Journal of Scientific & Industrial Research
Vol. 84, January 2025, pp. 24-35
DOI: 10.56042/jsir.v84i1.5976

# An Ensemble Stacked Bi-LSTM with ResNet50 Method for Glaucoma Classification in IoT Framework

Sudeshna Pattanaik[1], Subhasikta Behera[3], Santosh Kumar Majhi[4]*, Rosy Pradhan[2]* & Pratyusa Dwibedy[3]

[1]Department of Computer Science & Engineering, [2]Department of Electrical Engineering, Veer Surendra Sai University of Technology, Burla 768 018, Odisha, India

[3]Department of Computer Science & Engineering, National Institute of Technology, Rourkela 769 008, Odisha, India

[4]Department of Computer Science and Information Technology, Guru Ghasidas Viswavidyalaya, Bilaspur 495 009, Chhattisgarh, India

*Received 04 October 2023; revised 13 June 2024; accepted 06 November 2024*

Rural areas in India face significant healthcare challenges, particularly in managing diabetic complications such as glaucoma due to the lack of timely medical facilities. This study proposes an IoT-based healthcare framework designed to connect rural populations with distant healthcare units, enabling medical professionals to provide necessary interventions promptly. The framework employs an ensemble learning-based Bidirectional Long Short-Term Memory (Bi-LSTM) architecture integrated with ResNet50 for glaucoma classification and detection. The methodology involves pre-processing input images, extracting features, and balancing the dataset using the Synthetic Minority Oversampling Techniques (SMOTE). The balanced dataset is then fed into the model, and the results are classified using a sigmoid function. The framework was validated on four datasets such as ACRIMA, Fundus, ORIGA, and Retinal image datasets. Key findings demonstrate that the proposed model achieves superior performance compared to other models for datasets considered, as evidenced by metrics for ACRIMA datasets such as precision (97%), specificity (99%), accuracy (99%), AUC (97%), recall (97%), and F1-score (97%). For fundus dataset, it obtains accuracy of 99%, precision of 92%, recall of 96%, specificity of 94%, F1-score of 95% and AUC of 90%; accuracy of 99%, precision of 95%, recall of 97%, specificity of 93%, F1-score of 94% and AUC of 88% for ORIGA dataset. For retinal datasets, it yields 97% of accuracy, 93% of precision, 93% of recall, 98% of specificity, 93% of F1-score and AUC of 93%. The study's uniqueness lies in its practical utility for addressing healthcare disparities in rural areas through IoT and machine learning, offering promising solutions for real-world applications.

**Keywords:** Artificial intelligence, Bi-directional LSTM, Data balancing, Image pre-processing, IoT healthcare

## Introduction

Glaucoma is termed as a disease where an ocular nerve and visual field is damaged, mostly due to the elevation of intraocular pressure. It is mostly caused by the inability of aqueous humour to flow out of the trabecular meshwork. The most predominant types of glaucoma are categorized into angle-closure and open-angle.[1] In open-angle type, an aqueous humour can access the spongy tissue called trabecular meshwork and increase in pressure obtains from increased resistance to outflow in the open-angle itself. In angle-closure glaucoma, the trabecular meshwork is blocked via iris, leading to high intraocular pressure.[2] Glaucoma is generally characterised by increased intraocular pressure inside the eye, thinning of Retinal Nerve Fibre Layer (RNFL) resulting in expansion ratio of Cup-to-Disk. Glaucoma has been analogous to reduced life quality and mental condition.[3] It is a genetic disorder and affects patients in the later stages of their lives elderly individuals over 60 years tend to be highly vulnerable to glaucoma.[4] According to statistics, this disease impacts around 111.8 million lives across the globe by 2040.(5) It has also been estimated that Asian countries would comprise 47% of all cases of glaucoma. Diagnosis usually takes place through tonometry tests, visual field tests or through imaging tests of the optic nerve. Glaucoma is notorious for exhibiting symptoms only at the advanced phases, hence the name “Silent Thief of Sight”.[6] Hence, development in diagnostic tools for detecting glaucoma in early stage is necessary.[7,8]

Detection of glaucoma is complicated as elevated intraocular pressure is not a reliable indicator of glaucoma, since patients suffering from glaucoma

*Authors for Correspondence
E-mail: smajhi_cse@ieee.org, rosypradhan_ee@vssut.ac.in

may have normal eye pressure.[9] In clinical practice, glaucoma diagnosis and assessment are carried out through manual methods. A detailed study about glaucoma requires a comprehensive medical history, field of vision, photography of the fundus and optic nerve imaging tests. Consequently, it becomes very time-consuming and requires specialized knowledge. However, machine learning-based technology and deep learning methods have shown promising potential in automating the manual methods to a significant extent. Recent advances in deep learning have made it possible to detect glaucoma only through images.[10]

In literature some studies are carried out for glaucoma classification and detection. Detecting glaucoma only from medical images represents certain challenges. Removal of features from colour images is considered difficult due to artery interweavement. Localizing the affected Optic Disc (OD) by using the size of the cup is labour intensive and time-consuming. The commonly used approaches are texture-based which conduct classification of non-OD regions and OD regions pixel-wise.[11] Therefore, advanced approaches need to be looked at to classify glaucoma from images effectively. Different deep learning methods such as Deep CNN, LSTM, RNN, transfer learning are used for glaucoma classification and detection.[12–20] Also the methods like evolving ensemble model and finite element methods are used for glaucoma identification.[21,22]

In the presented work, the authors have proposed a new Bi-LSTM with a transfer learning module, where multiple convolutions are used along with various convolution kernel sizes for extracting image features. A novel deep learning technique is utilized to classify glaucoma on tinge fundus images with the help of an ensemble-based technique. First, the image data set is pre-processed by incorporating mechanisms such as data augmentation, normalization, transformation, and sampling for detecting glaucoma. Second, a Convolution Neural Network (CNN) framework is utilized in order to extract features that are relevant. Next, the features that are extracted are fed to ensemble-based models for classification. Finally, various performance metrics are used for performance comparison of the implemented approach with the existing approaches, and datasets are considered to appraise the robustness of the current research. Four different open-source datasets, namely ACRIMA, Fundus, ORIGA, and Retinal are used for the simulations, which contain normal as well as glaucoma images. Analysis of this study's findings can be summarised as follows:

- This paper focuses on development of Internet of things (IoT) based assistive healthcare systems for robust and responsive medical treatment and prediction of the glaucoma disease.
- Ensemble Stacked Bi-LSTM with ResNet50 and transfer learning has been utilized to classify images of glaucoma.
- ACRIMA, FUNDUS, ORIGA and RETINAL standard image datasets were deployed in order to evaluate the performance of an implemented model.

