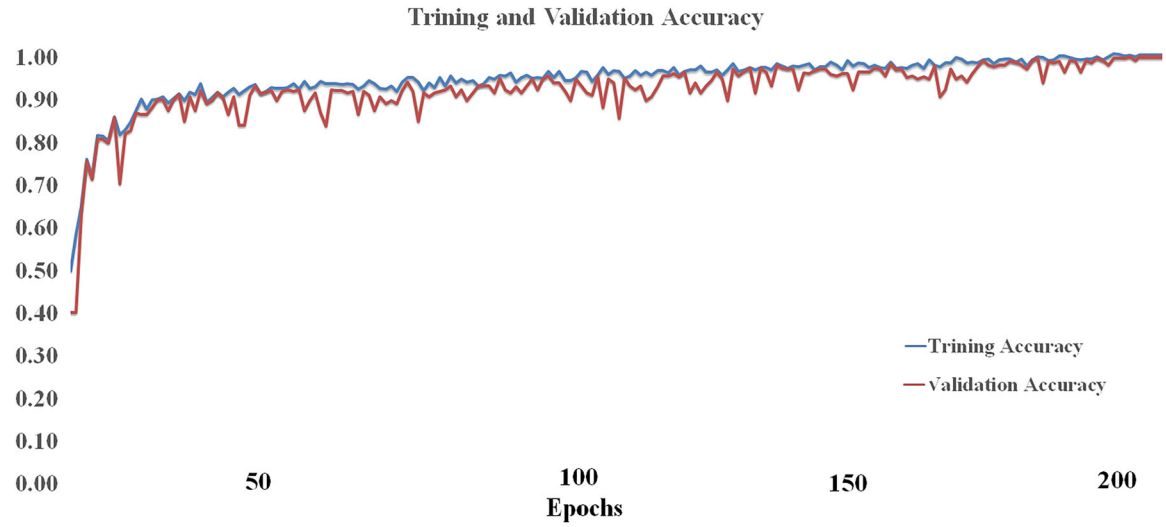

FIGURE 4 Accuracy in ACRIMA dataset.

Additionally, given the disparity in image quantities between the ACRIMA dataset and the RIM-ONE and Sjchoi86 datasets, Specifically, the images in the ACRIMA dataset were artificially augmented through the utilization of data augmentation techniques.

The data augmentation step is applied before preprocessing stage, i.e. before normalization and unsharp masking.

Data augmentation techniques including operations such as rotation range of 10°, zooming range of 0.2, shearing range of 0.2, horizontal flip, height shift range, and width shift range of 10% are shown in Figure 3. This figure depicts the influence of data augmentation on a subset of fundus images, potentially enhancing the generalization and resilience of a machine learning model. The visualization showcases the original fundus image alongside its diverse augmented versions, serving to exemplify the capacity of data augmentation in generating varied renditions of the original image. This process aims to furnish the model with a more extensive array of instances for learning purposes.

Note that the data augmentation is performed only for training data. Therefore, the validation and testing conditions are the same as in the previous experiment. Figure 4, shows the training and validation accuracy of a convolutional Neural Network (CNN) over time. The x-axis represents the number of training epochs, while the y-axis represents the accuracy of the model on the training and validation datasets. According to Figure 4, is observed how the accuracy of the CNN model changes as it is trained on the training data and then evaluated on the validation data. We aim to observe a

concurrent increase in both the training and validation accuracy over time, signifying the model's adeptness in learning and effectively generalizing to novel data. Conversely, if the training accuracy demonstrates persistent improvement while the validation accuracy reaches a plateau or diminishes, this could suggest potential over-fitting of the model to the training data, thus hindering its generalization capabilities.

4.3 | Effect of using MAS block

In this subsection, we delve into the impact of incorporating the Multi-Scale Attention (MAS) Block within our proposed framework. To thoroughly evaluate its impact, we compared the performance of our CNN model with and without the MAS Block. The results demonstrated a substantial increase in accuracy when utilizing the MAS Block, underlining its pivotal role in enhancing the model's predictive capabilities.

