Grenze International Journal of Engineering and Technology, January Issue

Glaucoma Detection using Deep Learning on Retinal Fundus Images

Arpita Paria[1], Kavyashree I Pattan[2], Nandini K[3], Annapurna Shobitha[4], Lakshmanan M[5] and Shivamma D[6]
[1-4]Assistant Professor, Dept. of CSE, School of Engineering, Dayananda Sagar University, Harohalli, Bengaluru South Dt., Karnataka, India
Email: arpitaporia11@gmail.com, kavyashreepattan-cse@dsu.edu.in, nandini-cse@dsu.edu.in, sannapurna.work@gmail.com
[5]Assistant Professor, Dept. of CSE (AI&ML), School of Engineering, Dayananda Sagar University, Harohalli, Bengaluru South Dt., Karnataka, India
Email: lakshmanan1909@gmail.com
[6]Assistant Professor, Department of CSE (Data Science), School of Engineering, Dayananda Sagar University, Harohalli, Bengaluru South Dt., Karnataka, India
Email: shivammad-cse@dsu.edu.in

***Abstract*— Glaucoma is one of the most prevalent permanent blindness in the world, and it may be hard to detect at the initial stages or even be asymptomatic because of slow progression. Retinal fundus imaging has been used as a popular technique for detecting the retinal optic nerve head and nearby structures in a very accurate manner; it has become a widely used tool due to its timely diagnosis. The publicly available datasets, namely, ACRIMA and ORIGA, and a hybrid dataset aimed at enhancing the model robustness with respect to different image sources, the given research is a deep learning-based blind glaucoma detection model. The model used transfer learning based on VGG16 on models trained on ACRIMA and EfficientNetB0, augmented with contrast preprocessing on ORIGA, trained with data augmentation on class imbalance and overfitting. The experimental assessment revealed that the custom CNN achieved training accuracy of 96.28% and validation accuracy of 93.57% on ACRIMA, and the transfer learning model trained on the combined dataset gained through training accuracy of 95.71% on the ACRIMA test set, which demonstrates that generalization is strong across datasets. Findings demonstrate that the suggested approach is correct and could be used in early glaucoma detection as well as in large-scale clinical screening.**

***Keywords*— Glaucoma detection, Convolutional Neural Network (CNN), Retinal fundus images, ORIGA dataset, ACRIMA dataset.**

I. INTRODUCTION

In the world, glaucoma is an important cause of permanent blindness and a serious ocular condition without any apparent symptoms until significant vision impairment has developed, and it primarily affects people over 60. The biggest problem with glaucoma is that once vision is lost, it cannot be restored, which makes early detection and diagnosis critical to preventing long-term damage. Recent studies estimate that around 76 million people were living with glaucoma in 2020, and this number is expected to rise to over 111 million by 2040. Given its slow progression and lasting effects, there is a pressing need for screening methods that are quick, accurate, and widely accessible—especially in regions where advanced eye care is not easily available.

Grenze ID: 01.GIJET.12.1.243

Currently, Multiple diagnostic procedures, which include optic nerve head assessment, intraocular pressure measurement, and visual field analysis, are used to detect glaucoma. Although these tests are standard and useful, and have some drawbacks. Retinal fundus photography has become increasingly popular as a diagnostic tool because it is simple, non-invasive, and captures detailed images of the back of the eye, allowing doctors to identify glaucoma-related changes more easily. In the upcoming years, artificial intelligence, especially deep learning, has started to improve medical diagnosis. Among AI methods, Convolutional Neural Networks have proven particularly effective for analyzing images.

CNNs can distinguish inconspicuous designs in retinal photographs that will be troublesome for human eyes to take note of, and they do so without requiring manual input to choose highlights. Numerous considers have investigated glaucoma location utilizing CNNs with retinal fundus images. Few approaches center on fragmenting the optic glass and circle to estimate the Cup-to-Disc Proportion (CDR), whereas others bypass division and straightforwardly classify pictures utilizing CNN models. One of the typical issues in this area is the limited availability of large well-labeled datasets.