## Material and Methods

This section provides detailed information on the proposed methodology. It begins with data capture using an IoT-based contact lens sensor. The pre-processing phase involves processing the input images to eliminate distortion and noise, ensuring that they are standardized, which enhances quality and consistency for subsequent analysis. To address class imbalance issues, the SMOTE is employed.[23] Next, feature extraction is performed using a CNN, which automatically learns and identifies essential features relevant to glaucoma classification. Finally, the extracted features from the CNN are fed into a classifier to determine whether an image is normal or glaucomatous. Every step is elucidated in the consecutive sub-sections.

### Description of Dataset

In this paper, to validate the implemented method, four open accessible databases are taken. The specifications of every dataset were recapitulated in Table 1.

Several publicly known databases which is shown in Table 1 with labeled glaucoma images can be used to assess glaucoma classification performance, especially through standardized and automated methodologies such as those employing deep learning approaches. This study utilized four datasets: ACRIMA, FUNDUS, ORIGA, and RETINAL. The ACRIMA dataset[24] consists of 309 normal and 396

Table 1 — Various dataset description related to retinal database

| Dataset | Image | | Resolution | Format |
|---|---|---|---|---|
| | Normal | Glaucoma | | |
| ACRIMA[24] | 309 | 396 | 1054 × 1054 | JPG |
| FUNDUS[24] | 482 | 202 | 3072 × 2048 | JPG |
| ORIGA[24] | 482 | 168 | 2896 × 1984 | JPG |
| RETINAL[25] | 300 | 255 | 2464 × 1632 | JPG |

glaucomatous images acquired with the consent of patient. These images were seized using the Topcon TRC retinal camera and IMAGEnet® capture system and annotated by two glaucoma experts with eight years of experience. The specialists selected all patients based on their criteria and clinical findings during assessments. Most images in this database were taken from both the left and right eyes, dilated, and centered on the optical disc. A few images were discarded due to artifacts, noise, and poor contrast. Again, The FUNDUS dataset[24] is split into training and validation sets, individually further organized into positive and negative classes. This dataset possesses a total of 684 fundus images. The ORIGA dataset[24] is an open-source collection with 650 retinal images analyzed by researchers from the Singapore Eye Research Institute (SERI). Another is the RETINAL dataset[25] which consists of 300 healthy images and 255 glaucomatous images collected from the Aravind Eye Hospital in Madurai. In total, this study utilized 2,594 images, with 1,573 normal and 1,021 glaucomatous instances. Of these, 1,815 images were used for training and 779 for validation to accurately evaluate the model.

**Data Pre-processing**

Healthy and glaucoma images are presented in Fig. 1. This section describes the pre-processing techniques used in this paper. It begins with data augmentation techniques, where SMOTE is used to increase the number of data points.[26] The augmented data is then fed into pertained architectures for feature extraction, followed by classification. The detailed steps are provided below.

***Data Augmentation***

To enhance resolution of the image and to emphasize the necessary features, the pre-processing of colour images which are taken as input for detection of glaucoma.[27] Firstly, to expand the size of dataset artificially, an image data generator is utilized by generating the updated form of the dataset's image. It transforms training dataset images to the equivalent class. Transforms consists of different operations such as zooming, shifting, shearing, clipping and more (Fig. 2). This study initially focuses on preferred augmentation methods of data such as fill mode random shift, dropout, random rotation etc.

***SMOTE***

SMOTE[28] is a widely used oversampling method to deal with the imbalance problem. SMOTE involves the following procedure:

1. K-nearest neighbours for each minority class variable is obtained.

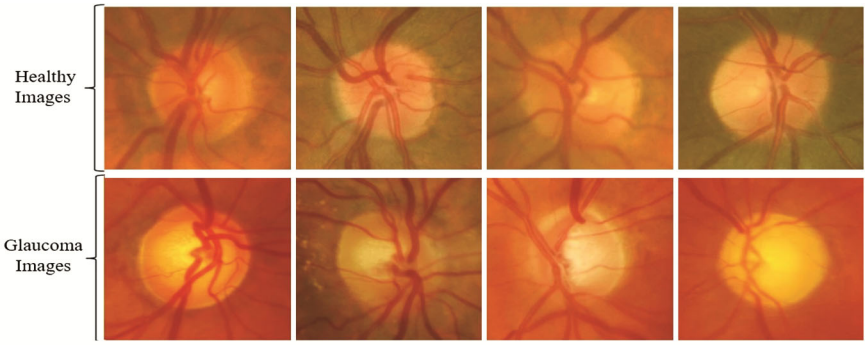


Fig. 1 — Healthy and glaucoma images

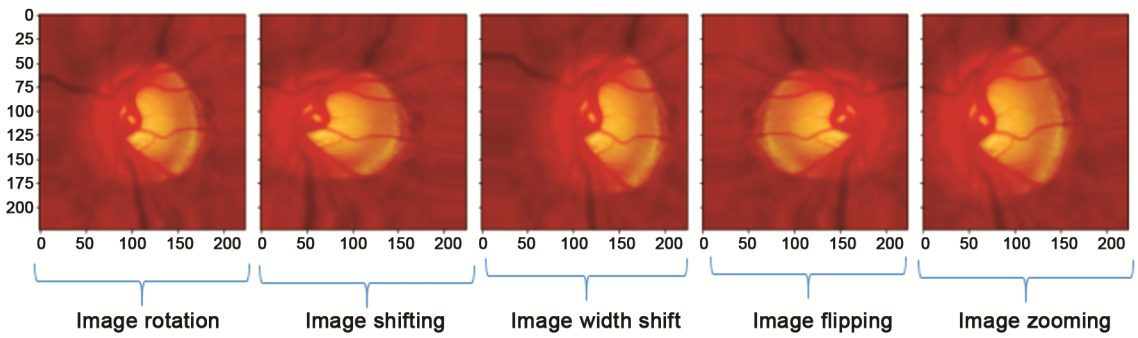


Fig. 2 — Data augmentation techniques

2. N-nearest neighbours for each minority class variable are randomly sampled.
3. Synthesized points (variables) are randomly selected from the line segments joining the N-nearest neighbours for each minority variable.

**Pre-trained CNN Architecture for Feature Extraction**

Five different deep learning architectures i.e., VGG16, ResNet50, GoogleNet, MobileNet, and DenseNet have been used in the work.

**Classification**

The classification of fundus[28] glaucoma images is done with a deep ensemble model. This work uses a total of 2594 images to train the model. This image is categorically bifurcated i.e., positive and negative. To accurately evaluate the model, the dataset is partitioned into a train (70%) and test (30%) respectively. In this, the training dataset is used for training the model by providing different types of images and the testing dataset is utilized for validating the model. At last, different classification criterions are utilized for performance measurement of distinct models. The following section briefly illustrates the proposed model.