Our experiments demonstrated that the MAS Block effectively captures multi-scale features within fundus images, allowing the model to discern intricate patterns crucial for accurate diagnosis. By applying multiple filters simultaneously in a series of parallel branches, the MAS Block facilitates the extraction of features at various spatial resolutions. This adaptability is especially crucial in analyzing complex structures and subtle variations present in medical images. The MAS Block's ability to capture both local and global context contributes significantly to the model's improved performance in distinguishing normal from abnormal areas within fundus images. As shown in Table 7, utilizing the MAS Block resulted in an increase in accuracy by 3.29%.

Furthermore, the integration of the MAS Block into our framework showcases its efficiency in optimizing the model's ability to generalize across diverse fundus images. Despite the additional complexity introduced by the MAS Block, the increase in accuracy is substantial, demonstrating its robustness and utility in real-world scenarios. These findings underscore the MAS Block's instrumental role in advancing the accuracy and efficacy of our proposed model for glaucoma diagnosis.

4.4 | Performance metrics of the proposed model

Here, the overall performance of our model will be discussed. The confusion matrix consists of actual and predicted classification results. The performance of such systems is usually assessed using the data in this matrix. This matrix is used to extract some important criteria such as accuracy, precision, recall, F-measure score and MCC etc. Table 8 shows the Confusion Matrix for a two-class classifier.

To facilitate the evaluation of the performance of the glaucoma diagnosis system, three criteria for evaluating the Accuracy (Equation (2)), Precision (Equation (3)), Recall (Equation (4)), and F-measure score (Equation (5)) and MCC (Matthews Correlation Coefficient) (Equation (6)) can be used.[55]

The F-measure score is commonly used in imbalanced datasets because it provides a balanced evaluation of a classifier's performance, taking into account both precision and recall. In imbalanced datasets, where the number of instances in different classes is uneven, accuracy alone may not be a reliable measure of a model's performance. When a dataset is imbalanced, accuracy can be misleading because a classifier can achieve high accuracy by simply predicting the

TABLE 7 The effect of MAS block.

	Accuracy %	Precision %	Recall %	F-measure score %	MCC %
Baseline without MAS block	90.14	90.47	92.68	91.77	79
Baseline with MAS block	**97.18**	**100**	**95.12**	**97.06**	**94.43**

TABLE 8 Confusion matrix for classification of two classes.

	Predicted positive	Predicted negative
Actual positive	TP	FN
Actual negative	FP	TN

TABLE 9 Results of performance evaluation of our model and pre-trained models.

#	Model name	Accuracy %	Precision %	Recall %	F-Measure score %	MCC %
1	DensNet201	70.42	78.79	65.00	70.02	42.18
2	InceptionResNetV2	81.69	90.91	75.00	80.79	64.95
3	InceptionV3	63.38	62.07	90.00	75.99	24.40
4	MobileNet	73.24	78.38	72.50	74.80	46.35
5	Resnet101	84.51	100	72.50	81.71	73.15
6	Resnet152V2	74.65	82.35	70.00	74.58	50.28
7	VGG 16	84.51	96.77	75.00	82.62	71.77
8	VGG 19	83.10	96.67	72.50	80.78	69.55
9	Xception	63.38	93.75	37.50	49.73	40.68
10	ConvNeXtTiny	64.79	94.12	40.00	52.33	42.73
11	ConvNeXtBase	76.06	89.66	65.00	73.26	55.82
12	ConvNeXtLarg	76.06	87.10	67.50	74.36	54.59
13	**Our model**	**97.18**	**100**	**95.12**	**97.06**	**94.43**

majority class for most instances, while performing poorly on the minority class. This can lead to a false sense of the model's effectiveness. The F-measure score, which combines precision and recall into a single metric, is particularly useful in imbalanced datasets because it considers both false positives (precision) and false negatives (recall), providing a more comprehensive assessment of the classifier's ability to correctly identify instances from all classes.

