Mathematical Biosciences and Engineering

http://www.aimspress.com/journal/MBE

MBE, 18(5): 5321–5346.
DOI: 10.3934/mbe.2021270
Received: 10 March 2021
Accepted: 02 June 2021
Published: 16 June 2021

Research article

An ensemble framework based on Deep CNNs architecture for glaucoma classification using fundus photography

Aziz-ur-Rehman[1,*], Imtiaz A.Taj[2], Muhammad Sajid[3] and Khasan S. Karimov[1,4]

[1] Faculty of Electrical Engineering, GIK Institute of Engineering Sciences and Technology, Topi 23640, District Swabi, KPK, Pakistan

[2] Department of Electrical Engineering, Capital University of Science and Technology Islamabad Expressway, Kahuta Road, Zone-V Islamabad, Pakistan

[3] Department of Electrical Engineering, Mirpur University of Science and Technology (MUST), Mirpur 10250 (AJK), Pakistan

[4] Centre for Innovative and New Technologies of Academy of Sciences of the Republic of Tajikistan, 734015, Rudaki Ave., 33. Dushanbe Tajikistan

* **Correspondence:** Email: azizur80@hotmail.com; Tel: +923338109594.

Abstract: Glaucoma is a chronic ocular degenerative disease that can cause blindness if left untreated in its early stages. Deep Convolutional Neural Networks (Deep CNNs) and its variants have provided superior performance in glaucoma classification, segmentation, and detection. In this paper, we propose a two-staged glaucoma classification scheme based on Deep CNN architectures. In stage one, four different ImageNet pre-trained Deep CNN architectures, i.e., AlexNet, InceptionV3, InceptionResNetV2, and NasNet-Large are used and it is observed that NasNet-Large architecture provides superior performance in terms of sensitivity (99.1%), specificity (99.4%), accuracy (99.3%), and area under the receiver operating characteristic curve (97.8%) metrics. A detailed performance comparison is also presented among these on public datasets, i.e., ACRIMA, ORIGA-Light, and RIM-ONE as well as locally available datasets, i.e., AFIO, and HMC. In the second stage, we propose an ensemble classifier with two novel ensembling techniques, i.e., accuracy based weighted voting, and accuracy/score based weighted averaging to further improve the glaucoma classification results. It is shown that ensemble with accuracy/score based scheme improves the accuracy (99.5%) for diverse databases. As an outcome of this study, it is presented that the NasNet-Large architecture is a feasible option in terms of its performance as a single classifier while ensemble classifier further improves the generalized performance for automatic glaucoma classification.

Keywords: Deep Convolutional Neural Network; transfer learning; fundoscopy; Optic Nerve Head; performance metrics; ensemble; accuracy based weighted voting and averaging

1. Introduction

According to the World Health Organization (WHO), an increasing number of glaucoma patients have been reported worldwide in recent years [1]. Glaucoma prevention and treatment has been a major focus of international directives including the WHOs Vision 2020 campaign. It is estimated that the number of people with glaucoma is expected to rise from 64 million to 76 million in 2020 and 111.8 million in 2040, with Africa and Asia being affected more heavily than the rest of the world as there is a shortage of trained ophthalmologist for its diagnosis [2, 3]. Hence, this disease is considered as a major public health concern, and its early diagnosis is important for preventing blindness.

Glaucoma is the second most common cause of irreversible vision loss and its diagnosis is a challenging research field because real world glaucoma images are acquired in the presence of several factors such as, illumination changes, background interference, light variation, etc [4]. With the rapid advancement in imaging technologies, several retinal imaging modalities like Heidelberg Retina Tomography (HRT) and Optical Coherence Tomography (OCT) have been developed. Although, these are used in developed and under-developed countries for image based diagnosis of various ocular diseases, i.e., Macular degeneration, Diabetic retinopathy and Glaucoma. However, these are costlier technologies and are not affordable for smaller public health units. Therefore, the most popular and widely used imaging device is the Fundoscopy [5].

Fundoscopy enables ophthalmologist to examine the Optic Nerve Head (ONH) for the glaucoma diagnosis and a typical image of ONH is shown in Figure 1. There are various parts of ONH which can be considered to classify between normal and glaucoma eyes. Four main changes can be observed in ONH associated with glaucoma, including ONH cupping, Neuroretinal rim thinning, Nerve Fibre Layer (NFL) thickness and Parapapillary Atrophy (PPA). These changes are detected manually through analysis of ONH images to diagnose glaucoma [6]. However, identification of glaucomatous signs in ONH require specialist ophthalmologists with years of experience and practice. Therefore, the development of automatic glaucoma assessment algorithms based on fundus image analysis, will be very helpful in reducing overall workload of ophthalmologists and also make the diagnosis more feasible and efficient even in smaller health units.

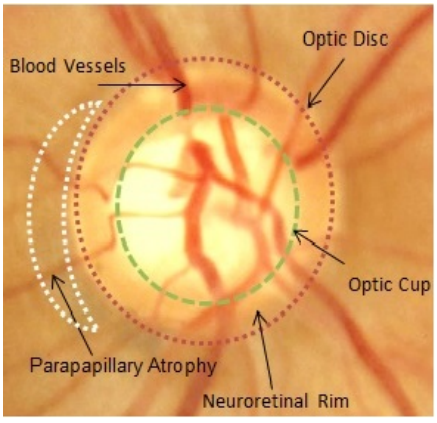


Figure 1. The most prominent area, Optic Nerve Head to diagnose the glaucoma initially.

Recently, Deep CNNs have attracted a lot of interest and are seen to have great potential for the

solution of computer vision tasks like image classification and semantic segmentation [7]. These networks are more robust to discover meaningful features in the images that are usually ignored in conventional image processing techniques. Moreover, different intermediate steps such as feature extraction and selection are embedded within the networks and these networks can perform feature learning and classification, simultaneously. Due to their popularity, numerous algorithms have been developed during the past few years on the detection and classification for glaucomatous fundus images. However, to train deep CNNs a large amount of annotated data is required [8, 9]. Therefore, a lot of research effort is being put on the training methodology of the networks. We can divide these methodologies into two categories, i.e., training a deep network from scratch (full training) and transfer learning with fine-tuning.

Training from scratch requires large amount of labeled data which is extremely difficult to find in medical imaging. It is expensive to collect images both in terms of time and budget for disease identification. Besides, the training of Deep CNNs is time consuming that usually requires extensive memory and computational resources. Furthermore, the designing and adjustment of the hyper-parameters are the challenging tasks with reference to over-fitting and other issues. On the contrary, transfer learning with fine-tuning [10–12] is the easiest way to overcome such problems. The transfer learning is commonly defined as the ability of the network to exploit the knowledge learned in one domain, to another that share some common characteristics. Therefore, it is more popular method in machine learning and data mining for regression, clustering, and classification problems.

In an ensemble framework, various classifiers are trained to solve the same problem. The classification capability of an individual optimized Deep CNN is not generalized for diverse databases. Generally, the classification performance of grouping of various Deep CNNs is better than single architecture [13]. To make the most of the single classification capability, a promising solution would be to create an ensemble of several Deep CNN models.

1.1. Background

During the past few years, extensive research has been carried out for the development of automatic glaucoma assessment systems based on fundus image analysis through transfer learning [14–16]. In the ImageNet Large Scale Visual Recognition Challenge (ILSVRC), millions of labeled images selected from 1000 different classes have been successfully tested on different medical image analysis [17]. It is tested that CNNs outperform the previous implementations with computer aided screening systems for medical images. Two individual CNN architectures are used in [18] to segment the optic disc and optic cup to find the cup-to-disc ratio (CDR). In [19], the authors have developed a CNN architecture to automate the detection process of glaucoma. Another Deep CNN algorithm has been developed in [20] to detect glaucoma through extraction of different features with the combination of different classifiers, i.e., Random Forest, Support Vector Machine and Neural Network. Feature learning through deep learning algorithm from retinal fundus images has also proposed in [21] to detect glaucoma. The authors have considered CNN model to learn features with linear and nonlinear activation functions. In the study [22], authors have implemented a Glaucoma-Deep system. The deep-belief network has been used to select the most discriminative features. In the work [23], two different types of CNNs, i.e., Overfeat and VGG-S, have been used as feature extractors. Contrast-Limited Adaptive Histogram Equalization (CLAHE) and vessels deletion

have considered to investigate the performance of these networks. The authors have proposed a joint segmentation of optic disc, optic cup and glaucoma prediction in [24]. CNN feature sharing for different tasks ensured better learning and over-fitting prevention. Inception-v3 architecture has presented in [25] to detect glaucomatous optic neuropathy. Local space average color subtraction has been applied in pre-processing to accommodate for varying illumination. A framework on a dataset of fundus images collecting from different hospitals has presented in [26] by incorporating both domain knowledge and feature learned from a deep learning model. The assessment of deep learning algorithms with transfer learning have also been addressed in [27, 28], with greater number of images than previous methods with high accuracy, sensitivity and specificity. Recently, in [29], authors have implemented deep learning based segmentation to identify glaucomatous optic disc. In [30], the authors have combined CNN and Recurrent Neural Network (RNN) to extract the spatial features in a fundus image and also the temporal features embedded in a fundus video,i.e., sequential images. ResNet50 deep CNN model has been explored with 48 full convolutional neural network layers. A deep learning approach based on deep residual neural network (ResNet101) for automated glaucoma detection using fundus images is proposed in [31]. Chronic eye disease diagnosis using ensemble-based classifier has presented in [32]. This study tends to achieve an early and accurate diagnosis of glaucoma based on an ensemble classifier by integrating the principal component analysis with rotation forest tree. Ensemble learning based CNN has been proposed in [33] to segment retinal images. The output of the classifier is subject to an unsupervised graph cut algorithm followed by a convex hull transformation to obtain the final optic cup and disc segmentation. A novel disc-aware ensemble network based on the application of different CNNs is presented in [34] for automatic glaucoma screening. The authors have introduced a deep learning technique to gain additional image-relevant information and screen glaucoma from the fundus image directly. More recently, a multi-class multi-label ophthalmological disease detection using transfer learning has been investigated in [35] with four pre-trained Deep CNN architectures. ResNet, InceptionV3, MobileNet, and VGG16 are implemented to detect eight types of ocular diseases. Besides, in [36], an artificial intelligence and telemedicine based screening tool has been developed to identify glaucoma suspects from color fundus Images. An ensemble of five pre-trained Deep CNN architectures is presented to detect glaucoma on the basis of cup to disc ratio (CDR). Two Xception architectures, InceptionResNetV2, NasNet, and InceptionV3 are used in the proposed ensemble to calculate the final CDR value for glaucoma detection.