***Proposed Framework***

The proposed ensemble framework consists of Bi-LSTM with ResNet50. Ensemble modelling is a process, where multiple weak learner models were formulated to generate an output, either by incorporating various modelling algorithms or by using training data sets. The ensemble learning model then combines the results of all base models and generates one final prediction for the testing data. A Bi-LSTM architecture is a sequential processing model that contains two LSTM models in which the first LSTM model takes input in the onward direction and further several models in a direction backwardly. The Bi-LSTMs raise the quantity of available information effectively on the network, with improvement in the available content to the algorithm. The training of proposed model is performed by utilizing the input parameters for epochs of 50 and 10 as batch size and rate of learning as 0.0001. The function named Relu which is a function of activation is used in the layers in between the layers of input and output called hidden layer and the function of activation called "sigmoid" is enforced in the layer of output in order to classify the output obtained.

The process, as illustrated in Fig. 3, includes each component from the pre-processing phase to the fully connected layer. In this framework, the transfer learning model ResNet50 is initialized with a Bi-LSTM, demonstrating the integration of advanced deep learning techniques to enhance model performance. In the first step, the image of original version is considered for purpose of prediction in order to feed the model. Pre-processing of input images are performed as the formats of images are appropriate. In the phase of pre-processing, techniques for augmentation of data such as parameter of rotational range, shear range etc are utilized for transforming an image data. As an image sizes are not of equal size, the technique called normalization that

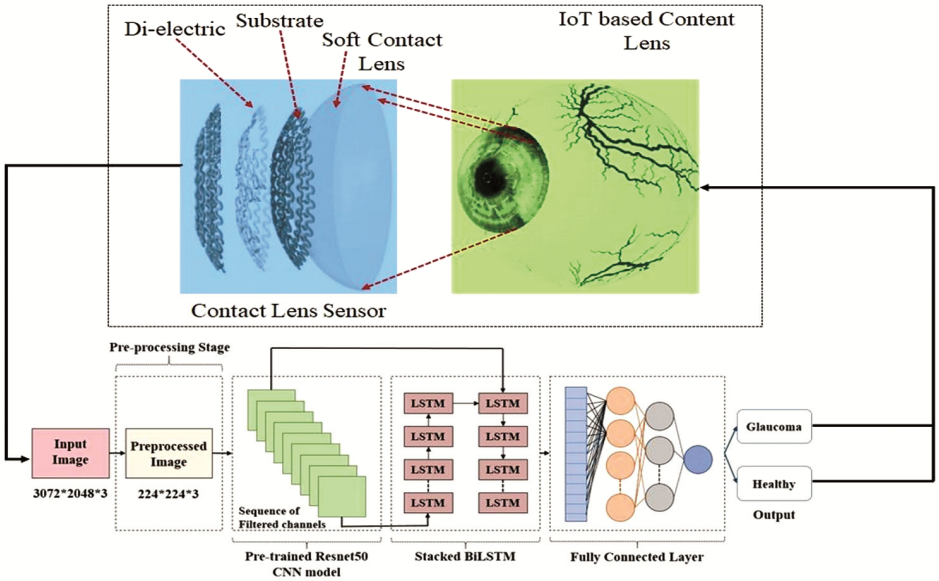


Fig. 3 — Proposed ensemble-based CNN architecture

is utilized in order to make each image of same size. Next to the pre-processing phase of an image, the technique called SMOTE is adapted for an image data re-sampling as the information provided are imbalanced. In order to stabilize the information provided as data, a technique is being utilized called a method of oversampling in which data were randomly produced which is synthetic data for minority class. After this phase of data balancing, the data is bifurcated in order to validate the designed model. Thereafter, split image is provided to the proposed architecture that comprises of Bi-LSTM along with ResNet50. Firstly, the image is pre-processed and put into the model called resnet50 with parameters of weight and an outcome obtained by using this framework is fed as input to the Bi-LSTM architecture. After, both the input and output is merged for this framework and a function called "Sigmoid" is being used in order to predict the outcome. For fine-tuning of the model, a learning rate of 0.0001 is taken with loss function binary cross-entropy. Then the model fits with 50 epochs and a batch size of 10.

***Algorithm for Proposed Architecture***

Proposed algorithm for the framework is illustrated in a sequel.

*Step - 1: Given the image data into the proposed Model*

i) At the very first, the input layer is convoluted and mapping of features from an initial layer is achieved using the convolution kernels. From the given function feature map size can be reclaimed as in Eq. (1).

$$n_f = \frac{n_i + 2_p - f}{s} \quad \dots (1)$$

where, $n_i$ implies the input size, p refers to padding representing kernel size and s shows stride.

ii) Using the convolution process, it extracts the characteristics of the given image which can be done through the function given in Eq. (2).

$$\propto^1 = \delta\,(w^t \alpha^{t-1} + b^t) \quad \dots (2)$$

where, $\propto^1$ denotes the result of $t_{th}$ layer,$\alpha^{t-1}$ denotes outcome, and b denotes bias of convolutional layer and δ denotes the activation function.

iii) For classification, the image is fed into resnet50 architecture using Eq. (3).

$$y = F(x, w) + x \quad \dots (3)$$

where, x, w, y means input, weight and output of resnet model.

*Step - 2: Forward Propagation of Proposed Model*

In this approach, the i(t) is input, f(t) is forget, o(t) is output gate, and c(t) refers to cell state vector, respectively. Followed by $W_{p1}$, $W_{p2}$, $W_{p3}$ refers to the links W, $W_{reci}$ and $b_i$, i = 1, 2, 3, 4 are the input links, recurrent link and bias respectively. g(•) and h(•) represents relu function, o represents element wise product and shows sigmoid function.