In this study, we employed Equation (5) to compute the F-measure score. The parameter β regulates the emphasis on recall in relation to precision. A higher β value assigns greater weight to recall, whereas a lower β value assigns greater weight to precision. Our proposed model aims to reduce the occurrence of false positives (FP). To accomplish this, we have specified the β as 1.2 in Equation (5), thereby highlighting the significance of recall over precision.

The variables accuracy, precision, recall, F-measure score and MCC (Matthews Correlation Coefficient) are performance metrics commonly used in classification tasks to evaluate the quality of a machine learning model. The variables TP, TN, FP, and FN stand for true positives, true negatives, false positives, and false negatives, respectively.

The obtained variables are shown in Table 9 (#13) according to the data of the evaluation section of the proposed model for the data in Table 3.

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TN} + \text{FP} + \text{TP} + \text{FN}} \tag{2}$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{3}$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{4}$$

$$\text{F} - \text{Measure score} = \left(1 + \beta^2\right) \frac{\text{Precision} \times \text{Recall}}{\beta^2 \times (\text{Precision}) + \text{Recall}} \tag{5}$$

$$\text{MCC} = \frac{(\text{TP} \times \text{TN} - \text{FP} \times \text{FN})}{\sqrt{(\text{TP} + \text{FP})(\text{TP} + \text{FN})(\text{TN} + \text{FP})(\text{TN} + \text{FN})}} \tag{6}$$

5 | DISCUSSION

Artificial intelligence, such as machine learning and deep learning, has been widely used in various fields, including cancer treatment, biomedical information, industrial electronics, the Internet of Things, machine translation, and more.

In ophthalmology, artificial intelligence has shown great potential in diagnosing various eye diseases, including diabetic retinopathy, glaucoma, retinopathy, and OCT (optical coherence tomography). These algorithms have performed well in image classification and evaluation and have even been comparable to human clinicians. However, challenges remain in clinical and technical aspects, including algorithm interpretability by physicians and patients, and potential over-fitting due to small and imbalanced datasets. The proposed method has solved the problem of over-fitting to an extent and achieved promising results in diagnosing glaucoma using fundus images, but this limitation has not been fully resolved. Although the data will be increased by using the image augmentation technique to solve the over-fitting problem, there is a possibility that the model may make an error in the testing section and make an incorrect diagnosis, which is caused by not training the proposed model with a publicly available dataset with a large number of glaucoma data samples.

In this study, we addressed the diagnosis of glaucoma using a deep learning approach. First, we applied raw data to our model based on the obtained results. Because of the small amount of data in the training phase, the phenomenon of over-fitting occurred in the model. To reduce the effect of over-fitting, we presented the use of the data augmentation technique and adding MAS Block. And trained the proposed model after adding dataset images on the dataset images in Table 3.

The results of the performance of the proposed model with data augmentation and adding MAS Block were evaluated using measures like accuracy, precision, recall, F-measure score and MCC (Table 9). Figure 4 show the amount of accuracy measured in the model. Afterward, we compared the performance of the proposed model with 12 pre-trained networks. In order to evaluate and compare the performance of the proposed model with pre-trained networks, we used the same test dataset for both the proposed model and the pre-trained networks. The results of which are shown in Table 9. Based on the findings in Table 9, the performance of Resnet101 and VGG16 models is higher than in comparison with other pre-trained networks. While the performance of the proposed model is more than in comparison with these two models. It should be noted that the results of Table 9 are based on the implementation of the methods by the researchers of this article and testing on the same test set.

Extensive research has been done on the diagnosis of glaucoma using fundus images on images of the ACRIMA dataset. The results of different methods of diagnosing glaucoma are shown in Table 11. These results are based on reports presented in different articles, which can have different test data sets. As an example, the accuracy of InceptionV3 is three different values, which include of 93.87, 98.25, and 98.5. These results are also given in this article to make a fair comparison. Another important parameter for comparing different methods is the number of parameters of models. In Table 10, the number of parameters of pre-trained models.