It is evident from the above literature that almost all previously proposed methods employ CNN for the detection of glaucoma. However, limited work has been carried out to ensemble Deep CNNs for glaucoma diagnosis.

1.2. Paper contribution

The research contributions of presented work are summarized as follows:

1) We propose a new two-staged scheme based on Deep CNNs architectures for glaucoma classification using fundus images. This new scheme comprises four Deep architectures, i.e., AlexNet, InceptionV3, InceptionResNetV2 and NasNet-Large obtained through extensive experimental results search. We also assess their individual performance on both publicly and local hospital datasets. Our results clearly demonstrate the effectiveness of the newly proposed scheme based on Deep CNN architecture.

2) We propose a new ensemble framework for better glaucoma classification. Four different pre-trained Deep CNN models are fused together in parallel and the output scores / probabilities are optimized using five different voting techniques to find the best Deep CNNs combination. We also propose two novel voting techniques to achieve much better results as compared to existing state-of-the-art for glaucoma classification.

The rest of the paper is organized as follows: the proposed methodology is presented in Section 2. Experiments and results are given in Section 3. Results related discussion is presented in Section 4, while conclusion and future work are summarized in Section 5.

2. Proposed methodology

The proposed automatic glaucoma diagnosis using fundus images is shown in Figure 2. There are five steps: 1) data collection; 2) pre-processing; 3) Deep CNN feature learning; 4) Ensemble (4x Deep CNNs); 5) classification/diagnosis. These steps are explained in the following subsections:

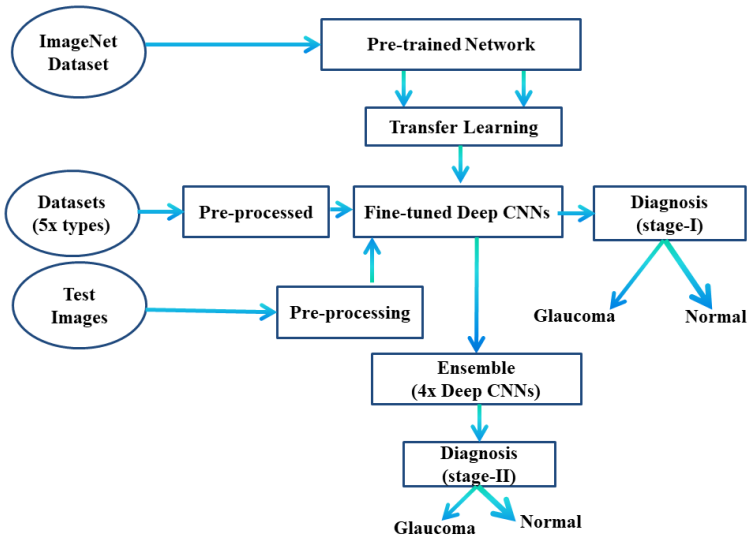


Figure 2. The proposed ensemble classifier based on Deep CNNs architecture for glaucoma diagnosis.

2.1. Data collection

Both public and private datasets are used to train, validate and test the different Deep CNNs. In this way, we can create the diversity in the images and the results will be generalized. The description about each dataset is given as follows:

ACRIMA is the most newly available public dataset for the classification of glaucoma through Deep learning. It consists of 705 fundus images (309 normal and 396 glaucomatous). All images are annotated by two glaucoma experts with eight years of experience. It is only be used for classification tasks because optic disc and optic cup annotations are not provided [37].

The ORIGA-light is another public dataset which is annotated by experienced professionals from Singapore Eye Research Institute. There are 650 fundus images (482 normal and 168 glaucomatous) with a resolution of 3072×2048 pixels. This dataset is widely used as a benchmark for the diagnosis of glaucoma [38].

The RIM-ONE is a very popular publicly available dataset for ONH segmentation and glaucoma detection. The database was created by collaboration of three Spanish hospitals, i.e., Hospital Universitario de Canarias, Hospital Clnico San Carlos and Hospital Universitario Miguel Servet. There are 455 fundus images, 261 normal and 194 glaucoma [39].

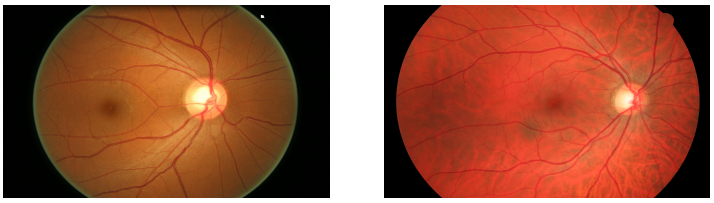
Another 124 fundus images, are collected from local private hospital, i.e., Armed Forces Institute of Ophthalmology (AFIO), Military Hospital, Rawalpindi, Pakistan. Similarly, 55 images are also acquired from local private hospital, i.e., Hayatabad Medical Complex (HMC), Peshaware, Pakistan. These images are annotated by two glaucoma experts with ten years of experience in glaucoma department. The number of normal and glaucoma images in each database are listed in Table 1.

Table 1. Key statistics of fundus images used to train/test the Deep CNNs.

Dataset	Normal	Glaucoma	Total
ACRIMA	309	396	705
ORIGA-light	482	168	650
RIM-ONE	261	194	455
AFIO	85	39	124
HMC	40	15	55
	1177	**812**	**1989**

The images of different datasets are shown in Figure 3. The difference between normal and glaucoma images are also presented in Figure 4. The first row shows the normal while the second row is the glaucoma images.

(a) ACRIMA **(b)** RIM-ONE **(c)** AFIO



(d) ORIGA-Light **(e)** HMC

Figure 3. Examples of fundus images of different datasets.

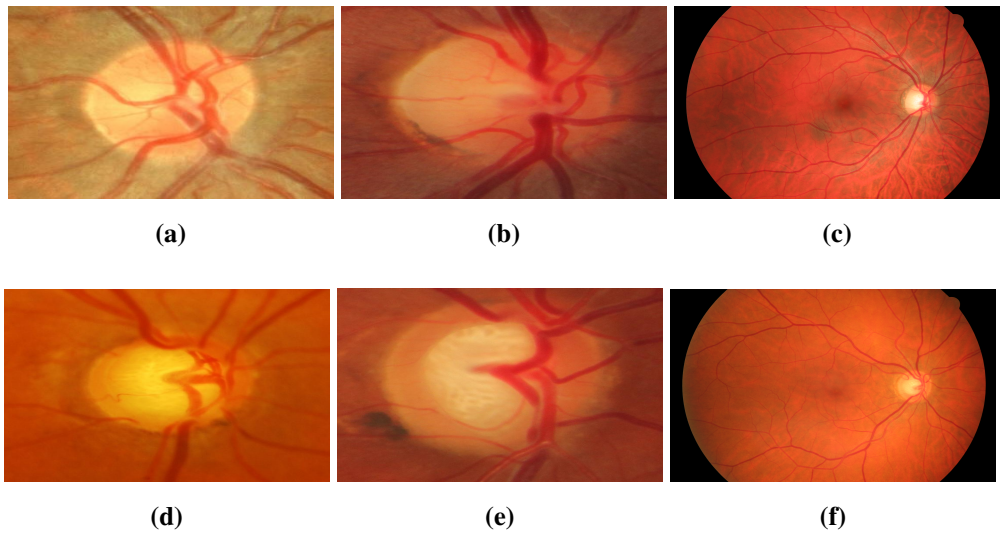


Figure 4. Normal and glaucoma images, (a)–(c) are normal while (d)–(f) are glaucoma images.

2.2. Pre-processing

The data pre-processing steps include the image patch extraction centered at ONH and data augmentation for the training of Deep CNNs.

First of all, we process all the images into a standard format that are used to train Deep CNNs. Image patches, centered at ONH, are extracted at same size according to the requirement of different Deep CNNs. Bicubic interpolation [40] is considered for resizing. It is examined that output pixel values are weighted average of pixels in 4-by-4 neighborhood pixels. The fundus images are cropped by evaluating the bounding box of 1.5 times the optic disc radius. Meanwhile, the illumination and contrast enhancement procedures are avoided to make the Deep CNNs learning more dynamic. In case of images from the local hospitals, i.e., AFIO and HMC, we have to localize ONH at the center of fundus images. It is to note that glaucoma disease mainly affects the ONH and its surrounding area. The cropping of images around ONH turned out to be more effective as compared to whole image, used for the training of Deep CNNs [41]. Moreover, the computational cost is also reduced during network learning.

The data augmentation technique is also explored during training of Deep CNNs to increase the training images and minimize over-fitting. The fundus images are invariant for flipping, rotation and translation. Hence, these three steps have been considered to increase the data for training of Deep CNNs.

2.3. Deep CNN feature learning

Currently, Deep CNNs are applied to a wide variety of applications in image segmentation and classification tasks. In this work, we also implement ImageNet trained Deep CNNs for the diagnosis of glaucoma through retinal fundus images. We have explored four types of Deep CNNs, i.e., AlexNet, InceptionV3, InceptionResNetV2 and NasNet-Large for the classification of normal and glaucomatous images. The basic architecture as well as feature extraction scheme of each network are illustrated in

the forthcoming paragraphs.