Therefore, an LSTM is employed in order to compute every word and the final word of a sentence is embedded with y(m) function as a correct vector for the whole sentence. The following are the LSTM method's forward pass:

$$y_g(t) = g(w_4 l(t) + w_{rec4} y(t-1) + b_4) \quad \dots (4)$$

$$i(t) = \sigma(w_3 l(t) + w_{rec3} y(t-1) + w_{p3} c(t-1) + b_3) \quad \dots (5)$$

$$f(t) = \sigma(w_2 l(t) + w_{rec2} y(t-1) + w_{p2} c(t-1) + b_2) \quad \dots (6)$$

$$c(t) = f(t)\, o\, c(t-1) + i(t) o y_g(t) \quad \dots (7)$$

$$o(t) = \sigma(w_2 l(t) + w_{rec1} y(t-1) + w_{p1} c(t-1) + b_1) \quad \dots (8)$$

$$c(t) = o(t)\, o\, h(t) \quad \dots (9)$$

$$y(t) = o(t)\, o\, h(c(t)) \quad \dots (10)$$

*Step -3: Backward propagation for optimise and get the result from the model*

i) While training the image in forward propagation data, the result is not optimised due to the uneven initialization of the parameters. Consequently, it generates a few results that are hard to maximize the output of the model. Therefore, the model is optimized using Eq. (11).

ii) $$L(A) = \min \prod_{q=1}^{M} \{-\log \prod_{q=1}^{M} p(D_r^+)\} = \min \sum_{q=1}^{M} l_q(\wedge) \quad \dots (11)$$

iii) Following the optimization, the gradient or cost function of our model is estimated using Eq. (12).

$$\nabla L(\wedge) = -\sum_{q=1}^{M} \sum_{p=1}^{M} \sum_{\tau=0}^{T} \alpha_{p,q} \frac{\partial \Delta_{p,q,\tau}}{\partial \wedge} \quad \dots (12)$$

iv) Then a sigmoid function is used in the last layer to predict the output using Eq. (13).

$$p(x_i) = \frac{1}{\sum_{i=1}^{n} (1 + e^{-x_i})} \quad \dots (13)$$

Algorithm 1 presents the steps for the proposed framework.

**Algorithm 1:**

1. *Input the image data into the proposed Model*
   Convolute the input layer and feature map the initial layer using Eq. (1);

Extract the features from images using Eq. (2);
Classify the images using Eq. (3);

2. *Forward Propagation of Proposed Model*
   Apply the LSTM for computing using Eq. (4);
   Evaluate i(t), f(t), o(t), c(t) using Eq. (5), (6), (8), and (7) respectively;
3. *Backward Propagation for Optimise and Get the Result from the Model*
   Optimise the model for Backward propagation using Eq. (11);
   Estimate the gradient or cost function of the model using Eq. (12);
   Generate the output using a sigmoid function (Eq. (13));

Recently, IoT technology is growing rapidly that provides immense potentiality in various life disciplines such as governmental control, border management, smart traffic management, and eHealth Care, especially in the proactive personal eHealth system. We extended the Proposed IoT Framework for Healthcare system, shown in Fig. 3, which broadcast the medical data from various contact lens sensors through a fog network via communication technology such as Bluetooth Low Energy that systematically collects the clinical care of glaucoma data that feedback to the input of the proposed ensemble learning of the proposed work.

#### Implementation Details of the IoT Framework

To implement the IoT framework for Glaucoma detection, the system will utilize various IoT devices, including sensors and cameras, to capture retinal images and other health data from patients. This data will be processed locally on microcontrollers before being transmitted to central servers for in-depth analysis. Communication between devices will rely on protocols such as MQTT (Message Queuing Telemetry Transport) or HTTP/HTTPS, with MQTT chosen for its efficiency in low-bandwidth and high-latency environments, making it suitable for rural areas with limited internet connectivity. To ensure robust data security, the framework will employ end-to-end encryption to protect data during transmission, secure authentication methods to control access to devices and servers, and regular updates to firmware and software to address potential vulnerabilities. Additionally, data anonymization techniques will be used to safeguard patient identities, ensuring compliance with healthcare regulations such as HIPAA (Health Insurance Portability and Accountability Act) and GDPR (General Data Protection Regulation). These measures collectively ensure the reliability, security, and effectiveness of the IoT-based healthcare system.

## Result and Discussion

The implementation of a designed model is carried out using Python where Keras is used and tensor flow is used in backend. The simulations are carried out on Linux LTS 18.04. Various kinds of coloured retinal image consisting of training and validation set that is considered for classifying glaucoma. The metrics consisting of precision, accuracy, recall, AUC, F1-score and specificity were taken into consideration to appraise performance of an implemented method.

#### Performance Metrics

This section outlines the various performance metrics used to evaluate the proposed model. These metrics provide a comprehensive assessment of the model's effectiveness in classifying glaucoma from non-glaucoma images. Key metrics include accuracy, which measures the overall correctness of the model's predictions, precision, which evaluates the model's ability to correctly identify positive cases, and recall, which assesses the model's capability to detect all relevant cases. Additionally, the F1-score is used to balance precision and recall, while specificity evaluates the model's ability to correctly identify negative cases. Finally, the area under the curve (AUC) is used to assess the model's ability to distinguish between classes across various threshold settings. Together, these metrics offer a detailed understanding of the model's performance and reliability.

##### *Accuracy*

It is defined as barometer for quality values and termed as AM (Arithmetic Mean) of weight. It is defined as reverse of the precision value. It is described as number of correct predictions made by the model over all. It is a good measure when the target variable is balanced. Evaluation of accuracy is based on the given formula in Eq. (14).

$$\text{Accuracy} = \frac{\text{TN}+\text{TP}}{\text{TP}+\text{FP}+\text{TN}+\text{FN}} \quad \ldots (14)$$

where, TN = True Negative, TP = True Positive, FP = False Positive, FN = False Negative

##### *Precision*

Precision refers to percentage of relevant results. It refers to the count of actual positives over the sum

count of positive predictions. It can be represented as in Eq. (15).

$$\text{Precision} = \frac{\text{TP}}{\text{TP+FP}} \quad \ldots (15)$$

where, TN = True Negative, TP = True Positive, FP = False Positive, FN = False Negative

***Recall***

It is defined as proportion of true negative where it got predicted as negative or true negative. This proportion is also called a false-positive rate. It can be represented as in Eq. (16).

$$\text{Recall} = \frac{\text{TP}}{\text{TP+FN}} \quad \ldots (16)$$

where, TN = True Negative, TP = True Positive, FP = False Positive, FN = False Negative

***F1-Score***

F1 score considers both recall and precision. It shows the balance existing between recall and precision and produces a better computation for incorrectly classified cases as compared to the metric of accuracy.

$$\text{F1 Score} = 2 \times \frac{(\text{Precision} * \text{Recall})}{(\text{Precisio+Recall})} \quad \ldots (17)$$

***AUC Curve***

AUC is applied in classification analysis for determining the model which performs the best class prediction amongst a set of considered models. The ROC curves are popular examples of AUC. AUC ranges in value from 0 to 1. A model with AUC 0 performs predictions which are 100% wrong and models with AUC 1 perform predictions which are 100% correct.