As per the findings presented in Table 11, some methods such as DenseNet[56] have higher accuracy compared to the proposed method. However, these methods have more parameters compared to the proposed method (Table 10), which

TABLE 10 Number of parameters for pre-trained models.

Model name	# Parameters (in millions)
DenseNet201	20.2
InceptionResNetV2	59.9
InceptionV3	23.9
MobileNet	4.3
Resnet50	25.6
Resnet101	44.7
Resnet152V2	60.4
VGG 16	138.4
VGG 19	143.7
Xception	22.9
ConvNeXtTiny	28.6
ConvNeXtBase	88.5
ConvNeXtLarg	197.8
Our model	**3.4**

TABLE 11 Comparison of different results of glaucoma diagnosis in ACRIMA dataset.

Title	Network	Accuracy (%)
53	Xception	70.00
	ResNet-50	84.00
	GoogLeNet	87.00
	ResNet-152	83.00
5	Xception	96.05
47	Inception V3	98.25
	VGG19	92.64
	ResNet50	95.58
57	Alex Net	99.50
	InceptionV3	98.50
	InceptionResNetV2	99.00
	NasNet-Large	99.50
56	AlexNet	96.23
	GoogleNet	91.51
	InceptionV3	93.87
	XceptionNet	98.11
	ResNet50	95.75
	SqueezeNet	95.75
	ShuffleNet	96.23
	MobileNet	98.58
	DenseNet	99.53
	InceptionResNet	96.23
	NasNet-Large	96.23

requires more computation power and resources. For the DenseNet method, the number of parameters of model is about 6 times of the proposed method. It is also noteworthy that Sreng et al.[56] merely acknowledged the presence of a significant class imbalance in the ACRIMA dataset and did not take steps to address it during their research.

6 | CONCLUSIONS

Fundus images are currently used as a complement to other important clinical parameters in the diagnosis and progression of glaucoma. This paper demonstrates a process of classifying fundus images using CNNs that has a deep learning architecture. Although there are various algorithms that classify images, CNN is the standard method. One of the challenges on CNN is the need for a large number of images to train the model. In this article, we used of the data augmentation technique and MAS Block to reduce model over-fitting, improve image classification accuracy, and increase model accuracy by 97.18%.

The patient's clinical findings, and other ophthalmologic images such as OCT, Pentacam, Perimetry, and genetic testing are data of a different nature, each offering a different and potential perspective on the patient's condition. This type of data, along with a comprehensive glaucoma diagnosis application could be a new approach to enhance diagnostic accuracy. An accurate and early diagnosis of glaucoma is essential for early treatment before the patient begins to permanently lose vision. One of the limitations of this study is the lack of access to the OCT and Perimetry data of the patients. Today, with the advent of smartphones and the possibility of installing various applications, the authors of this study suggest the

development of a proposed model using OCT data And perimetry data of each patient on smartphones as a next step that can be made available to the public as a glaucoma screening program.

CONFLICT OF INTEREST STATEMENT

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

PEER REVIEW

The peer review history for this article is available at https://www.webofscience.com/api/gateway/wos/peer-review/10.1002/eng2.12866.

DATA AVAILABILITY STATEMENT

Data sharing is not applicable to this article as no datasets were generated during the current study.