AlexNet, first proposed in [42] , is an extremely powerful model in achieving high accuracies on the challenging databases. It is very similar architecture as LeNet [43] but much deeper with more filters per layer with stacked CNNs layers. In this model, there are 5 convolutional and 3 Fully Connected (FC) layers with max pooling, Rectified Linear Unit (ReLu), and dropout layers. The input to this model is an RGB image of size 256×256. The basic structure of this network is shown in Figure 5.

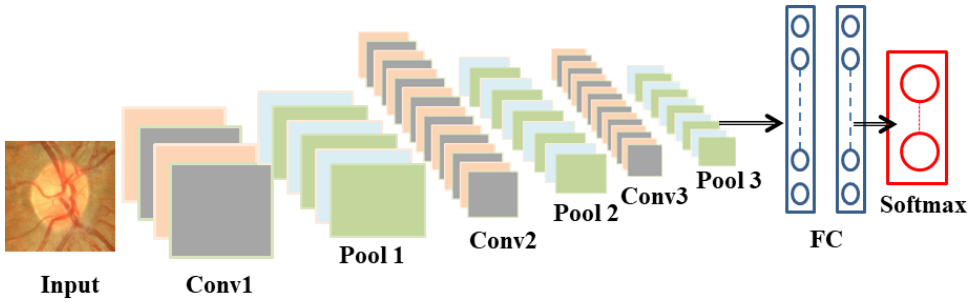


Figure 5. The basic architecture of AlexNet used for glaucoma classification.

InceptionV3 [44] is an extended network of the GoogLeNet [45] with good classification performance in several biomedical applications with transfer learning [46, 47]. InceptionV3 originally contains 11 Inception modules and one FC layer. Each Inception module also has 4 to 10 multiscale convolutional operations conducting with $1 \times 1, 3 \times 3$, or 5×5 filters. Max-pooling is used as the spatial pooling operation in the inception modules and FC is for final classification. This design reduces the computational complexity as well as number of parameters to be trained. It is trained on more than a million images from the ImageNet database. The network has an image input size of 299×299. The fundamental architecture of InceptionV3 is illustrated in Figure 6.

We also consider the state-of-the-art InceptionResnetV2 network [48] to extract deep spatial features. It is a combination of two recent networks, one is Residual Connections [49] and another is Inception structure [44]. The engaged InceptionResNetV2 network includes Stem, InceptionResNet, Reduction layers, average pooling layer and a FC layer. The Stem includes preliminary convolution operations executed before entering the Inception blocks. The InceptionResNet layers have residual connections with convolution operations while the Reduction layers are responsible for changing the width and height of the image. The spatial features are extracted from the convolutional layers and the pooling layers decrease the dimensionality of individual feature map, but hold the most significant features. Furthermore, the convolutional layers are followed by the batch normalization layer and ReLU, which is a nonlinearity function and helped to decrease the training time. The network has an image input size of 299×299. Figure 7 illustrates the architecture of deep spatial feature extraction using InceptionResNetV2 network.

Neural Architecture Search Network Large (NASNet-Large) is basically composed of two kinds of layers or cells, i.e., Normal and Reduction Cells. During the forward path, the width and height of feature map is reduced half by Reduction Cell while the Normal Cell retain these two dimensions same as the input feature map. The Normal Cells are stacked between Reduction Cells as shown in Figure 8. Each cell in Normal / Reduction Cell is composed of a number of blocks. Each block is built from the set of popular operations in Deep CNNs with various kernel size e.g.: convolutions, max pooling,

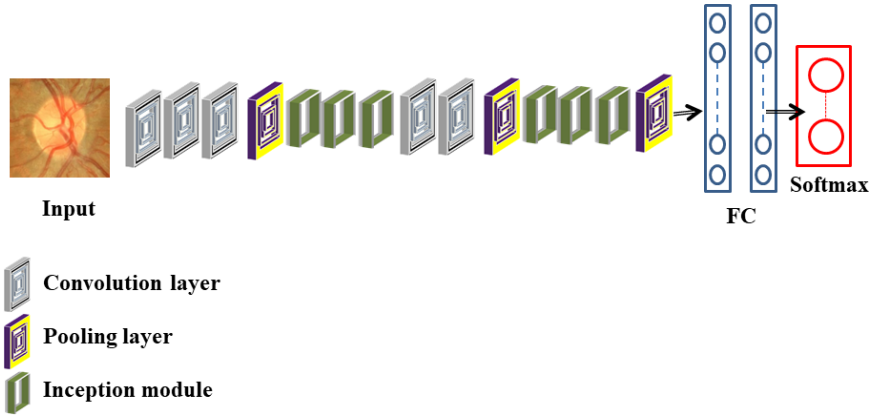


Figure 6. The InceptionV3 basic structure used for deep glaucoma feature learning.

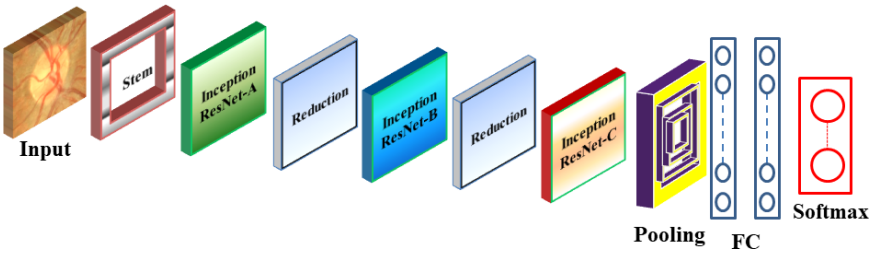


Figure 7. The InceptionResNetV2 network implemented for Deep spatial feature extraction.

average pooling, dilated convolution, depth-wise separable convolutions. Finding best architecture for Normal Cell and Reduction Cell with 5 blocks is described in [50]. NASNet-Large with N equals to 6 aims to get maximum possible accuracy. With the help of this arrangement, the network is able to learn rich feature representations for a wide range of images [51]. This network has an image input size of 331×331.

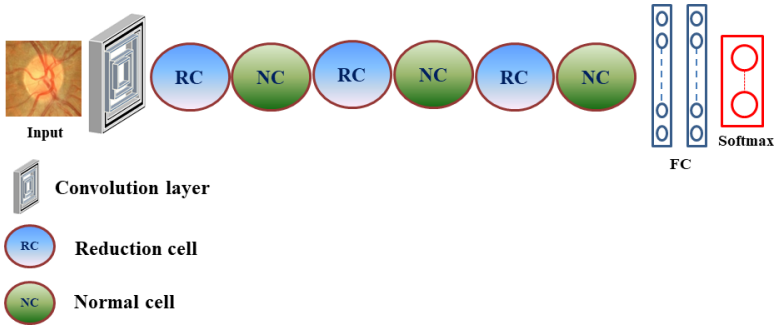


Figure 8. The structure of the NasNet-Large architecture considered for glaucoma diagnosis.

It is to note that the inputs in the above mentioned Deep CNN architectures are pre-processed color fundus images with centered at ONH . The region of interest is also kept same for all the models while the last FC layer is modified with softmax classifier to study only two classes, i.e., normal and glaucoma.

2.4. *Ensemble (4x Deep CNNs)*

In case of an ensemble framework, a set of diverse classifiers are aggregated and are trained to solve the same problem. All the above mentioned architectures are combined in one classifier as presented in Figure 9. The final decision is given by any of the five voting techniques, i.e., Majority Voting (MV), Proportional Voting (PV), Averaging (AV), Accuracy based Weighted Voting (AWV), and Accuracy/Score based Weighted Averaging (ASWA).

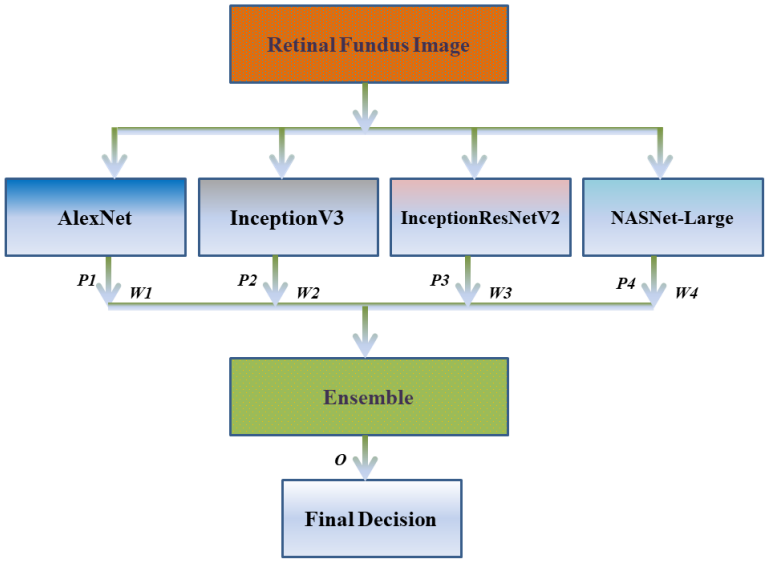


Figure 9. The proposed ensemble classifier for automatic glaucoma classification.

In MV [52], each model makes a prediction (vote) for each test instance and the final output prediction is the one that receives more than half of the votes. Here V_{ij} is the vote for *jth* class with reference to *ith* classifier and $N(V_j)$ is the total number of votes for *jth* class. We predict the class label, O, via highest number of votes. Mathematically, it is written as

$$N(V_j) = \sum_{i=1}^{N} V_{ij} \tag{2.1}$$

$$O = Max\{N(V_1), N(V_2), ..., N(V_j)\} \tag{2.2}$$

where $N(V_1)$, $N(V_2)$, and $N(V_3)$ are the total number of votes for class 1, class 2, and class 3, respectively.