## Discussion

Training the colour fundus images using an integrated CNN and Recurrent Neural Network (RNN) has an ability to extract features both spatially and temporally thereby improving the designed model's performance. In recent research, four various kinds of datasets are taken into consideration for model training. The hybrid Bi-LSTM and ResNet50 architecture give significantly high accuracy as compared to other models whereas the Bi-LSTM and VGG-16 model gives moderate accuracy. ResNet-50 is an optimal choice for real-time applications in resource-constrained IoT environments due to its balance of depth and efficiency. It offers strong feature extraction capabilities and good classification accuracy while ensuring shorter training times and reduced energy consumption, which is ideal for battery-powered devices. ResNet-50's architecture minimizes the risk of over fitting and allows for quick inference times, making it easy to deploy and integrate into existing systems. In contrast, deeper networks like ResNet-101 and ResNet-152 have higher computational costs and longer training times, making them less practical for such environments. The concept provides solid evidence that our proposed model will categorize the glaucoma image. With respect to classification the proposed Bi-LSTM with ResNet50 performs better than others. Comparative analysis of the implemented techniques against existing methods, using different metrics across various datasets is presented in Table 2. The powerful deep learning techniques have the ability to categorize the glaucoma on the basis of colour fundus image. However, they are not improved yet for low precision and false-positive rate of high proportion. The drawback is surpassed by improving the results of classification by using the designed model as compared to other existing models. For classification of images the approach of transfer learning is used for all datasets. Methods of transfer learning are selected as it reduces high-rise in performance significantly and statistically as compared to various other models of CNN. Dataset considered is small so naive models didn't performed well. A technique of oversampling is utilized for generating minority classes randomly in order to balance all the images. As the imbalanced data influence the precision and recall value. This problem of imbalanced image is discarded by balancing the data and bifurcated into a train and test set of data. The training and testing set of data are considered to train and evaluate model's performance respectively. In the below Fig. 4 it shows the flow chart of the proposed method. Although, the proposed method achieves high accuracy, its implementation, especially in rural areas, presents several potential challenges. For instance, limitations in infrastructure, unreliable internet connectivity, and varying levels of user acceptance can affect the implementation and effectiveness of the IoT-based healthcare framework. Addressing these challenges requires specific strategies, such as using solar-powered devices, deploying local edge servers, employing data compression techniques, and engaging with local communities to build trust and ensure successful adoption.

Table 2 — Comparative analysis of implemented and techniques existing with respect to different metrics with distinct datasets

| Dataset | Architecture | Accuracy (%) | Precision (%) | Recall (%) | Specificity (%) | F1-Score (%) | AUC (%) |
|---|---|---|---|---|---|---|---|
| ACRIMA | Xception[27] | 96 | 95 | 94 | 85 | 93 | 96 |
| | DeepLabv3 with MobileNet[28] | 98 | 85 | 86 | 95 | 94 | 99 |
| | Deep CNN [29] | 96 | 97 | 92 | 97 | 96 | 96 |
| | Deep Learning Technique [30] | 54 | 60 | 68 | 72 | 31 | 73 |
| | **Proposed Method** | **99** | **97** | **97** | **99** | **97** | **97** |
| FUNDUS | ResNet and GoogleNet[31] | 90 | 65 | 42 | 94 | 43 | 84 |
| | Deep Belief Network (DBN)[32] | 99 | 84 | 83 | 98 | 84 | 89 |
| | Ensemble of 4 CNNs [33] | 97 | 85 | 89 | 83 | 84 | 98 |
| | **Proposed Method** | **99** | **92** | **96** | **94** | **95** | **90** |
| ORIGA | Xception | 87 | 82 | 84 | 81 | 93 | 83 |
| | CNN | 90 | 83 | 86 | 91 | 88 | 82 |
| | Ensemble of 4 CNNs | 95 | 87 | 86 | 92 | 95 | 91 |
| | **Proposed Method** | **99** | **95** | **97** | **93** | **94** | **88** |
| Retinal | Support Vector Machine (SVM)[34] | 87 | 88 | 84 | 89 | 85 | 87 |
| | Semi-supervised transfer learning for CNN | 92 | 86 | 84 | 82 | 90 | 91 |
| | Feed-Forward Neural Network[34] | 92 | 70 | 75 | 85 | 84 | 90 |
| | Attention-CNN[34] | 95 | 80 | 78 | 86 | 84 | 88 |
| | **Proposed Method** | **97** | **93** | **93** | **98** | **93** | **93** |

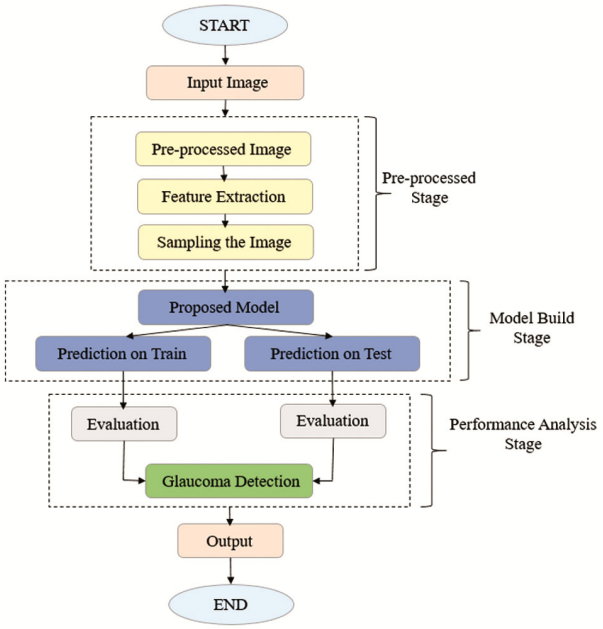


Fig. 4 — Flow chart of proposed model

The confusion matrix for different datasets has been given in Fig. 5. The confusion matrix for ACRIMA, FUNDUS, ORIGIA, and RETINAL has given in Fig. 5(a), Fig. 5(b), Fig. 5(c), Fig. 5(d) respectively. The proposed architecture was tested using four separate datasets: ACRIMA[24], Fundus[24], ORIGA[24], and Retinal.[25] Out of a total of 2,594 images, approximately 400 images were correctly identified as healthy, and 378 images were correctly classified as glaucomatous. However, 72 images were misclassified, indicating areas for further improvement in classification accuracy.

The accuracy versus epoch graph is presented in Fig. 6. The performance of the proposed model across 50 epochs for ACRIMA, FUNDUS, ORIGIA, and RETINAL has given in Fig. 6(a), Fig. 6(b), Fig. 6(c), Fig. 6(d) respectively. This evaluation considers all layers of the model and assesses its performance on both the training and testing datasets. The graph reveals the relationship between accuracy and the number of epochs, showing that a minimum number of epochs can yield better results for the proposed ensemble-based model. Additionally, the model's validation process addresses potential over fitting and under fitting conditions, ensuring robust performance.