To overcome it, the exchange learning is often used by analysts, i.e. CNNs trained on very huge dataset of images, such as ImageNet, are adapted to the glaucoma location problem. This step accelerates the preparation process and often increases accuracy. In addition, picture enlargement strategies like rotation, flipping, and zooming are applied in preparing to widen the dataset collection and provide support the models to generalize better than unknown pictures.

The purpose of this ponder was to design and evaluate CNN-based glaucoma discovery models on two public datasets, ACRIMA and ORIGA. The experiment examined the effect of data set combinations on show demonstrations and experimented with two exchange learning architectures, VGG16 linked to ACRIMA images, and EfficientNetB0 linked to CLAHE-processed ORIGA images. Finally, the model performance was surveyed by use of precision measurements, perplexity lattices and precision movement charts to calculate the effectiveness of the models to identify glaucoma and manage masked information.

II. Literature Review

Rohit Thanki et al. [1] report about the use of deep neural networks with machine learning algorithms in classifying retinal fundus images. They provide accurate diagnosis and decision-making tools to ophthalmology with a special emphasis on glaucoma detection using convolutional neural networks (CNNs). Hongyu Jiang et al. [2] use deep learning and predictive analytics to study the use of genre-based emojis in video comments. Despite the fact that their mechanisms of handling visual and textual data are not directly related to medical imaging, they still show procedures that can be generalized to the medical image classification. Afolabi O. Joshua et al. [3] evaluate Cup-to-Disc Ratio (CDR) method in detection of glaucoma in fundus pictures by the importance of optic disc and cup segmentation as a diagnostic tool in the automated screening systems. S. Iqbal et al. [4] provide an in-depth review of the advancements made on the fundus image analysis with preprocessing, segmentation and classification methods. The fact of their publication brings to the fore the growing significance of AI-based models and the significance of powerful datasets in ophthalmic imaging. According to Mr. Yedukrishnan S et al. [5], a CNN based glaucoma detector that focuses more on deep features of the optic nerve head is proposed and has demonstrated promising results in detecting glaucoma at an earlier phase using retina fundus images.

To identify cup and optic disc positions in fundus images, Sandra Virbukaitė [6] describes the ensemble CNN architectures to train the discrete optic disc and cup positions in different fundus images. The ensemble approach enhances the strength and accuracy of the glaucoma detection because of the power of several models. The binary classification problem to separate both glaucomatous and non-glaucomatous cases into binary classes is addressed by the spatial properties of the color fundus images accessible under a CNN-based classification pipeline presented by P. Elangovan and M.K. Nath [7] with extremely high accuracy. Dr. G. P. Ramesh and others include a super pixel-based algorithm to compute the CDR that retains the anatomical features of the preprocessing stage and further enhances the accuracy of a glaucoma diagnosis by enhancing the quality of the segmentation [8].

The estimation of the cup-to-disc ratio approach presented by Mandeep Singh et al. [9] is simplified and is geared towards computational efficiency to fit large-scale screening programmes and specifically under resource-limited circumstances. The article by Ahmed Almazroa et al. [10] looks at a broad spectrum of optic disk and cup segmentation techniques to identify glaucoma and provide a detailed comparison, which may assist in identifying the best practices and questions in automated systems.

III. Proposed Methodology

This research project proposes the application of a machine-based variant of diagnosing glaucoma with deep learning algorithms associated with retinal fundus photography. The ponder is an ordered and clear process, and the first step will include getting pictures of the publicly available datasets. The images are up-sized and post-processed to make the show more lively following the collection of the data. On the basis of classification, this data was divided into two groups on the grounds of glaucomatous highlights and those that were normally regarded as normal pictures. These were some of the deep structures of learning that were contemplated in this reflection, which included a custom-built convolutional neural network (CNN), as well as transfer learning models, which are based on popular architectures, i.e., VGG16 and EfficientNetB0. The model was trained on the prepared databases, and the performance was measured by the generally recognized assessment parameters, such as overall accuracy and confusion matrices, and training accuracy versus time. In this paper, Figure 1 shows a simplified topography of the entire workflow to be utilized.