In PV, the training accuracy of each classifier are summed-up with respect to their prediction and the maximum result will be the final outcome. A_{ij} is the accuracy for *jth* class with respect to *ith* classifier and $T(A_j)$ is the sum of training accuracies for *jth* class. We can write as

$$T(A_j) = \sum_{i=1}^{N} A_{ij} \tag{2.3}$$

$$O = Max\{T(A_1), T(A_2), ..., T(A_j)\} \tag{2.4}$$

where $T(A_1), T(A_2)$, and $T(A_3)$ are the sum of training accuracies for class 1, class 2, and class 3, respectively.

However, in case of equal votes, the AV is considered. The score vector for each predicted class are summed-up and averaged. S_{ij} is the score for *jth* class according to *ith* classifier and S_j is the average score for *jth* class. The output class, O, is the one corresponding to the highest value, such as

$$S_j = \frac{1}{N}\sum_{i=1}^{N} S_{ij} \tag{2.5}$$

$$O = Max\{S_1, S_2, ..., S_j\} \tag{2.6}$$

where S_1, S_2, and S_3 are the averaged score for class 1, class 2, and class 3, respectively.

In AWV, accuracy of each classifier is modified according to the following relation,

$$\alpha_i = \frac{e^{-10(1-ACCi)}}{\sum_{i=1}^{k} e^{-10(1-ACCi)}} \tag{2.7}$$

where α_i is the updated accuracy of *i*th classifier, i.e., $ACCi$. The final decision, O, is assigned to the maximum value, given as

$$(AWV)_j = \sum_{i=1}^{N} \alpha_i \times W_{ij} \tag{2.8}$$

$$O = Max\{(AWV)_1, (AWV)_2, ..., (AWV)_j\} \tag{2.9}$$

where W_{ij} is the weight of *jth* class with reference to *ith* classifier and $(AWV)_1, (AWV)_2$, and $(AWV)_3$ are the accuracy based weighted votes for class 1, class 2, and class 3, respectively.

In ASWA, the probabilities / Scores are used to calculate the weighted average based on accuracy. Mathematically, it is written as

$$(ASWA)_j = \sum_{i=1}^{N} \alpha_i \times S_{ij} \tag{2.10}$$

$$O = Max\{(ASWA)_1, ..., (ASWA)_j\} \tag{2.11}$$

where α_i is already calculated from Eq (2.7), and S_{ij} is the probability / score of *jth* class with respect to *ith* classifier. The final output decision is evaluated according to the Eq (2.11). $(ASWA)_1, (ASWA)_2$, and $(ASWA)_3$ are the accuracy/score based weighted averaging of class 1, class 2, and class 3, respectively.

2.5. Classification/diagnosis

The classification task is performed by the Softmax layer. During training a network, this layer updates the weights through back propagation. This process is based on the loss function used in the training stage. Keeping in view the underlying binary classification problem, the Cross Entropy (C.E) has been used as a loss function, as shown in Eq (2.12).

$$\begin{aligned} C.E &= -\sum_{t}^{C=2}(t_i \log(S_i)) \\ &= -t_1 \log(S_1) - (1-t_1)\log(1-S_1) \end{aligned} \tag{2.12}$$

where t_i and S_i are the ground truth and the Deep CNNs score for each class i in C.

We have analyzed the fundus images in two stages. In stage one, the performance of each Deep CNN has been evaluated for glaucoma diagnosis. In stage two, all four Deep CNNs are grouped together as a single classifier to further improve the accuracy of the system. Five voting schemes with two newly proposed techniques are considered to calculate the final decision.

3. Experiments and results

3.1. Experimental protocols

First, all fundus images are divided into three groups, training, validation and testing images. The sixty percent of all is for training while thirty percent of remaining for validation and rest for testing purpose. Table 2 shows the distribution of images in each group.

Table 2. Distribution of the fundus images for training, validation and testing sets.

Class	Training	Validation	Testing
Normal	706	141	330
Glaucoma	487	98	227

The validation group is selected to monitor the number of epochs in the training process of different deep architectures. All the images used in validation, shared with the training set after selecting hyper-parameters and cross validation experiments. Thus, the training set has 847 normal and 585 glaucoma images. But, the testing sets are kept same during all experiments. The distribution of test images in different datasets is presented in Table 3.

Table 3. The distribution of test images in each datasets.

Dataset	Normal	Glaucoma
ACRIMA	87	110
ORIGA-Light	135	47
RIM-ONE	73	55
AFIO	24	11
HMC	11	4

Commonly, the transfer learning strategy is applied in Deep CNNs to use the knowledge learned while classifying natural images to classify retinal images with glaucoma. Hence, the transfer learning has been employed in this work that involves, replacing and retraining the softmax layer and also fine-tuning the weights of the pre-trained network. We have carried out several experiments to achieve the optimal performance of each Deep CNNs with different number of fine-tuned layers and number of epochs. Initially, we have considered the last weighted layers of Deep CNNs for the number of fine-tuned layers, while keeping the initial layers in a freezing state. After that, the number of fine-tuned layers is increased until updating all the remaining layers in the Deep CNNs. Secondly, the effect of number of epochs have been evaluated for the best performance of each model. It is to note that we get maximum performance for each Deep CNNs for 30 epochs. The Stochastic Gradient Descent

(SGD) is used as the optimizer. The other hyper-parameters, the learning rate, the batch size, and the momentum are optimally selected to get the best results in different sets of experiments. The objective of optimal setting is to adjust the weights of each Deep CNN such that the training loss is minimum. By minimizing the loss, we can achieve the optimal parameters resulting in the best model performance. Batch size is the number of training examples used in the estimation of error gradient for the learning algorithm. We take its value smaller because smaller batch size make it easier to fit one batch worth of training data in memory, offering a regularizing effect and lower the generalization error. The learning rate decides how far to move the weights in the direction of gradient to get the minimum loss. However, an optimal value is required to reach a minimum loss quickly. This is because smaller learning rate will take tiny steps to and hence a large time is required to reach the minimum loss function. While, the higher learning rate will result in overshooting the minimum loss and the Deep CNN may not converge. Besides, momentum is another parameter used to optimize the learning rate. It aims to calculate the weighted average of weights between the average of previous values and the current value. Thus, we have set the batch size to 6, the learning rate to $1 \times e^{-4}$ and the momentum to 0.9 for all the networks.

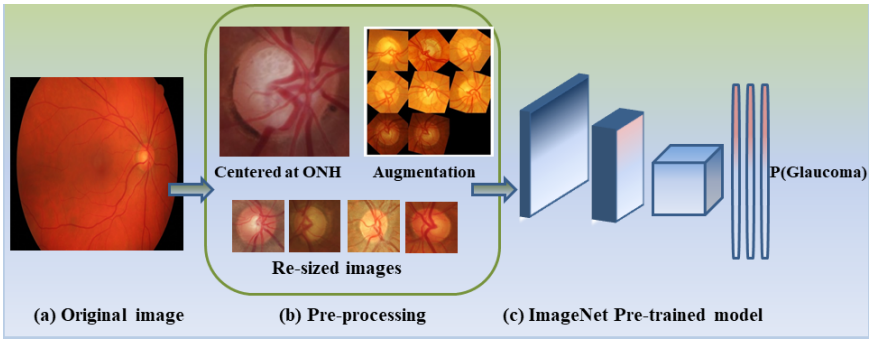


Figure 10. The flowchart of training process for Deep CNNs.

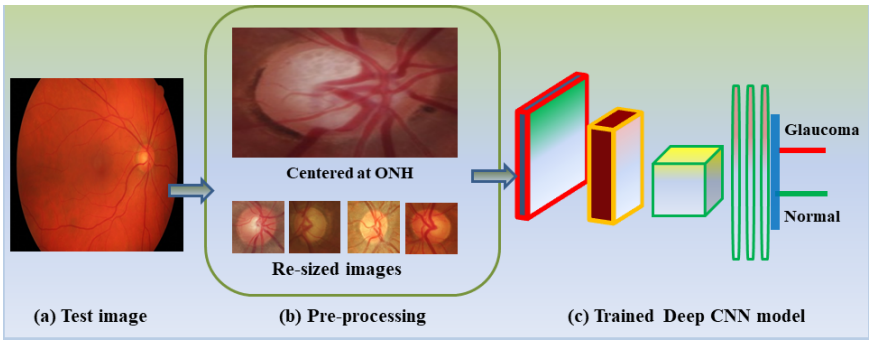


Figure 11. The flowchart of testing process for Deep CNNs.