The accuracy comparison for these datasets is provided in Table 3, showcasing the model's effectiveness and reliability in accurately diagnosing glaucoma across different conditions. This analysis underscores the model's versatility and consistent performance, reinforcing its suitability for diverse datasets. Furthermore, the ROC (Receiver Operating Characteristic) curves for glaucoma classification, shown in Fig. 7, illustrate the model's performance across four distinct datasets. These curves indicate that the AUC is consistently high and nearly identical for each dataset, reflecting the model's robust ability to distinguish between classes across various data sources.

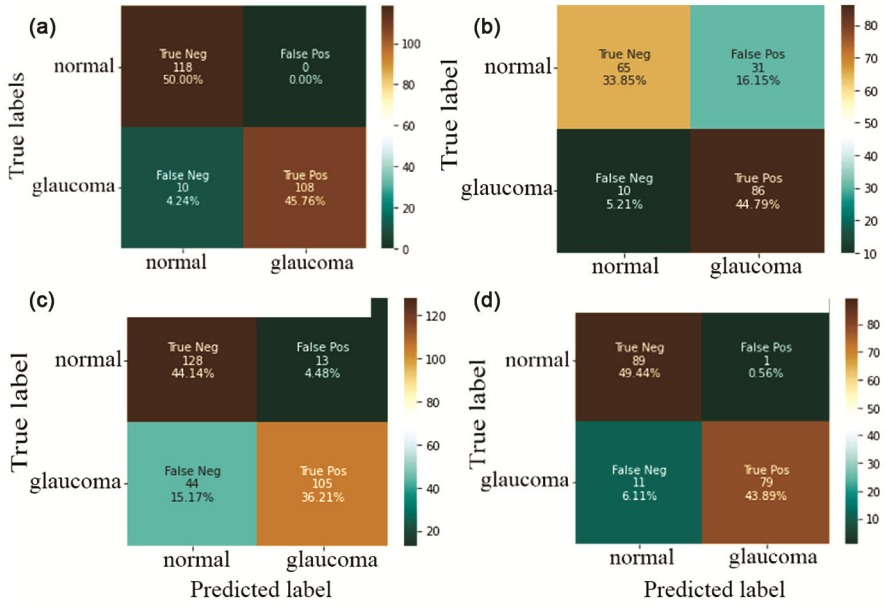


Fig. 5 — Confusion matrix of different datasets: (a) ACRIMA, (b) FUNDUS, (c) ORIGA, (d) RETINAL

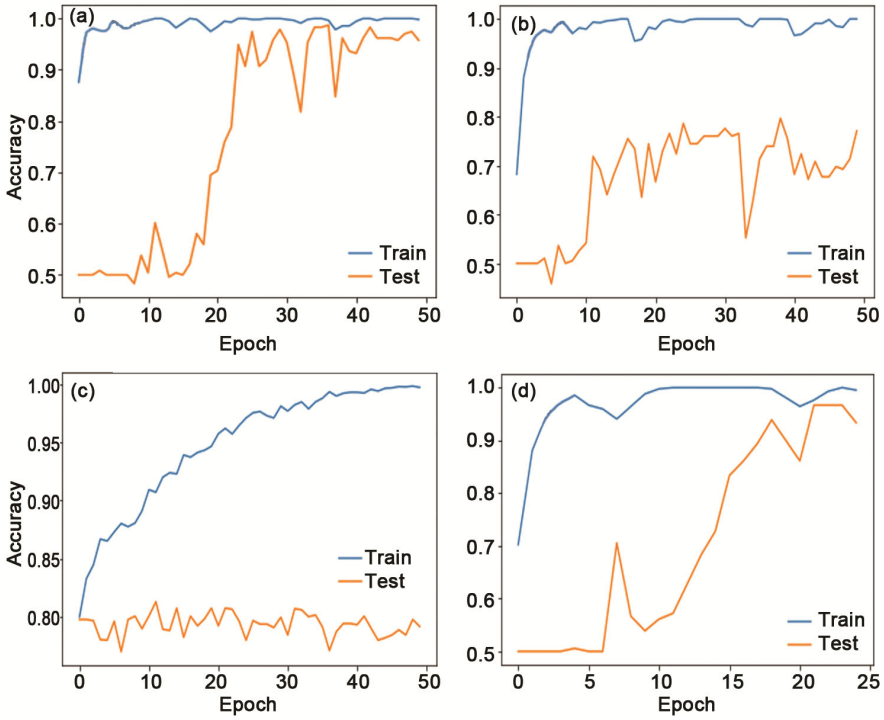


Fig. 6 — Accuracy versus epoch graph: (a) ACRIMA, (b) FUNDUS, (c) ORIGA, (d) RETINAL

**Comparative Study of Glaucoma Images**

The purpose of glaucoma detection is assessed in comparison to different existing methods and various methods based on an ensemble technique. The comparison of performance for the proposed model with various ensemble techniques for precision, accuracy, AUC, and recall are shown in Table 3. The designed model for classification based on ensemble methods reveals significant values statistically as compared to methodologies already exist for detecting glaucoma. Evidence from the results obtained outlines that the implemented model based on ensemble methods perform better than the existing models. The model's conduct is computed by evaluating a sector below the curve of recipient operating features, sensitivity, accuracy and specificity parameters.

Table 3 — Performance measure of proposed model with different ensemble techniques

| Dataset | Architecture | Accuracy (%) | Precision (%) | Recall (%) | Specificity (%) | F1-Score (%) | AUC (%) |
|---|---|---|---|---|---|---|---|
| ACRIMA | Bi-LSTM + VGG16 | 95 | 93 | 92 | 93 | 91 | 93 |
| | Bi-LSTM + DenseNet | 92 | 92 | 91 | 93 | 92 | 93 |
| | Bi-LSTM + MobileNet | 97 | 79 | 79 | 84 | 79 | 79 |
| | Bi-LSTM + GoogleNet | 94 | 92 | 94 | 93 | 94 | 94 |
| | **Proposed Method** | **99** | **97** | **97** | **99** | **97** | **97** |
| FUNDUS | Bi-LSTM + VGG16 | 86 | 86 | 91 | 87 | 88 | 80 |
| | Bi-LSTM + DenseNet | 90 | 89 | 85 | 87 | 86 | 85 |
| | Bi-LSTM + MobileNet | 95 | 89 | 87 | 92 | 86 | 82 |
| | Bi-LSTM + GoogleNet | 86 | 81 | 82 | 85 | 84 | 82 |
| | **Proposed Method** | **99** | **92** | **96** | **94** | **95** | **90** |
| ORIGA | Bi-LSTM + VGG16 | 88 | 86 | 79 | 87 | 80 | 87 |
| | Bi-LSTM + DenseNet | 87 | 90 | 88 | 91 | 85 | 82 |
| | Bi-LSTM + MobileNet | 96 | 84 | 80 | 87 | 85 | 83 |
| | Bi-LSTM + GoogleNet | 88 | 84 | 80 | 98 | 80 | 82 |
| | **Proposed Method** | **99** | **95** | **97** | **93** | **94** | **88** |
| RETINAL | Bi-LSTM + VGG16 | 95 | 88 | 87 | 95 | 87 | 87 |
| | Bi-LSTM + DenseNet | 95 | 92 | 92 | 97 | 92 | 92 |
| | Bi-LSTM + MobileNet | 92 | 92 | 92 | 96 | 92 | 92 |
| | Bi-LSTM + GoogleNet | 91 | 93 | 93 | 92 | 93 | 93 |
| | **Proposed Method** | **97** | **93** | **93** | **98** | **93** | **93** |

