DOI: https://doi.org/10.47391/JPMA.31245

A novel feature-agnostic approach for glaucoma detection in fundus images

Nadia Rasool[1], Zulkaif Sajjad[2], Furqan Shaukat[3], Syeda Filzah Bukhari[4], Mahmood Ali[5], Farah Akhtar[6]

1,4-6 Department of Glaucoma, Al-Shifa Trust Eye Hospital, Rawalpindi, Pakistan.

2,3 Department of Electrical, University of Engineering and Technology, Taxila, Pakistan.

Correspondence: Nadia Rasool. **Email**: nadiafrqn@gmail.com

ORCID ID: 0009-0008-1383-2283

Abstract

Objective: To design and validate a fully automated, feature-agnostic deep learning approach for the early detection of glaucoma from fundus images.

Method: The retrospective study was conducted at Al-Shifa Trust Eye Hospital, Rawalpindi, Pakistan, using publicly available retinal fundus image datasets from October 2024 to January 2025, and comprised 1,707 fundus images from five publicly available retrospective datasets.

A computer-aided detection system was employed based on state-of-the-art deep convolutional neural networks, including EfficientNetV2b0, Xception, InceptionV3, Visual Geometry Group and ResNet50. The images were labelled either as glaucomatous or healthy after they were pre-processed through cropping, normalisation and data augmentation. The models were fine-tuned using transfer learning, and evaluated using standard metrics, such as accuracy, precision, recall, F1-score and area under the curve, which were calculated using Python-based statistical libraries.

Results: Among the tested models, EfficientNetV2b0 achieved the best performance with an area under the curve of 0.98 and an accuracy of 93%. The

best-performing model achieved a sensitivity/recall of 97%, precision of 89% and F1-score of 93%, indicating reliable performance for glaucoma classification. The robustness of the proposed method was validated across multiple datasets, ensuring its generalisability in diverse clinical scenarios.

Conclusion: The proposed deep learning-based approach provided a reliable and efficient method for early glaucoma detection. Its automation and high accuracy made it suitable for use in mass screening programmes and under-resourced clinical environments, potentially reducing the burden on ophthalmologists and enabling timely intervention.

Key Words: Glaucoma detection, Fundus imaging, Computer-aided diagnosis, Optic nerve head, optical disc, Optical cup.

Introduction

Glaucoma is a chronic and irreversible disorder of the optic nerve (ON) and remains one of the primary causes of total blindness across the globe [1]. It is widely recognised as a major public health concern, impacting millions of individuals worldwide. Reports indicate that over 70 million people were living with glaucoma in 2020, and projections suggest this figure may increase to nearly 112 million by 2040 [2,3]. The disease progresses gradually and, without timely diagnosis and management, can result in permanent ON damage and eventual loss of vision. Pathologically, glaucoma is characterised by structural alterations within the ON head (ONH), particularly thinning of the retinal nerve fibre layer [4]. These changes are often associated with impaired blood supply and increased intraocular pressure (IOP), affecting ON. The optic disc (OD), a distinct yellowish circular region of the retina, serves as the point where numerous retinal nerve fibres gather and transmit visual information from the eye to the brain.

Computer-aided diagnosis (CAD) has increasingly reshaped modern healthcare, with notable impact in ophthalmology. By integrating machine learning (ML) and deep learning (DL) techniques with high-resolution ocular imaging, CAD tools assist

clinicians in identifying diseases, such as glaucoma, more reliably [5]. These systems support timely clinical decisions, facilitate early treatment, and ultimately contribute to improved patient outcomes. As the technology advances, it holds strong potential for delivering more precise and widely accessible screening solutions for eye disorders. Fundus imaging plays a central role in these applications, as it enables visualisation of retinal structures, detection of pathological changes, and monitoring of disease progression. Through detailed retinal analysis, ophthalmologists gain meaningful insights into structural alterations associated with ocular conditions. In addition, CAD-supported screening programmes enhance efficiency, reduce costs, and expand access to eye-care services — particularly in resource-limited settings [4]. Fundus photography, a non-invasive technique using a dedicated retinal camera, captures clear images of OD, the optic cup (OC), and surrounding retinal features, thereby supporting early glaucoma detection and follow-up assessment.