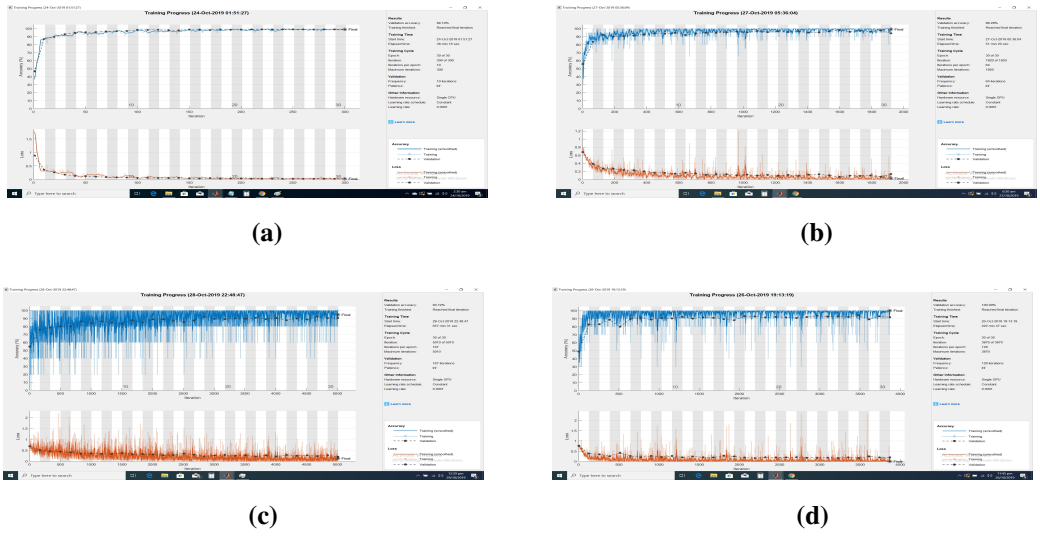
We have also assessed the performances of Deep CNNs using 10 cross validation technique. Due to limited training data, over-fitting is a well-known problem, occurred in Deep CNNs. To avoid over-fitting and increase the robustness of the architectures, we have pondered the dropout technique as proposed in [53] that temporally remove units along with all its incoming and outgoing connections in deep neural networks. Similarly, we have also employed data augmentation technique during training of all the networks. The fundus images are augmented by using random rotations between 0 to 360

degrees, and random translation of maximum 10 pixels in both x and y direction of the image. The input images are also resized according to the default input size of each Deep CNNs. Hence, we have evaluated the performance of each Deep CNNs using all datasets and experiments have been carried out to compare the best of the above mentioned four Deep CNNs. The training and testing procedures are displayed in Figures 10 and 11, respectively.

3.2. Experimental results and evaluation

3.2.1. Deep CNNs learning results

The number of epochs has been evaluated for each of the selected Deep CNN during the training process. The validation accuracy/loss during fine-tuning process of all four Deep CNNs for ACRIMA dataset are illustrated in Figure 12. It is observed that after 30 epochs, the validation accuracy reaches its maximum values, i.e., 99.10, 99.28, 94.72 and 100% for AlexNet, InceptionV3, InceptionResNetV2 and NasNet-Large, respectively. Similarly, the validation accuracy/loss results for other Deep CNNs are also evaluated.



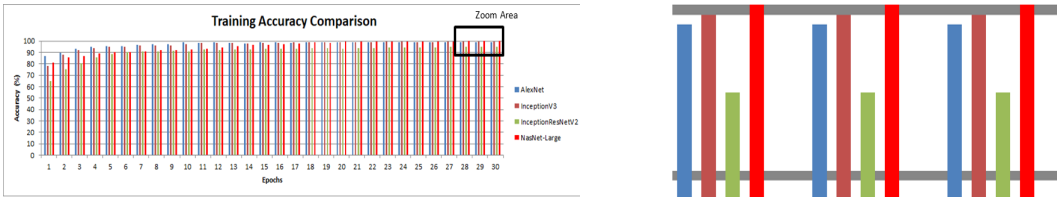
(a) **(b)**

(c) **(d)**

Figure 12. Fine-tunning process of Deep CNNs for ACRIMA dataset (a) AlexNet (b) InceptionV3 (c)InceptionResNetV2 (d) NasNet-Large.

It is observed that the NasNet-Large has maximum accuracy, i.e., 100% for ACRIMA dataset while for ORIGA-Light it has minimum value, i.e., 87.10% as compared with other networks. On average, all Deep CNNs perform well on ACRIMA dataset and the lowest results have been considered on ORIGA-Light dataset. This is because ACRIMA is a newly developed dataset for classification of glaucoma images while ORIGA-Light is designed for the segmentation of optic cup and optic disc. The superiority of the NasNet-Large model has also been observed during training for all other datasets. The training accuracy based comparison among 4x Deep CNNs for ACRIMA dataset is illustrated in Figure 13. It is investigated that the NasNet-Large outperforms all other networks during training. It is also applicable for other datasets.

Furthermore, Deep CNNs training time has been reduced through transfer learning. The weights in pre-trained networks are used as the starting point for the training process. Hence, we have minimized



(a) Training accuracy comparison chart

(b) Zoom area for last 3 epochs

Figure 13. The training accuracy based comparison among 4x Deep CNNs for ACRIMA dataset.

the training epochs, i.e., 30 for all Deep CNNs and achieved best validation results, as displayed in Table 4.

Table 4. The training accuracies of Deep CNNs for all the datasets.

Deep CNNs	ACRIMA	ORIGA-Light	RIM-ONE	AFIO	HMC
AlexNet	99.10	75.50	86.70	90.90	98.50
InceptionV3	99.28	78.50	86.40	84.50	90.50
InceptionResNetV2	94.72	76.50	90.60	85.50	95.50
NasNet-Large	100.00	87.10	94.70	97.70	98.60

3.2.2. Deep CNNs testing results

Commonly, a single evaluation metric is not appropriate to evaluate the performance of a given algorithm due to the presence of some imbalanced classes in the dataset or a large number of training labels [54]. Therefore, the performance of Deep CNNs are reported in terms of five distinct metrics including Accuracy (ACC), Sensitivity (SEN), Specificity (SP), F1 score and Area Under the Curve (AUC) as proposed in the previous studies [55]. These performance parameters are calculated using the following equations:

$$ACC = \frac{TP + TN}{TP + FP + TN + FN} \tag{3.1}$$

$$SEN = \frac{TP}{TP + FN} \tag{3.2}$$

$$SP = \frac{TN}{TN + FP} \tag{3.3}$$

$$F1 = 2\frac{precision.recall}{precision + recall} \tag{3.4}$$

where the precision and recall are expressed as

$$Precision = \frac{TP}{TP + FP} \tag{3.5}$$

$$Recall = \frac{TP}{TP + FN} \tag{3.6}$$

In the above equations, the True Positive (TP) is defined as the number of glaucoma images classified as glaucoma and True Negative (TN) is the number of normal images classified as normal. False Positive (FP) is the number of normal images identified as glaucoma images and False Negative (FN) is the number of glaucoma images classified as normal.

AUC is the area under the Receiver Operating Curve (ROC) and it provides the probability that the model ranks a positive example more highly than a negative example. ROC is a plot between two parameters, i.e., True Positive Rate (TPR) and False Positive Rate (FPR). TPR is synonym for recall while FPR can be calculated as

$$FPR = \frac{FP}{FP + TN} \tag{3.7}$$

The confusion matrices of each Deep CNNs have been evaluated for the test images in each dataset. Figures 14 and 15 show the confusion matrices for ACRIMA test images and total test images, respectively. Similarly, for other test images, ACC, SEN and SP results are also calculated according to Eqs (3.1), (3.2) and (3.3).

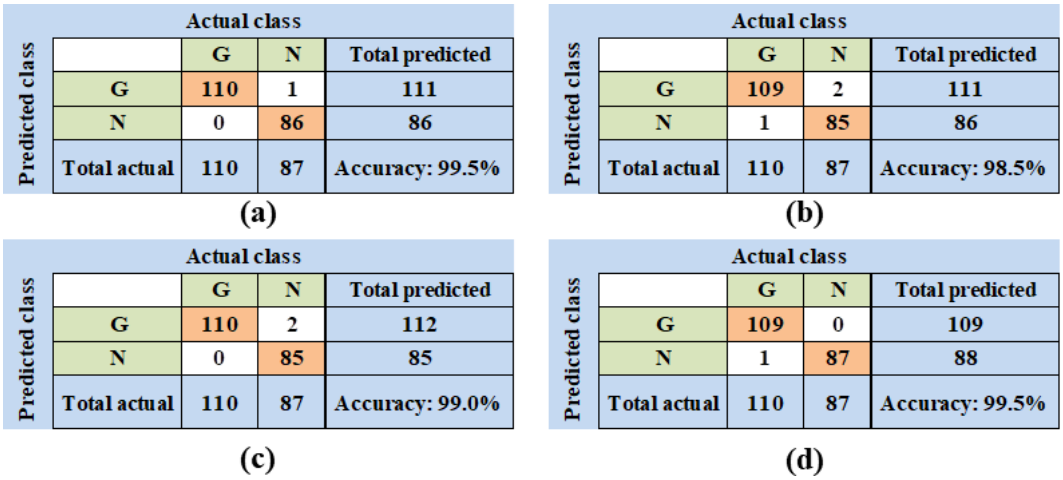


Figure 14. The test results of Deep CNNs for ACRIMA dataset (a) AlexNet (b) InceptionV3 (c) InceptionResNetV2 (d) NasNet-Large.

It is observed that NasNet-Large achieves best results over ACRIMA dataset, i.e., ACC (99.5%), SP (100%) and SEN (99%). The InceptionResNetV2, gives ACC (99%), SP (97.7%) and SEN (100%). Now for ORIGA-Light test images, it can be seen that all the Deep CNNs show poor performance except NasNet-Large with ACC (88%), SP (91%) and SEN (79%). The AlexNet, InceptionV3, and InceptionResNetV2 achieve worst results in the range of 60–66% SEN.

All the Deep CNNs perform well on RIM-ONE test images. Like, NasNet-Large again gives better results in terms of ACC (94.5%), SP (96%) and SEN (92.7%). While, the AlexNet shows lowest results with ACC (87.5%), SP (90.4%) and SEN (83.6%) as compared with other networks.