ORCID

Mansoor Fateh https://orcid.org/0000-0003-2133-3480

REFERENCES

1. Ham Y-C, Li X, Wong TY, Quigley HA, Aung T, Cheng C-Y. Global prevalence of glaucoma and projections of glaucoma burden through 2040: a systematic review and meta-analysis. *Ophthalmology*. 2014;121(11):2081-2090.
2. Bourne RRA. Worldwide glaucoma through the looking glass. *Br J Ophthalmol*. 2006;90(3):253-254.
3. Selvathi D, Prakash NB, Gomathi V, Hemalakshmi GR. Fundus image classification using wavelet based features in detection of glaucoma. *Biomed Pharmacol J*. 2018;11(2):795-805.
4. Soanes M, Essa K, Butt H. Testing the viability of measuring intraocular pressure using soundwaves from a smartphone. *Eng Rep*. 2021;3(7):12355-12369.
5. Diaz-Pinto A, Morales S, Naranjo V, Köhler T, Mossi JM, Navea A. CNNs for automatic glaucoma assessment using fundus images: an extensive validation. *Biomed Eng Online*. 2019;18:1-19.
6. Sivaswamy J, Krishnadas SR, Joshi GD, Jain M, Ujjwaft Syed Tabish A. Drishti-gs: retinal image dataset for optic nerve head (onh) segmentation. In: *2014 IEEE 11th International Symposium on Biomedical Imaging (ISBI)*. IEEE; 2014:53-56.
7. Morales S, Naranjo V, Angulo J, Alcañiz M. Automatic detection of optic disc based on PCA and mathematical morphology. *IEEE Trans Med Imaging*. 2013;32(4):786-796.
8. Wong DWK, Liu J, Lim JH, et al. Level-set based automatic cup-to-disc ratio determination using retinal fundus images in ARGALI. *2008 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. IEEE; 2008:2266-2269.
9. Goodfellow I, Bengio Y, Courville A. *Deep Learning (Adaptive Computation and Machine Learning Series)*. MIT press; 2016:321-359.
10. Szegedy C, Ioffe S, Vanhoucke V, Alemi A. Inception-v4, inception-resnet and the impact of residual connections on learning. In: *Proceedings of the AAAI Conference on Artificial Intelligence*. 2017;31(1).
11. Zhang Y, Zhang K, Ding Y, et al. Deep transfer learning from ordinary to capsule esophagogastroduodenoscopy for image quality controlling. *Eng Rep*. 2023:e12776.
12. Zhu X, Rangayyan RM. Detection of the optic disc in images of the retina using the Hough transform. In: *2008 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. IEEE; 2008:3546-3549.
13. Yin F, Liu J, Ong SH, et al. Model-based optic nerve head segmentation on retinal fundus images. In *2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. IEEE; 2011:2626-2629.
14. Aquino A, Gegúndez-Arias ME, Marín D. Detecting the optic disc boundary in digital fundus images using morphological, edge detection, and feature extraction techniques. *IEEE Trans Med Imaging*. 2010;29(11):1860-1869.
15. Joshi GD, Sivaswamy J, Krishnadas SR. Optic disk and cup segmentation from monocular color retinal images for glaucoma19 assessment. *IEEE Trans Med Imaging*. 2011;30(6):1192-1205.
16. Abramoff MD, Alward WLM, Greenlee EC, et al. Automated segmentation of the optic disc from stereo color photographs using physiologically plausible features. *Invest Ophthalmol Vis Sci*. 2007;48(4):1665-1673.
17. Cheng J, Liu J, Yanwu X, et al. Superpixel classification based optic disc and optic cup segmentation for glaucoma screening. *IEEE Trans Med Imaging*. 2013;32(6):1019-1032.
18. Cheng J, Tao D, Wong DWK, Liu J. Quadratic divergence regularized SVM for optic disc segmentation. *Biomed Opt Express*. 2017;8(5):2687-2696.
19. Brezinski ME, Tearney GJ, Bouma BE, et al. Imaging of coronary artery microstructure (in vitro) with optical coherence tomography. *Am J Cardiol*. 1996;77(1):92-93.
20. Joshi GD, Sivaswamy J, Karan K, Krishnadas SR. Optic disk and cup boundary detection using regional information. In *2010 IEEE International Symposium on Biomedical Imaging: from Nano to Macro*, IEEE; 2010:948-951.
21. Wong DWK, Liu J, Lim JH, Li H, Wong TY. Automated detection of kinks from blood vessels for optic cup segmentation in retinal images. *Medical Imaging 2009: Computer-Aided Diagnosis*. SPIE; 2009;7260:459-466.