A. Convolution Neural Network.

Image classification applications typically involve the competence of Convolutional neural networks (CNNs) to be able to automatically learn spatial representations of input images, which saves the need to manually extract features. This research applies several layers of learnable filters that extract patterns of data, textures such as edges, and other relevant visual features. The CNN architecture proposed in this paper has several convolutional layers with ReLU activation functions to add non-linearity. These are followed by max-pooling layers, which help to downsize the feature map without losing significant details to avoid overfitting and enhance efficiency. Once the features have been extracted, the network is linked with dense layers, which interpret the patterns detected and make a classification. The last layer is a Softmax, which classifies pictures as either glaucomatous or normal. The function categorical cross-entropy loss is employed to improve the model, and the ADAM optimizer is applied to allow adaptive and efficient learning in training.

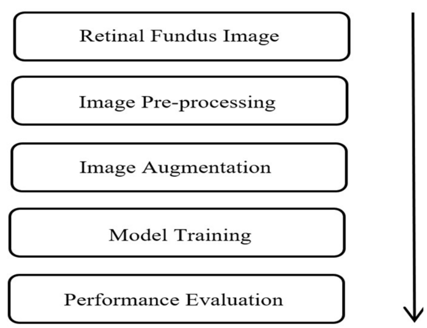


Figure 1. Workflow of a glaucoma detection framework using deep learning.

Retinal Fundus Image

This study employed two publicly accessible retinal fundus image datasets, ACRIMA and ORIGA, both of which include labeled images representing glaucomatous and healthy eyes. ORIGA dataset was used for cross-dataset testing to assess the model's capacity to generalize to previously unknown data, whereas the ACRIMA dataset was used for model training and validation. These datasets were chosen due to their clinical significance, availability of verified annotations, and widespread use in previous glaucoma detection research.

Image Preprocessing

The consistency of the input data for the deep learning models was enhanced by preprocessing. The input specifications of CNN architectures, like VGG16 and EfficientNetB0, all retinal images were scaled to 224 × 224 pixels. Using common preprocessing techniques, pixel intensity levels were standardized to a similar scale. In order to improve local contrast, especially near the optic nerve head, and help the EfficientNetB0 model identify characteristics linked to glaucoma, Contrast Limited Adaptive Histogram Equalization (CLAHE) was also used.

Image Augmentation

To mitigate the challenges posed by limited dataset sizes and to improve model robustness, various image augmentation methods were applied during training. These augmentations artificially expanded the training set by creating modified versions of existing images through:

Rotation: Within a specific range of angles, images were rotated at random.
Flipping: Horizontally and vertically flipping images to create mirror images.
Zooming and Shifting: To simulate the variations observed in clinical image acquisition, images are slightly resized and repositioned. These adjustments improved model's ability to handle the unpredictability present in real-world retinal pictures, to decrease the chance of overfitting, and increased the diversity of the data.

Model Training
The study developed and trained three distinct deep learning models: A custom CNN trained on the ACRIMA dataset to establish baseline performance. A transfer learning model based on VGG16, trained on a combined dataset from ACRIMA and ORIGA, with evaluation conducted on the ACRIMA test partition. An EfficientNetB0 transfer learning model, incorporating CLAHE preprocessing, was trained on the combined dataset and evaluated using the ORIGA test set. Transfer learning enabled the models to utilize pre-existing feature extraction knowledge from networks pretrained on large-scale datasets like ImageNet. The final layers of these models were replaced and fine-tuned to adapt specifically to the glaucoma detection task.