The performance metrics are also evaluated for both the local datsets, i.e., AFIO and HMC. It is noticed that NasNet-Large provides maximum results in terms of ACC, SP, and SEN for AFIO test

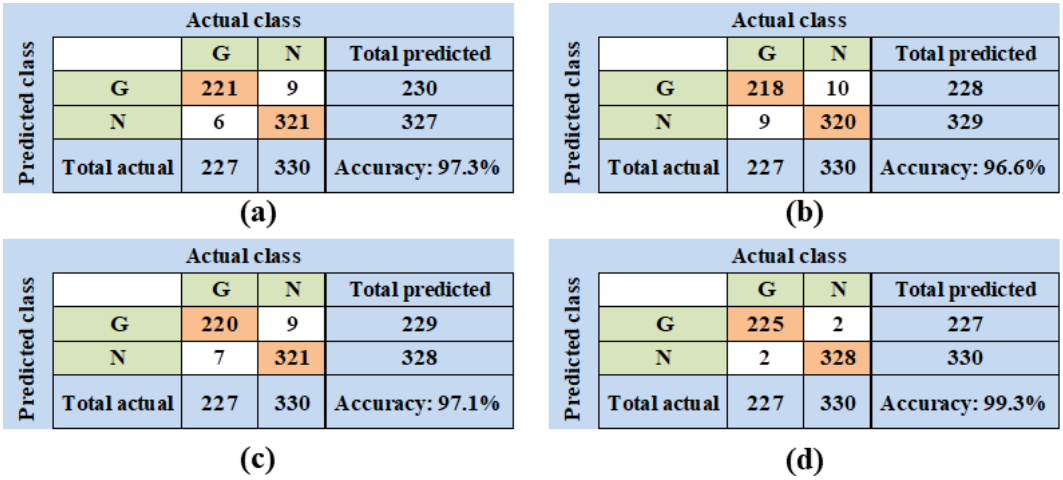


Figure 15. The test results of Deep CNNs for total test images (a) AlexNet (b) InceptionV3 (c) InceptionResNetV2 (d) NasNet-Large.

images while InceptionResNetV2 has 88.6% ACC, 83.3% SP, and 100% SEN. Similarly, NasNet-Large gives maximum results, i.e., 100% in all performance metrics while AlexNet has 93.3% ACC, 100% SP, and 75.0% SEN with HMC test images.

In case of total test set of images, the NasNet-Large again performs well as compared to other networks. It provides 99.3% ACC, 99.4% SP, and 99.1% SEN, while InceptionV3 shows lowest results in terms of ACC, SP, and SEN. These results are displayed in Table 5.

Table 5. The test results of the Deep CNNs for all the test images selected from each dataset.

Deep CNNs	Datasets								
	ACRIMA			ORIGA-Light			RIM-ONE		
	ACC	SP	SEN	ACC	SP	SEN	ACC	SP	SEN
AlexNet	99.5	98.9	100.0	75.8	81.5	60.0	87.5	90.4	83.6
InceptionV3	98.5	97.7	99.1	78.6	83.0	66.0	92.2	94.5	89.1
InceptionResNetV2	99.0	97.7	100.0	76.9	82.2	61.7	90.6	94.5	85.5
NasNet-Large	99.5	100.0	99.1	87.9	91.1	78.7	94.5	95.9	92.7
	AFIO			HMC			TOTAL		
	ACC	SP	SEN	ACC	SP	SEN	ACC	SP	SEN
AlexNet	91.4	95.8	81.8	93.3	100.0	75.0	97.3	97.3	97.4
InceptionV3	91.4	91.7	91.0	86.7	100.0	50.0	96.6	97.0	96.0
InceptionResNetV2	88.6	83.3	100.0	93.3	91.0	100.0	97.1	97.3	96.9
NasNet-Large	94.3	91.7	100.0	99.9	100.0	100.0	99.3	99.4	99.1

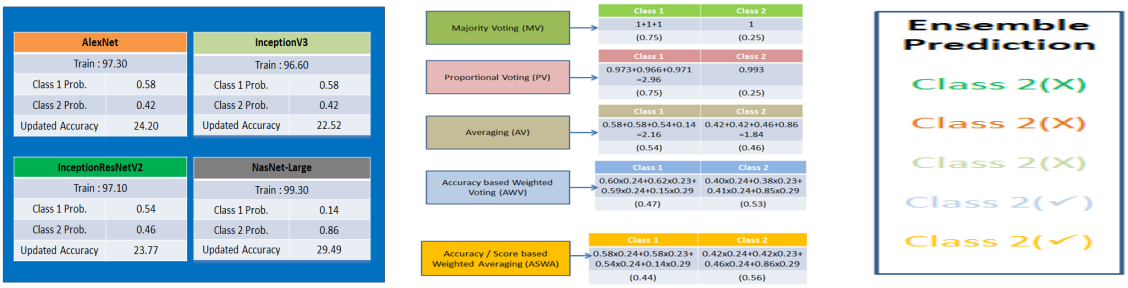
3.2.3. Deep CNNs ensemble results

It is noted that the classification accuracies have been improved using an ensemble of all four Deep CNNs. For ACRIMA dataset, NasNet-Large and AlexNet give 99.5% accuracy while ensemble

with AWV provides 99.6% accuracy. In case of ORIGA-Light, the accuracy has been increased from 87.9 to 88.3% with ASWA. Similarly, for RIM-ONE, AFIO, and HMC datasets, the results have been improved. Now, if we consider total test set of images, the result is increased from 99.3 to 99.5% with ASWA. The AV provides lowest results. This is because it is simple averaging. However, the AWV and ASWA provide better results. This is because the weights are updated according to the accuracy and score of each Deep CNNs. The results of five voting schemes are displayed in Table 6. Figure 16 also provides an overview of newly developed AWV and ASWA schemes that makes it different from MV, PV, and AV.

Table 6. The accuracies of Deep CNNs ensemble framework for different voting schemes.

Dataset	MV	PV	AV	AWV	ASWA
ACRIMA	99.5	99.4	99.1	99.6	99.5
ORIGA-Light	87.9	88.0	79.8	88.2	88.3
RIM-ONE	94.5	94.5	91.2	95.1	95.2
AFIO	94.3	94.3	91.4	96.1	96.2
HMC	99.9	99.9	96.3	99.8	99.9
Total	99.3	99.3	98.5	99.4	99.5



(a) 4x classifiers **(b)** Voting results **(c)** Final predictions

Figure 16. Illustration of five voting techniques with test instance of class 2.

Besides, we have also considered the combinations of two and three Deep CNNs to increase the persuasiveness of the proposed ensemble framework. The results of different voting schemes with ensemble of 2x, 3x, and the proposed 4x Deep CNNs for total test set of images are presented in Table 7. It is observed that the AlexNet and NasNet-Large provide better results as comapred to other networks in 2x Ensemble framework. While, in the case of 3x Ensemble network, AlexNet, InceptionResNetV2 and NasNet-Large show superior results with other combinations. However, our proposed 4x Ensemble framework outperforms in all five voting schemes as compared to 2x, and 3x Deep CNNs ensemble.

The evaluation parameters of classification performance, i.e., sensitivity, specificity, accuracy, and AUC of the ImageNet trained Deep CNNs have been displayed in Table 8, where the performance comparison of proposed work with [22, 26, 37], and [56–61] is presented. In [57–60], the authors have used CNN based architectures for the classification of glaucoma using RIM-ONE, and ORIGA-Light

Table 7. The accuracies of various combinations of Deep CNNs for five voting schemes with total test set of images.

	Deep CNNs	MV	PV	AV	AWV	ASWA
2*x* Ensemble	A+B	98.70	98.60	97.80	99.00	98.80
	A+C	98.80	98.50	97.60	99.00	98.90
	A+D	99.00	98.90	97.80	99.20	99.00
	B+C	98.80	99.00	96.90	98.90	98.80
	B+D	98.80	98.70	97.50	98.90	98.90
3*x* Ensemble	A+B+C	98.90	98.80	98.10	98.80	98.90
	A+B+D	98.80	99.00	98.10	98.80	99.00
	A+C+D	99.00	99.10	98.20	99.00	99.00
	B+C+D	98.90	99.10	98.10	98.80	98.90
4*x* Ensemble	A+B+C+D	99.30	99.30	98.50	99.40	99.50

A = AlexNet, B = InceptionV3, C = InceptionResNetV2, & D = NasNet-Large

images. Similarly, in another study proposed by [61], the authors have considered newly developed ACRIMA dataset for the glaucoma classification and achieved highest 96.0% AUC. While, in our study, the NasNet-Large performed well and gives 99.1% sensitivity, 99.4% specificity, 99.3% accuracy and 97.8% AUC for the total test set of images. This is because the cropped images centered at ONH are to be more effective as compared to whole image as well as computational cost is reduced during network learning. It is also helpful in improving the identification of glaucomatous damages in early stages. Furthermore, a data augmentation technique is also considered during training of Deep CNNs to increase the training images and minimize over-fitting problem. The classification accuracy has been further increase from 99.3 to 99.5% with ensemble framework. The results show that AWV and ASWA provide better results as compared with other voting techniques. This is because weights are updated according to accuracy and scores of each Deep CNNs. These results indicate that the proposed study provides better performance than its previous state-of-the-art. To the best of our knowledge, there is no existing related Deep CNNs based ensemble framework for the diagnosis of glaucoma using fundus images.

4. Results related discussion

The experimental results presented in this study suggest the following key observations:

- From the above mentioned results, the proposed automatic glaucoma diagnosis system is more useful and effective. This is because Deep CNNs have the ability to learn glaucoma specific features automatically. While in traditional methods, the feature extraction strategies are manual that limit the success of overall system. Moreover, the illumination and textural variations in retinal fundus images are also the problems in classification of glaucoma. Conversely, the presented work uses automatic glaucoma features extraction and thus achieve superior classification accuracy across a wide range of publicly and locally available datasets. Hence, the results are generalized for diverse set of images.

Table 8. The comparison of proposed work with existing state-of-the-art.