22. Xu Y, Duan L, Lin S, et al. Optic cup segmentation for glaucoma detection using low-rank superpixel representation. In: *Medical Image Computing and Computer-Assisted Intervention–MICCAI* 2014*: 17th International Conference, Boston, MA, USA, September 14–18, 2014, Proceedings, Part I 17*, Springer International Publishing; 2014:788-795.
23. Rostami M, Muhammad U, Forouzandeh S, Berahmand K, Farrahi V, Oussalah M. An effective explainable food recommendation using deep image clustering and community detection. *Intell Syst Appl*. 2022;16:200157-200170.
24. Chen J, Chen D, Liu G. Using temporal convolution network for remaining useful lifetime prediction. *Eng Rep*. 2021;3(3):12305-12321.
25. Bock R, Meier J, Nyúl LG, Hornegger J, Michelson G. Glaucoma risk index: automated glaucoma detection from color fundus images. *Med Image Anal*. 2010;14(3):471-481.
26. Russakovsky O, Deng J, Hao S, et al. Imagenet large scale visual recognition challenge. *Int J Comput Vis*. 2015;115:211-252.
27. Srivastava N, Hinton G, Krizhevsky A, Sutskever I, Salakhutdinov R. Dropout: a simple way to prevent neural networks from overfitting. *J Mach Learn Res*. 2014;15(1):1929-1958.
28. Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. *arXiv Preprint* arXiv:1409.1556. 2014.
29. Rostami M, Oussalah M, Farrahi V. A novel time-aware food recommender-system based on deep learning and graph clustering. *IEEE Access*. 2022;10:52508-52524.
30. Chen X, Yanwu X, Wong DWK, Wong TY, Liu J. Glaucoma detection based on deep convolutional neural network. In *2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, IEEE; 2015:715-718.
31. Alghamdi HS, Tang HL, Waheeb SA, Peto T. Automatic optic disc abnormality detection in fundus images: a deep learning approach. In: *Proceedings of the Ophthalmic Medical Image Analysis International Workshop*, vol. 3, no. 2016. University of Iowa, 2016.
32. Yosinski J, Clune J, Bengio Y, Lipson H. How transferable are features in deep neural networks? *Adv Neural Inf Process Syst*. 2014;27.
33. Carneiro G, Nascimento J, Bradley AP. Unregistered multiview mammogram analysis with pre-trained deep learning models. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*, Cham: Springer International Publishing; 2015:652-660.
34. Tajbakhsh N, Shin JY, Gurudu SR, et al. Convolutional neural networks for medical image analysis: full training or fine tuning? *IEEE Trans Med Imaging*. 2016;35(5):1299-1312.
35. Bar Y, Diamant I, Wolf L, Greenspan H. Deep learning with non-medical training used for chest pathology identification. *Medical Imaging 2015: Computer-Aided Diagnosis*. Vol 9414. SPIE; 2015:215-221.
36. Szegedy C, Vanhoucke V, Ioffe S, Shlens J, Wojna Z. Rethinking the inception architecture for computer vision. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*; 2016:2818-2826.
37. Szegedy C, Liu W, Jia Y, et al. Going deeper with convolutions. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2015;1-9.
38. Muramatsu C, Nakagawa T, Sawada A, et al. Determination of cup-to-disc ratio of optical nerve head for diagnosis of glaucoma on stereo retinal fundus image pairs. *Medical Imaging 2009: Computer-Aided Diagnosis*. SPIE; 2009;7260:1028-1035.
39. Hatanaka Y, Noudo A, Muramatsu C, et al. Automatic measurement of cup to disc ratio based on line profile analysis in retinal images. In: *2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. IEEE; 2011:3387-3390.
40. Mookiah MR, Krishnan UR, Acharya CM, Lim AP, Suri JS. Data mining technique for automated diagnosis of glaucoma using higher order spectra and wavelet energy features. *Knowl-Based Syst*. 2012;33:73-82.
41. Pathan S, Kumar P, Pai RM, Bhandary SV. Automated segmentation and classifcation of retinal features for glaucoma diagnosis. *Biomed Signal Process Control*. 2021;63:102244.
42. Elangovan P, Nath MK. En-ConvNet: A novel approach for glaucoma detection from color fundus images using ensemble of deep convolutional neural networks. *Int J Imaging Syst Technol*. 2022;32(6):2034-2048.
43. Elangovan P, Nath MK. Glaucoma assessment from color fundus images using convolutional neural network. *Int J Imaging Syst Technol*. 2021;31(2):955-971.
44. Nath MK, Dandapat S. Differential entropy in wavelet sub-band for assessment of glaucoma. *Int J Imaging Syst Technol*. 2012;22(3):161-165.
45. Elangovan P, Nath MK, Mishra M. Statistical parameters for glaucoma detection from color fundus images. *Procedia Comput Sci*. 2020;171:2675-2683.
46. Madhusudhan M, Malay N, Nirmala SR, Samerendra D. Image processing techniques for glaucoma detection. In: *Advances in Computing and Communications: First International Conference, ACC* 2011*, Kochi, India*, July 22–24, 2011, Proceedings, Part III 1, pp. 365–373. Springer Berlin Heidelberg; 2011.
47. Shyamalee T, Meedeniya D. CNN based fundus images classification for glaucoma identification. In: *2022 2nd International Conference on Advanced Research in Computing (ICARC)*, IEEE; 2022:200-205.
48. Ronneberger O, Fischer P, Brox T. U-net: convolutional networks for biomedical image segmentation. In: *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5–9, 2015, Proceedings, Part III 18*, pp. 234–241. Springer International Publishing, 2015.
49. El-Amir H, Hamdy M, El-Amir H, Hamdy M. Data wrangling and preprocessing. *Deep Learning Pipeline: Building a Deep Learning Model with TensorFlow*. Apress; 2020:147-206.
50. Jiang K, Xiong Z, Yang Q, Chen J, Chen G. An interpretable ensemble method for deep representation learning. *Eng Rep*. 2023;e12725.
51. Yang S, Xing L, Li Y, Chang Z. Implicit sentiment analysis based on graph attention neural network. *Eng Rep*. 2022;4(1):12452-12467.
52. Budai A, Bock R, Maier A, Hornegger J, Michelson G. Robust vessel segmentation in fundus images. *Int J Biomed Imaging*. 2013;2013:154860-154870.