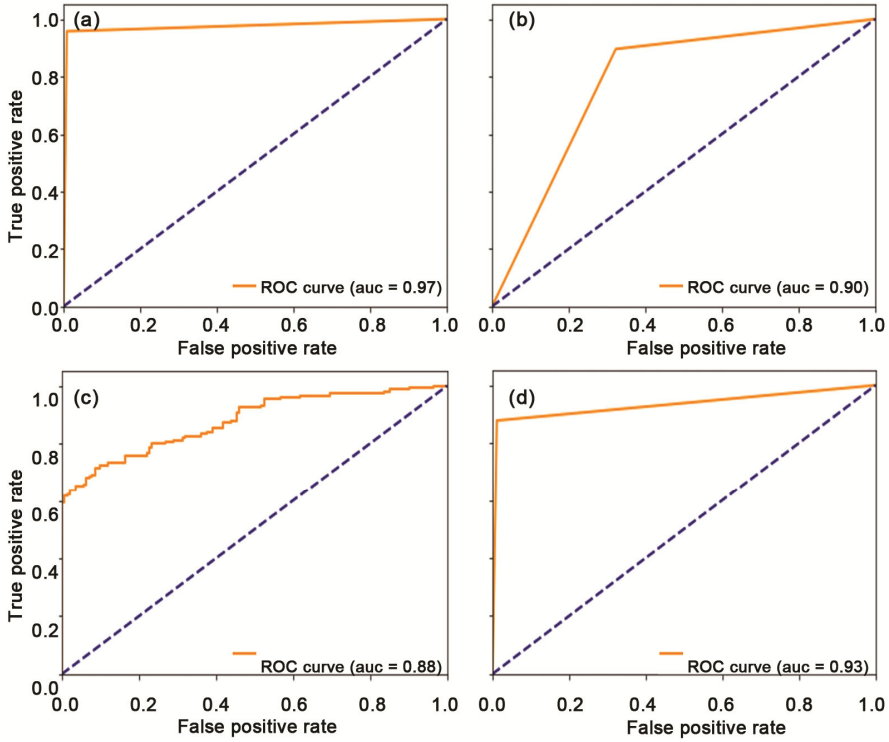


Fig. 7 — ROC curves for glaucoma classification: (a) ACRIMA, (b) FUNDUS, (c) ORIGA, (d) RETINAL

## Conclusions

The study demonstrates that ensemble models combining CNN architectures, particularly Bi-LSTM with ResNet50, effectively classify glaucoma from non-glaucoma fundus images. This model achieved high performance across datasets like ACRIMA, ORIGA, and RETINAL, with average specificity, F1-scores, and accuracy of 85%, 93%, and 99%,

respectively. The ACRIMA database recorded the highest accuracy of 0.99, while similar results were observed in other datasets like ORIGA (0.9967) and RETINAL (0.9775). The study emphasizes the potential of deep learning models to extend beyond glaucoma detection to other ocular diseases. Limitations include the dependence on publicly available datasets, which may not fully reflect real-world diversity. Future work will involve enhancing the IoT-based healthcare framework by using transfer learning and multi-task learning, expanding diagnostic capabilities to other eye conditions. A modular design, data augmentation, and collaboration with healthcare professionals will ensure the models remain flexible, accurate, and relevant for global vision health.

## References


1 Kumar B N, Chauhan R P & Dahiya N, Detection of Glaucoma using image processing techniques: A review, *Proc – 2016 Int Conf Microelectron*, *Comput Commun* (IEEE) 2016, 1–6, doi: 0.1109/MicroCom.2016.7522515.

2 Weinreb R N, Leung C K, Crowston J G, Medeiros F A, Friedman D S, Wiggs J L & Martin K R, Primary open-angle glaucoma, *Nat Rev Dis Primers*, **2(1)** (2016) 1–19, doi: doi.org/10.1038/nrdp.2016.67.

3 Civit-Masot J, Domínguez-Morales M J, Vicente-Díaz S & Civit A, Dual machine-learning system to aid glaucoma diagnosis using disc and cup feature extraction, *IEEE Access*, **8** (2020) 127519–127529, doi: 10.1109/ACCESS.2020.3008539.

4 Shinde R, Glaucoma detection in retinal fundus images using U-Net and supervised machine learning algorithms, *Intell-Based Med*, **5** (2021) 1–15, 100038, doi.org/10.1016/j.ibmed.2021.100038.

5 Deepa N, Esakkirajan S, Keerthiveena B & Dhanalakshmi S B, Automatic diagnosis of glaucoma using ensemble based deep learning model, *Proc - 2021 7th International conference on advanced computing and communication systems* (IEEE) 2021, 536–541, doi: 10.1109/ICACCS51430.2021.9441817.

6 Chen X, Xu Y, Wong D W K, Wong T Y & Liu J, Glaucoma detection based on deep convolutional neural network, In *2015 37th Ann Int Conf IEEE Eng Med Biol Soc* (IEEE) 2015, 715–718, doi: 10.1109/EMBC.2015.7318462.

7 David & Stalin D, Enhanced glaucoma detection using ensemble based CNN and spatially based ellipse fitting curve model, *J J Ambient Intell Humaniz Comput*, **14(4)** (2023) 3303–3314, https:doi.org/10.1007/s12652-021-03467-4.

8 Saxena A, Vyas A, Parashar L & Singh U, A glaucoma detection using convolutional neural network, *2020 Int Conf Electron Sustain Commun Syst* (IEEE) 2020, 815–820, doi: 10.1109/ICESC48915.2020.9155930.

9 Serener A & Serte S, Transfer learning for early and advanced glaucoma detection with convolutional neural networks, *2019 Medical technologies congress* (IEEE) 2019, 1–4, doi: 10.1109/TIPTEKNO.2019.8894965.