Performance Evaluation
The effectiveness of the models was measured in a number of conventional measures:
Accuracy: The proportion of correctly predicted images of the overall number of images imposed.
Loss: A measure of prediction error followed during validation and training.
Area Under the ROC Curve (AUC): The AUC is a measure of the ability of the model to distinguish between normal and glaucomatous cases at various decision thresholds.
Recall, Precision, F1-Score: These measures were computed in order to determine the trade-off between false positives and false negatives, which in medical diagnosis is critical as the result of a false negative may be disastrous.
Additional research involved plotting training accuracy and ill-fortune throughout training, disarray networks to display classification errors, and point by point classification reports to illustrate steady quality and generalization across datasets.

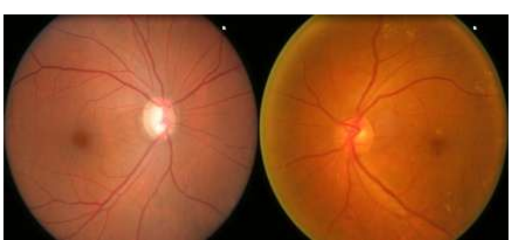
IV. RESULTS

The glaucoma discovery framework outlined in this ponder was built and tested through an organized set of tests utilizing freely accessible retinal fundus image datasets. This range depicts the test framework, how the data was confined, and the results recorded for each testing circumstance. To survey the models' capacity to distinguish between glaucomatous and normal eyes, execution was measured using estimations such as precision, perplexity grids, and planning precision designs over time.

A. Experimental Setup

Experiments were conducted using publicly available retinal fundus image datasets ACRIMA and ORIGA. Initially, ACRIMA was used alone for training and validation. To improve model robustness and generalization, images from both datasets were later combined to create a diverse training set. The models were then independently evaluated on the ACRIMA and ORIGA test sets to assess performance across different image distributions. Three models were examined: a custom CNN, VGG16, and an EfficientNetB0-based transfer learning model with CLAHE preprocessing. All models were trained under the same settings, including a batch size of 32, learning rate of 0.0001, categorical cross-entropy loss, and the ADAM optimizer. Data augmentation was applied consistently to enhance performance and prevent overfitting.

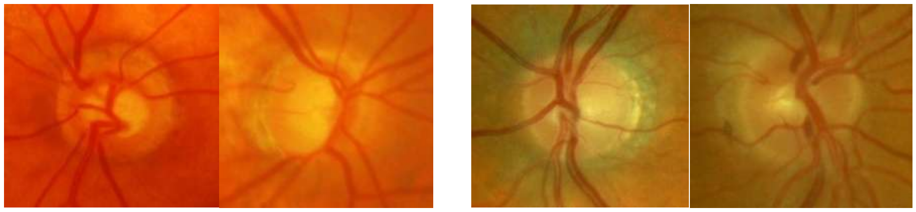
B. Dataset Ratio



(a) Glaucoma

(b) Normal

Figure 2. Sample ORIGA dataset eye: Sample glaucoma and normal images from the ORIGA dataset



(a) Glaucoma (b) Normal

Figure 3. Sample ACRIMA dataset eye: Glaucoma and normal images from ACRIMA dataset.

Publicly accessible retinal fundus image datasets, namely ACRIMA and ORIGA provided the samples for this research. Each dataset included clearly labeled cases of glaucomatous and normal eyes, organized into two categories for binary classification during training and evaluation. For the training phase, a combined dataset was created with a total of 1,225 images — 530 showing glaucomatous eyes and 695 representing normal cases. From this, 80% of the images were used for training, and the remaining 20% set aside for validation. To evaluate the trained models, two independent test sets were assembled. The ORIGA test set consisted of 194 images, split almost evenly between 98 glaucomatous and 96 normal eyes. In addition, a test subset from the ACRIMA dataset was prepared, containing 141 images with 79 glaucomatous and 62 normal cases. Using separate test sets made it possible to assess the models' performance both within the same dataset and on a completely different one. For visual reference, Figure 2 presents a glaucomatous eye image from ORIGA, while Figure. 3 displays a similar case from ACRIMA. These examples highlight optic nerve head changes typically associated with glaucoma, which the models were trained to detect.