Systems	Methods	Dataset	Sensitivity (%)	Specificity (%)	Accuracy (%)
[22]	CNN	Public & Private	84.5	98.0	99.0
[26]	Multi-branch NN	Private	92.0	90.0	91.0
[37]	Deep CNNs	Public	93.4	85.8	96.0 (AUC)
[56]	CNN	Private	98.0	98.0	98.0
[57]	CNN	RIM-ONE	80.0	88.0	85.0
[58]	RCNN	ORIGA-Light	NR	NR	87.4 (AUC)
[59]	GoogleNet	HRF & RIM-ONE	NR	NR	87.6
[60]	AlexNet	RIM-ONE	87.0	85.0	88.0
[61]	Deep Models	Public & Private	NR	NR	85.0 (AUC)ACRIMA
[62]	InceptionV3	Private	NR	NR	84.5 & 93.0 (AUC)
[63]	Ensemble Classifier	Private	86.0	90.0	88.0 & 94.0 (AUC)
Our	**NasNet-Large**$*$	Public & Private	**99.1**	**99.4**	**99.3** & **97.8% (AUC)**
Our	**Ensemble Classifier**$*$	Public & Private	**99.1**	**99.7**	**99.5**

- In pre-processing step, the region ONH has been extracted and used as the input image to Deep CNNs models. This is due to the fact that most of the initial changes have been occurred in ONH during early stage of glaucoma. Hence, the cropped images centered at ONH are to be more effective as compared to whole image as well as computational cost is reduced during network learning. It is also helpful in improving identification of glaucomatous damages in early stages. Furthermore, a data augmentation technique is also considered during training of Deep CNNs to increase the training images and minimize over-fitting problem.
- Transfer learning generally refers to a process where a model trained on one problem is used in some way on a second related problem. In our study, Deep CNNs training time has been reduced through transfer learning. The weights in pre-trained networks are used as the starting point for the training process. Hence, we have minimized the training epochs, i.e., 30 for all Deep CNNs and achieved best validation results, as displayed in Table 4. Additionally, a graphical comparison is also being presented among these models during the training process. It is examined that our proposed model, i.e., NasNet-Large achieves best results as compared to other Deep CNNs. These results are displayed in Figure 13.
- Generally, a single performance metrics can lead to inappropriate classification results due to some imbalance classes in the dataset or too small or large number of training subjects. From the literature survey of the existing methods on fundus images such as [37,58,61] show classification performance in terms of AUC only. In contrast, we have evaluated four distinct metrics including SEN, SP, ACC, and AUC. The results show the steady performance in glaucomatous classification across different metrics.
- We have also performed extensive experiments to evaluate the performance of four Deep CNNs (AlexNet, InceptionV3, InceptionResNetV2 and NasNet-Large). Comparing the results of these architectures, it is noted that NasNet-Large is significantly better than that of others networks. The results are consistent with the relative performance of these architectures on wide range of public and private fundus images. The AlexNet also provides better results as compared with

other networks. However, it shows lower performance on ORIGA-Light and RIM-ONE datasets as compared with others. From the results, it is observed that ORIGA-Light is more challenging dataset for each type of network. On the contrary, all the networks provide good results for ACRIMA dataset because it is a newly developed dataset for Deep learning classification tasks.

- We have also presented an ensemble classifier to further improve the classification accuracy. The final results have been evaluated with five voting techniques. Two newly developed voting schemes, i.e., AWV and ASWA provide the better results as compared with others as presented in Table 6. Moreover, we have also carried out ablation experiments to increase the persuasiveness of the experimental results as well as the validation of four Deep CNNs combination. The experimental results are displayed in Table 7. It is clearly observed that ensemble of four networks show better results as compared to ensemble of $2x$ and $3x$ networks. These results clearly demonstrate the effectiveness of the proposed ensemble framework to diagnose the glaucoma.
- The experimental results are also compared with other deep learning based diagnostic systems developed by other researchers. In [57–60], the authors have used CNN based architectures for the classification of glaucoma using RIM-ONE, and ORIGA-Light images. Similarly, in another study proposed by [61], the authors have considered newly developed ACRIMA dataset for the glaucoma classification and achieved highest 96.0% AUC. In [62], the authors have implemented InceptionV3 architecture for glaucoma detection and achieved maximum 84.5% accuracy with 93% AUC. More recently, an ensemble of AlexNet, ResNet-50, and ResNet-152 have investigated in [63] and achieved the highest accuracy of 88% with 0.94 AUC. However, our proposed ensemble framework have achieved superior results as compared with previously proposed methods. The results of SEN, SP, and ACC proposed in this study are displayed in Table 5. A detailed comparisons with existing state-of-the-art has been presented in Table 8.

5. Conclusions and future works

In this work, we have proposed an ensemble framework based on pre-trained Deep CNNs for glaucoma classification using fundus images. At first, four Deep CNNs, i.e., AlexNet, InceptionV3, InceptionResNetV2, and NasNet-Large are tested on five different datasets, three publicly available,i.e., ACRIMA, ORIGA-Light, and RIM-ONE, and others two collected from the local hospitals. Dropout and data augmentation techniques are also considered to improve the performance of Deep CNNs models. NasNet-Large is the best option with transfer learning and fine-tuning, with AUC (97.8%), SEN (99.1%), SP (99.4%) and ACC (99.3%). Secondly, for even better results, we also proposed an ensemble framework for automatic glaucoma classification. The AWV and ASWA based ensembling methods improve the accuracy with all datasets and total test images to 0.3%. Moreover, the proposed ensemble classifier has considerably better accuracy and robustness than the individual optimized Deep CNN models for automatic glaucoma diagnosis.

As a future work, the new architectures with more data can be explored and assessed to confirm the presented line of work. The performance of Deep CNNs can also be enhanced to extract deep features for the classification tasks. In this way, we can train even more robust glaucoma classifiers.

Acknowledgments

We would like to mention here the great struggle of Dr. Yousaf Jamal Mahsood, Associate Professor, Glaucoma Khyber Girls Medical College, HMC Peshawar, in collecting the fundus images and AFIO department, Military Hospital, Rawalpindi, Pakistan.

Conflict of interest

The authors declare no conflict of interest.

References

1. S. Kingman, Glaucoma is second leading cause of blindness globally, *Bull. World Health Organ.*, **82** (2014), 887–888.
2. Y. C. Tham, X. Li, T. Y. Wong, H. A. Quigley, T. Aung, C. Y. Cheng, Global prevalence of glaucoma and projections of glaucoma burden through 2040: a systematic review and Meta-analysis, *Ophthalmology*, **121** (2014), 2081–2090.
3. H. Quigley, A. T. Broman, The number of people with glaucoma worldwide in 2010 and 2020, *Br. J. Ophthalmol.*, **90** (2006), 262–267.
4. J. Fuente-Arriaga, E. M Felipe-Riverón, E. Garduño-Calderón, Application of vascular bundle displacement in the optic disc for glaucoma detection using fundus images, *Comput. Biol. Med.*, **47** (2014), 27–35.
5. M. D. Abramoff, M. K. Garvin, M. Sonka, Retinal imaging and image analysis, *IEEE Rev. Biomed. Eng.*, **3** (2010), 169–208.
6. M. S. Haleem, L. Han, J. Van Hemert, B. Li, Automatic extraction of retinal features from colour retinal images for glaucoma diagnosis: a review, *Comput. Med. Imaging Graphics*, **37** (2013), 581–596.
7. M. Shakeri, S. Tsogkas, E. Ferrante, S. Lippe, S. Kadoury, N. Paragios, et al., Sub-cortical brain structure segmentation using f-cnn's, in *2016 IEEE 13th International Symposium on Biomedical Imaging (ISBI)*, (2016), 269–272.
8. M. Jaderberg, A. Vedaldi, A. Zisserman, Speeding up convolutional neural networks with low rank expansions, preprint, arXiv:1405.3866.
9. J. Lemley, S. Bazrafkan, P. Corcoran, Smart augmentation learning an optimal data augmentation strategy, *IEEE Access*, **5** (2017), 5858–5869.
10. S. J. Pan, Q. Yang, A survey on transfer learning, *IEEE Trans. Knowl. Data Eng.*, **22** (2010), 1345–1359.
11. C. Li, D. Xue, X. Zhou, J. Zhang, H. Zhang, Y. Yao, et al., Transfer learning based classification of cervical cancer immunohistochemistry images, in *ACM International Conference Proceeding Series*, (2019), 102–106.

12. A. Ghoneim, G. Muhammad, M. S. Hossain, Cervical cancer classification using convolutional neural networks and extreme learning machines, *Future Gener. Comput. Syst.*, **102** (2020), 643–649.
13. H. Parvin, M. MirnabiBaboli, H. A. Rokny, Proposing a classifier ensemble framework based on classifier selection and decision tree, *Eng. Appl. Artif. Intell.*, **37** (2015), 34–42.
14. S. Maheshwari, V. Kanhangad, R. B. Pachori, Cnn-based approach for glaucoma diagnosis using transfer learning and lbp-based data augmentation, preprint, arXiv:2002.08013.
15. A. Singh, S. Sengupta, V. Lakshminarayanan, Glaucoma diagnosis using transfer learning methods, in *Applications of Machine Learning, International Society for Optics and Photonics*, (2019).
16. A. Serener, S. Serte, Transfer learning for early and advanced glaucoma detection with convolutional neural networks, in *2019 Medical Technologies Congress (TIPTEKNO)*, (2019), 1–4.
17. O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, et al., ImageNet large scale visual recognition challenge, *Int. J. Comput. Vision*, **115** (2015), 211–252.
18. H. N. Veena, A. Muruganandham, T. S. Kumaran, A novel optic disc and optic cup segmentation technique to diagnose glaucoma using deep learning convolutional neural network over retinal fundus images, *J. King Saud Univ. Comput. Inf. Sci.*, (2021), forthcoming.
19. X. Chen, Y. Xu, D. W. K. Wong, T. Y. Wong, J. Liu, Glaucoma detection based on deep convolutional neural network, in *2015 37th annual international conference of the IEEE engineering in medicine and biology society (EMBC)*, (2015), 715–718.
20. R. Asaoka, H. Murata, A. Iwase, M. Araie, Detecting preperimetric glaucoma with standard automated perimetry using a deep learning classifier, *Ophthalmology*, **123** (2016), 1974–1980.
21. X. Chen, Y. Xu, S. Yan, D. Wong, T. Wong, J. Liu, Automatic feature learning for glaucoma detection based on deep learning, in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, (2015), 669–677.
22. Q. Abbas, Glaucoma-deep: detection of glaucoma eye disease on retinal fundus images using deep learning, *Int. J. Adv. Comput. Sci. Appl.*, **8** (2017), 41–45.
23. J. Orlando, E. Prokofyeva, M. del Fresno, M. B. Blaschko, Convolutional neural network transfer for automated glaucoma identification, in *12th international symposium on medical information processing and analysis*, (2017).
24. A. Chakravarty, J. Sivswamy, A deep learning based joint segmentation and classification framework for glaucoma assesment in retinal color fundus images, preprint, arXiv:1808.01355.
25. Z. Li, Y. He, S. Keel, W. Meng, R. T. Chang, M. He, Efficacy of a deep learning system for detecting glaucomatous optic neuropathy based on color fundus photographs, *Ophthalmology*, **125** (2018), 1199–1206.
26. Y. Chai, H. Liu, J. Xu, Glaucoma diagnosis based on both hidden features and domain knowledge through deep learning models, *Knowl. Based Syst.*, **161** (2018), 147–156.