10 Gómez-Valverde J J, Antón A, Fatti G, Liefers B, Herranz A, Santos A, Sánchez C I & Ledesma-Carbayo M J, Automatic glaucoma classification using color fundus images based on convolutional neural networks and transfer learning, *Biomed Opt Express*, **10(2)** (2019) 892913, doi: 10.1364/BOE.10.000892.

11 Mansour, Romany F & Al-Marghilnai A, Glaucoma detection using novel perceptron based convolutional multi-layer neural network classification, *Multidimens Syst Signal Process*, **9** (2021) 1–19, doi: 10.1007/s11045-021-00781-0.

12 Zhang L & Chee P L, Intelligent optic disc segmentation using improved particle swarm optimization and evolving ensemble models, *Appl Soft Comput*, **92** (2020) 106328, doi: 10.1007/s11045-021-00781-0.

13 Veena H N, Muruganandham A & Kumaran T S, 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*, **8** (2021) 6187–6198, doi: 10.1016/j.jksuci.2021.02.003.

14 Khan M K & Anwar S, M-Net with Bidirectional ConvLSTM for Cup and Disc Segmentation in Fundus Images, *2020 IEEE-EMBS Conf Biomed Eng Sci*, (IEEE) 2021, 472–476, doi: 10.1109/IECBES48179.2021.9398745.

15 Claro M, Veras R, Santana A, Araujo F, Silva R, Almeida J & Leite D, An hybrid feature space from texture information and transfer learning for glaucoma classification, *J Vis Commun Image Represent*, **64** (2019) 102597, doi: 10.1016/j.jvcir.2019.102597.

16 Dash J & Bhoi N, An unsupervised approach for extraction of blood vessels from fundus images, *J Digit Imaging*, **31(6)** (2018) 857–868, doi: 10.1007/s10278-018-0059-x.

17 Raja H, Akram M U, Shaukat A, Khan S A, Alghamdi N, Khawaja S G & Nazir N, Extraction of retinal layers through convolution neural network (CNN) in an OCT image for glaucoma diagnosis, *J Digit Imaging*, **33(6)** (2020) 1428–1442, doi: 10.1007/s10278-020-00383-5.

18 Almubarak H, Bazi Y & Alajlan N, Two-stage mask-rcnn approach for detecting and segmenting the optic nerve head, optic disc, and optic cup in fundus images*, Appl Sci*, **10(11)** (2020) 3833, doi: 10.3390/app10113833.

19 Aamir M, Irfan M, Ali T, Ali G, Shaf A, Al-Beshri A, Alasbali T & Mahnashi M H, An adoptive threshold-based multi-level deep convolutional neural network for glaucoma eye disease detection and classification, *Diagnostics*, **10(8)** (2020) 602, doi: 10.3390/app10113833.

20 Shoba S G & Therese A B, Detection of glaucoma disease in fundus images based on morphological operation and finite element method, *Biomed Signal Process Control*, **62** (2020), doi: 10.1016/j.bspc.2020.101986.

21 Ajesh F & Ravi R, Hybrid features and optimization-driven recurrent neural network for glaucoma detection, *Int J Imaging Syst Technol*, **30(4)** (2020) 1143–1161, doi: 10.1016/j.bspc.2020.101986.

22 Ajitha S, Akkara J D & Judy M V, Automated Identification of glaucoma from fundus images using deep learning techniques, *Eur J Mol Clin Med*, **7(2)** (2020) 5449–5458, doi: 10.4103/ijo.IJO_92_21.

23 Natarajan D, Sankaralingam E, Balraj K & Karuppusamy S, A deep learning framework for glaucoma detection based on robust optic disc segmentation and transfer learning, *Int J Imaging Syst Technol*, **1** (2021) 230–250, doi: 10.1002/ima.22609.

24 Diaz-Pinto A, Morales S, Naranjo V, Köhler T, Mossi J M & Navea, CNNs for automatic glaucoma assessment using

fundus images: An extensive validation, *Biomed Eng Online*, **18(1)** (2019) 1–19, doi: https:10.1186/s12938-019-0649-y.

25 Sivaswamy J, Krishnadas S R, Joshi G D, Jain M, Ujjwal A, & Drishti S T, Retinal image dataset for optic nerve head (ONH) segmentation, *2014 IEEE 11th Int Symp Biomed Imag* (IEEE) 2014, 53–56.

26 Kirar B S & Agrawal D K, Current research on glaucoma detection using compact variational mode decomposition from fundus images, *Int J Intell Syst*, **12(3)** (2019) 1–10, doi: 10.22266/ijies2019.0630.01.

27 Abbas Q, Glaucoma-deep: detection of glaucoma eye disease on retinal fundus images using deep learning, *Int J Adv Comput Sci Appl*, **8(6)** (2017) 41–45.

28 Gheisari S, Shariflou S, Phu J, Kennedy P J, Agar A, Kalloniatis M & Golzan S M, A combined convolutional and recurrent neural network for enhanced glaucoma detection, *Sci Rep*, **11(1)** (2021) 1–11, doi: 10.1038/s41598-021-81554-4.

29 Elangovan P, Nath M K & Mishra M, Statistical parameters for glaucoma detection from color fundus images, *Procedia Comput Sci*, **171** (2020) 2675–2683, doi: 10.1016/ j.procs.2020.04.290.

30 Sreng S, Maneerat N, Hamamoto K &Win KY,Deep learning for optic disc segmentation and glaucoma diagnosis on retinal images, *Appl Scis*, **10(14)** (2020) 4916, doi: **10.3390/app10144916**.

31 Elangovan P & Nath M K, Glaucoma assessment from color fundus images using convolutional neural network, *Int J Imaging Syst Technol*, **31(2)** (2021) 955–971, doi: 10.1002/ima.22494.

32 Medeiros F A, Jammal A A & Mariottoni E B, Detection of progressive glaucomatous optic nerve damage on fundus photographs with deep learning, *Ophthalmology*, **128(3)** (2020) 383–392, doi: 10.1016/j.ophtha.2020.07.045.

33 Li Z, He Y, Keel S, Meng W R, Chang T & He M, Efficacy of a deep learning system for detecting glaucomatous optic neuropathy based on color fundus photographs, *Ophthalmology*, **125(8)** (2018) 1119–1206, doi: 10.1016/j.ophtha.2018.01.023.

34 Aamir M, Irfan M, Ali T, Ali G, Shaf A, Al-Beshri A, Alasbali T & Mah- nashi M H, An adoptive threshold-based multi-level deep convolutional neural network for glaucoma eye disease detection and classification, *Diagnostics*, **10(8)** (2020) 602, doi: 10.3390/diagnostics10080602.