B. Comparative Analysis

The performance of the proposed glaucoma detection models was compared across the three experimental cases, as summarised in Table 1.

TABLE I. COMPARATIVE ANALYSIS OF 3 CASES

Case	Training Accuracy	Validation Accuracy	Test Accuracy
Case 1	0.9628	0.9357	0.94
Case 2	0.9224	0.5837	0.9574
Case 3	0.8439	0.5796	0.8093

Case 1: ACRIMA Dataset

The model was trained and tested solely on the ACRIMA dataset, it showed consistently high performance. The training accuracy reached 96.28%, with a validation accuracy of 93.57% and a test accuracy of 94%. Since the data used for both training and testing came from the same source, the model could easily recognize familiar patterns and features, which helped maintain reliable and stable classification results without dealing with much variability.

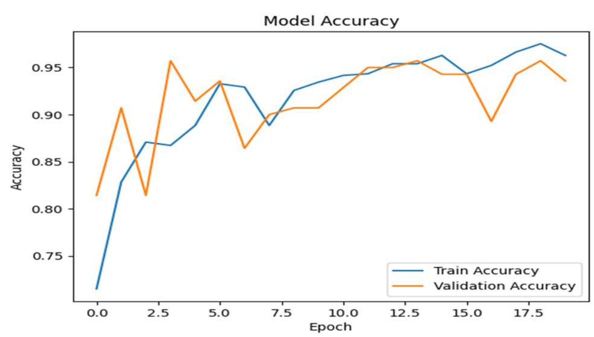


Figure 4. Accuracy graph for training and validation using the ACRIMA dataset

The model exhibited strong performance when both the training and validation images were drawn from the same dataset. These high-accuracy values demonstrate the effectiveness of the CNN architecture and the applied image augmentation methods in reliably capturing glaucoma-specific features within a uniform dataset shown in Figure 4.

Case 2 Combined Dataset Tested on ACRIMA Test Set.

In the second case, a model on the VGG16 transfer learning framework was trained on a pool of mixed data of both ACRIMA and ORIGA, in which data augmentation was used during training. At this stage the implementation of the model was examined on the ACRIMA test set to test its ability to extrapolate a familiar domain when trained on more diverse training data. This method resulted in the greatest test accuracy at 95.74 percent exceeding CNN-only model trained on ACRIMA data alone in Case 1. Although the accuracy of training was a few degrees lower at 92.24 percent and the approval accuracy dropped to 58.37 percent as the test performance advanced indicates that entrenching ORIGA images during training provides imperative added features. These highlights reconfigured the ability of model in identifying glaucomatous trends within ACRIMA dataset. The confusion matrix supported these findings, indicating a minimal number of false negatives and false positives. The application of data augmentation methods—including flips, rotations, zooms, and brightness adjustments—likely contributed significantly to this result by reducing overfitting and broadening the model's exposure to varied image characteristics.

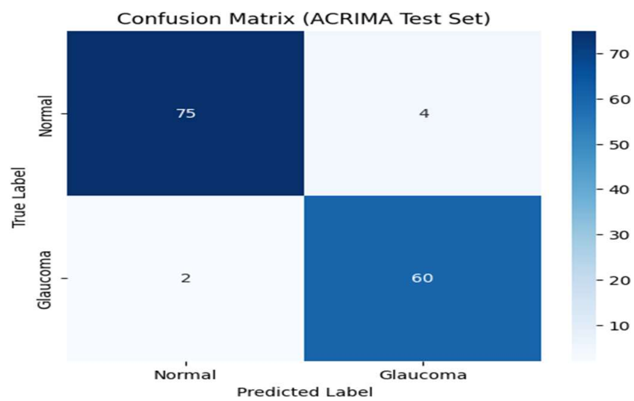


Figure 5. Confusion Matrix tested on ACRIMA test set for the combined dataset

Confusion matrix showing the model's performance on the ACRIMA dataset. Correct classifications are on the diagonal, which is 75 Normal, 60 Glaucoma, while off-diagonal values 4, 2 represent misclassifications shown in Figure 5.