53. Serte S, Serener A. A generalized deep learning model for glaucoma detection. In: *2019 3rd International Symposium on Multidisciplinary Studies and Innovative Technologies (ISMSIT)*, IEEE; 2019:1-5.
54. Fumero F, Alayón S, Sanchez JL, Sigut J, Gonzalez-Hernandez M. RIM-ONE: an open retinal image database for optic nerve evaluation. In: *2011 24th International Symposium on Computer-Based Medical Systems (CBMS)*, IEEE; 2011:1-6.
55. Anbalagan T, Nath MK, Vijayalakshmi D, Anbalagan A. Analysis of various techniques for ECG signal in healthcare, past, present, and future. *Biomed Eng Adv*. 2023;6:100089-100116.
56. Sreng S, Maneerat N, Hamamoto K, Win KY. Deep learning for optic disc segmentation and glaucoma diagnosis on retinal images. *Appl Sci*. 2020;10(14):4916-4934.
57. Taj IA, Sajid M, Karimov KS. An ensemble framework based on deep CNNs architecture for glaucoma classification using fundus photography. *Math Biosci Eng*. 2021;18(5):5321-5347.

How to cite this article: Khajeha HR, Fateh M, Abolghasemi V. Diagnosis of glaucoma using multi-scale attention block in convolution neural network and data augmentation techniques. *Engineering Reports*. 2024;e12866. doi: 10.1002/eng2.12866