27. M. Christopher, A. Belghith, C. Bowd, J. Proudfoot, M. Goldbaum, R. N. Weinreb, et al., Performance of deep learning architectures and transfer learning for detecting glaucomatous optic neuropathy in fundus photographs, *Sci. Rep.*, **8** (2018), 1–13.
28. N. Shibata, M. Tanito, K. Mitsuhashi, Y. Fujino, M. Matsuura, H. Murata, et al., Development of a deep residual learning algorithm to screen for glaucoma from fundus photography, *Sci. Rep.*, **8** (2018), 14665.
29. S. Liu, S. Graham, A. Schulz, M. Kalloniatis, B. Zangerl, W. Cai, et al., A deep learning-based algorithm identifies glaucomatous discs using monoscopic fundus photographs, *Ophthalmol. Glaucoma*, **1** (2018), 15–22.
30. S. Gheisari, S. Shariflou, J. Phu, P. Kennedy, A. Agar, M. Kalloniatis, et al., A combined convolutional and recurrent neural network for enhanced glaucoma detection, *Sci. Rep.*, **11** (2021), 1945.
31. F. Li, L. Yan, Y. Wang, J. Shi, H. Chen, X. Zhang, et al., Deep learning-based automated detection of glaucomatous optic neuropathy on color fundus photographs, *Graefe's Arch. Clin. Exp. Ophthalmol.*, **258** (2020), 851–867.
32. H. I. Elshazly, M. Waly, A. M. Elkorany, A. E. Hassanien, Chronic eye disease diagnosis using ensemble-based classifier, in *2014 International Conference on Engineering and Technology (ICET)*, (2014).
33. J. Zilly, J. Buhmann, D. Mahapatra, Glaucoma detection using entropy sampling and ensemble learning for automatic optic cup and disc segmentation, *Comput. Med. Imaging Graphics*, **55** (2017), 28–41.
34. H. Fu, J. Cheng, Y. Xu, C. Zhang, D. Wong, J. Liu, et al., Disc-aware ensemble network for glaucoma screening from fundus image, *IEEE Trans. Med. Imaging*, **37** (2018), 2493–2501.
35. N. Gour, P. Khanna, Multi-class multi-label ophthalmological disease detection using transfer learning based convolutional neural network, *Biomed. Signal Process. Control*, **66** (2021), 102329.
36. A. Bhuiyan, A. Govindaiah, R. T. Smith, An artificial intelligence and telemedicine based screening tool to identify glaucoma suspects from color fundus imaging, *J. Ophthalmol.*, **2021** (2021), 6694784.
37. 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, *Biomed. Eng. Online*, **18** (2019), 29.
38. Z. Zhang, F. S. Yin, J. Liu, W. K. Wong, N. M. Tan, B. H. Lee, et al., Origa-light: An online retinal fundus image database for glaucoma analysis and research, in *2010 Annual International Conference of the IEEE Engineering in Medicine and Biology*, IEEE, (2010).
39. F. Fumero, S. Alayón, J. L. Sanchez, J. Sigut, M. G.-Hernandez, Rim-one: An open retinal image database for optic nerve evaluation, in *2011 24th international symposium on computer-based medical systems (CBMS)*, IEEE, (2011).
40. P. R. Rajarapollu, V. R. Mankar, Bicubic interpolation algorithm implementation for image appearance enhancement, *Int. J.*, **8** (2017).

41. J. Orlando, E. Prokofyeva, M. del Fresno, M. B. Blaschko, Convolutional neural network transfer for automated glaucoma identification, in *12th international symposium on medical information processing and analysis*, (2017).
42. A. Krizhevsky, I. Sutskever, G. E. Hinton, Imagenet classification with deep convolutional neural networks, *Adv. Neural Inf. Process. Syst.*, **25** (2012), 1097–1105.
43. Y. LeCun, L. Bottou, Y. Bengio, P. Haffner, Gradient-based learning applied to document recognition, *Proceed. IEEE*, **86** (1998), 2278–2324.
44. C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, Z. Wojna, Rethinking the inception architecture for computer vision, in *Proceedings of the IEEE conference on computer vision and pattern recognition*, (2016), 2818–2826.
45. C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, et al., Going deeper with convolutions, in *Proceedings of the IEEE conference on computer vision and pattern recognition*, (2015), 1–9.
46. H. C. Shin, H. R. Roth, M. Gao, L. Lu, Z. Xu, I. Nogues, et al., Deep convolutional neural networks for computer-aided detection: Cnn architectures, dataset characteristics and transfer learning, *IEEE Trans. Med. Imaging*, **35** (2016), 1285–1298.
47. A. Kumar, J. Kim, D. Lyndon, M. Fulham, D. Feng, An ensemble of fine-tuned convolutional neural networks for medical image classification, *IEEE J. Biomed. Health Inf.*, **21** (2017), 31–40.
48. C. Szegedy, S. Ioffe, V. Vanhoucke, A. A. Alemi, Inception-v4, inception-resnet and the impact of residual connections on learning, in *Thirty-first AAAI conference on artificial intelligence*, (2017).
49. K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in *Proceedings of the IEEE conference on computer vision and pattern recognition*, (2016), 770–778.
50. B. Zoph, Q. V. Le, Neural architecture search with reinforcement learning, preprint, arXiv:1611.01578.
51. D. T. Bui, T. D. Tran, T. T. Nguyen, Q. L. Tran, D. V. Nguyen, Aerial image semantic segmentation using neural search network architecture, in *International Conference on Multi-disciplinary Trends in Artificial Intelligence*, (2018), 113–124.
52. S. Sabzi, R. Pourdarbani, D. Kalantari, T. Panagopoulos, Designing a fruit identification algorithm in orchard conditions to develop robots using video processing and majority voting based on hybrid artificial neural network, *Appl. Sci.*, **10** (2020), 383.
53. N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, R. Salakhutdinov, Dropout: a simple way to prevent neural networks from overfitting. *J. Mach. Learn. Res.*, **15** (2014), 1929–1958.
54. A. Moayedikia, K. L. Ong, Y. L. Boo, W. Yeoh, R. Jensen, Feature selection for high dimensional imbalanced class data using harmony search, *Eng. Appl. Artif. Intell.*, **57** (2017), 38–49.
55. D. M. W. Powers, Evaluation: from precision, recall and F-measure to Roc, informedness, markedness & correlation, preprint, arXiv:2010.16061.
56. U. Raghavendra, H. Fujita, S. V. Bhandary, A. Gudigar, J. H. Tan, U. R. Acharya, Deep convolution neural network for accurate diagnosis of glaucoma using digital fundus images, *Inf. Sci.*, **441** (2018), 41–49.

57. I. Memon, A. A. Ursani, M. A. Bohyo, R. Chandio, Automated diagnosis of glaucoma using deep learning architecture, *Eng. Sci. Technol. Res. J.*, **3** (2019), 58–62.

58. M. N. Bajwa, M. I. Malik, S. A. Siddiqui, A. Dengel, F. Shafait, W. Neumeier, et al., Two-stage framework for optic disc localization and glaucoma classification in retinal fundus images using deep learning, *BMC Med. Inf. Decis. Making*, **19** (2019), 136.

59. A. Cerentinia, D. Welfera, M. C. dOrnellasa, C. J. P. Haygertb, G. N. Dottob, Automatic identification of glaucoma using deep learning methods, in *MEDINFO 2017: Precision Healthcare Through Informatics: Proceedings of the 16th World Congress on Medical and Health Informatics*, (2018).

60. B. A. Bander, W. A. Nuaimy, M. A. A. Taee, Y. Zheng, Automated glaucoma diagnosis using deep learning approach, in *2017 14th International Multi-Conference on Systems, Signals & Devices (SSD)*, IEEE, 2017.

61. M. Christopher, K. Nakahara, C. Bowd, J. A. Proudfoot, A. Belghith, M. H. Goldbaum, et al., Effects of study population, labeling and training on glaucoma detection using deep learning algorithms, *Transl. Vision Sci. Technol.*, **9** (2020), 27.

62. J. M. Ahn, S. Kim, K. S. Ahn, S. H. Cho, K. B. Lee, U. S. Kim, A deep learning model for the detection of both advanced and early glaucoma using fundus photography, *Plos one*, **13** (2018), e0207982.

63. S. Serte, A. Serener, Graph-based saliency and ensembles of convolutional neural networks for glaucoma detection, *IET Image Process.*, **15** (2021), 797–804.

AIMS Press