Case 3 was the most difficult scenario. The model, trained on the combined dataset, was tested on the ORIGA test set. As expected, the test accuracy declined to 80.93%. This drop reflects the challenges of applying a model trained on mixed data to a different dataset, especially when the images vary in quality, resolution, and appearance of the optic disc. The training and validation accuracies in this case were 84.39% and 57.96%, respectively. Even so, the model still performed reasonably well, which indicates that training with images from multiple sources helped it learn some transferable features useful for analyzing the ORIGA images, though not as effectively as in the other two cases.

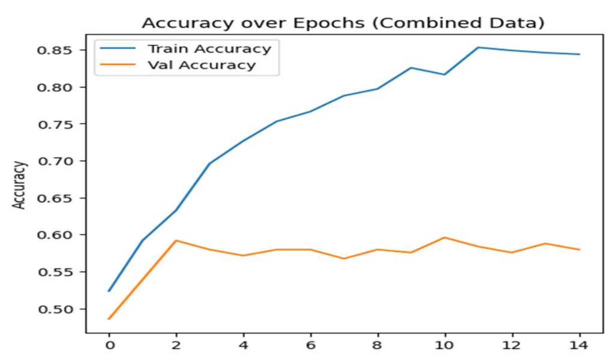


Figure 6. Accuracy graph for the combined dataset tested on ORIGA_test_set

Figure 6 shows that the improvement in the model's performance is due to the use of CLAHE preprocessing, which effectively enhanced the contrast in the fundus images.

C. Experimental Validation and Statistical Analysis

TABLE II. COMPARATIVE PERFORMANCE OF DIFFERENT DEEP LEARNING MODELS FOR GLAUCOMA DETECTION

Model	Dataset	Accuracy	Sensitivity	Specificity	Precision	F1 Score
Custom CNN	Acrima	96.28	88	97	94	0.91
VGG16	Acrima	95.71	95	96	94	0.95
EfficientNet-B0	Origa	81	82	79	80	0.81

Table II shows that the Custom CNN achieved the highest accuracy of 96.28% on the ACRIMA dataset, followed closely by VGG16 with 95.71%, while EfficientNet-B0 attained 81% on ORIGA. The results confirm the strong generalization and reliability of the proposed deep learning models for early glaucoma detection.

D. Frontend Implementation

To make the glaucoma detection more convenient and easy to use, a simple web-based interface was developed using the Streamlit framework. Through this interface, users can upload fundus images directly, which are then displayed on the screen for visual confirmation. Once an image is submitted, the system processes it using the trained deep learning model and returns a prediction indicating whether the image is normal or glaucomatous. This lightweight, user-friendly frontend dispenses with the need for specialized software and empowers clinicians, analysts, and healthcare staff to perform automated glaucoma screening rapidly and proficiently in any environment.

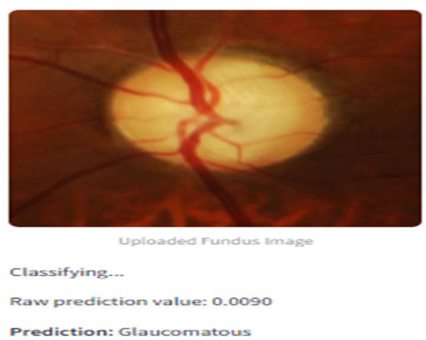


Figure 7. Glaucomatous Eye: Fundus image predicted as glaucomatous

Figure 7. Fundus image analyzed by the developed system, yielding a glaucomatous prediction with the corresponding confidence value.

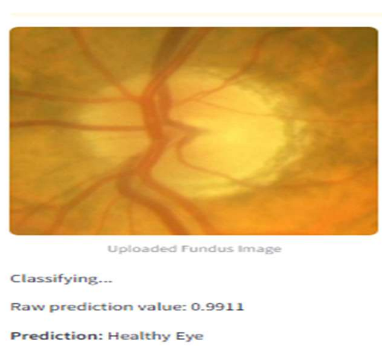


Fig 8. Healthy Eye: Fundus image predicted as healthy

Figure 8. The fundus picture was analyzed by the proposed model which diagnosed the case as a normal eye with the appropriate level of confidence.

V. CONCLUSION

This study explored the use of deep learning methods for detecting glaucoma automatically from retinal fundus images. The findings show that models tend to perform best when both training and testing are done on images from the same dataset, likely because of the uniformity in imaging conditions and features. Evaluating the models across different datasets proved more difficult, due to differences in image resolution, optic disc appearance, and how the images were captured. Despite these challenges, the combination of the EfficientNetB0 architecture and CLAHE preprocessing improved the model's sensitivity to glaucoma detection, which is especially important in clinical settings where missing true cases must be minimized.

Future work will focus on broadening the range of training data by including more publicly available glaucoma datasets. This will help the model work better with different types of images and a variety of patients. At the same time, we want to make the model's decisions easier to understand so doctors can trust it more. In the end, the goal is to build real-time tools that help healthcare workers spot glaucoma sooner and make their job of reviewing images less time-consuming.

REFERENCES

[1] A deep neural network and machine learning approach for retinal fundus image classification, Rohit Thanki Krian GmbH, Wolfsburg, Germany.

[2] Jiang, Hongyu & Guo, Ao & Ma, Jianhua. (2020). Genre-based Emoji Usage Analysis and Prediction in Video Comments. 10.1109/DASC-PICom-CBDCom-CyberSciTech49142.2020.00058.

[3] Afolabi O.Joshua, Gugulethu Mabuza-Hocquet, Fulufhelo V.Nelwamondo, "Assessment of the Cup-to-Disc ratio method for Glaucoma detection", Research Gate, 2020.

[4] S. Iqbal, T.M. Khan, K. Naveed, S.S. Naqvi, S.J. Nawaz, Recent trends and advances in fundus image analysis: A review, Comput. Biol. Med. (2022) 106277.

[5] Mr. Yedukrishnan S, Dr. V. Geetha, Ms. V. Jalaja Jayalakshmi, Glaucoma detection in fundus image using CNN, 2021 IEEE.

[6] Sandra Virbukait˙e, Glaucoma Identification Using Convolutional Neural Networks Ensemble for Optic Disc and Cup Segmentation

[7] P. Elangovan and M. K. Nath, ''Glaucoma assessment from color fundus images using convolutional neural network,'' Int. J. Imag. Syst. Technol., vol. 31, no. 2, pp. 955–971, Jun. 2021, doi: 10.1002/ima.22494.

[8] Dr.G.P. Ramesh, Malini, "An Efficacious Method of Cup to Disc Ratio Calculation for Glaucoma Diagnosis Using Super pixel", IJCSEC-International Journal of Computer Science and Engineering Communications, Vol.2 Issue.3, 2004.

[9] Mandeep Singh, JiwanpreetKaurVirk, Mooninder Singh, "A simple approach to Cup-to-Disk Ratio determination for Glaucoma Screening", IJCSC-International Journal of Computer Science &Communication, Vol.6 Issue.2, 2015.

[10] Ahmed Almazroa, Ritambhar Burman, Kaamran Raahemifar, and Vasudevan Lakshminarayanan, "Optic Disc and Optic Cup Segmentation Methodologies for Glaucoma Image Detection: A Survey", Hindawi Publishing Corporation Journal of Ophthalmology, 2015.