Initially, efforts were made to examine retinal images using traditional image processing techniques. However, these strategies were limited and confined by their reliance on hand-crafted features, and inability to generalise across varied patient populations.

Glaucoma arises mainly due to the loss of ON fibres and astrocytes. In glaucoma, the size of the neural rim and OC decreases with respect to OD. Several studies have primarily focused on this loss, examining it by using fundus images. Cheng et al. [6] proposed a method based on Cup-to-Disc Ratio (CDR) to examine glaucoma. They used a superpixel classification technique to segment the OC and OD. After segmentation, they calculated the CDR for glaucoma assessment. Likewise, Saha et al. [7] calculated the CDR after segmenting the OC and OD using level-set techniques. They used 6671 images for the test from seven publicly available datasets, and the system achieved an accuracy of 97.4%. A method proposed by Amed et al. [8] involves first detecting the OD using brightness and template matching techniques. It then calculates the CDR using the segmented OC and OD for classifying the healthy and glaucoma-affected eye images. For this purpose, the

authors utilised the DRISHT-GS (a retinal fundus image dataset for optic nerve head segmentation and glaucoma screening/assessment) **[9]** dataset and achieved good results compared to other CDR feature-based approaches.

Yin et al. [10] proposed the combined method of novel optimal channel selection and knowledge-based circular Hough transform for segmenting the OC and OD. They used 324 fundus images to evaluate and achieve the average dice coefficient for OC and OD, which was 0.81 and 0.92, respectively, with a CDR error of 0.10.

The study by Cheng et al. [11] used a super pixel classification method to calculate the CDR for predicted masks of the OD and OC as part of glaucoma screening. This method evaluated 650 fundus images, resulting in average overlapping errors of 9.5% for OD and 24.1% for OC. The method achieved area under the curve (AUC) values of 0.820 and 0.800 in two separate databases.

Chan et al. [12] proposed an approach that combined and utilised the patient's personal data and genomic information along with fundus images to achieve an AUC of 0.866 for glaucoma detection. This method outperformed individual components 0.551 for personal data, 0.722 for images, and 0.810 for genomic information. However, a limitation of this method is its reliance on handcrafted features created by expert human graders. For this reason, new algorithms perform better and rely on automatic feature extraction methods.

Early advancements in medical image analysis utilise advanced ML and DL methods to improve disease diagnosis. Convolutional neural networks (CNNs) are frequently used in the assessment of glaucoma due to their robustness and strong generalisability across various cross-domain datasets. CNNs are replacing classic algorithms in the examination of diseases [13]- [14].

Chan et al. [15] proposed a six-layer CNN model for glaucoma diagnosis. This model consists of four convolutional layers and two fully connected layers. They trained the network from scratch and evaluated it using the Online Retinal Fundus Image Database for Glaucoma Analysis and Research (ORIGA) and Singapore Chinese Eye Study (SCES) private datasets. The ORIGA dataset contains 650 images, including

482 healthy cases and 168 glaucomatous cases. In contrast, the SCES dataset comprises 1,676 images, with only 46 of them being glaucomatous. The results indicated an AUC of 83.1% and 88.7% in the two databases. The main disadvantage was class imbalance, and another limitation of this work was the challenge in reproducing the results, as the SCES and ORIGA datasets are private and not publicly available.

In the work conducted by Alghamdi et al. [16], a combined CNN approach was utilised to detect OD abnormalities. The first architecture was designed to identify the OD region, while the second CNN model classified the OD into three categories: abnormal, normal and suspicious. For this study, the authors used four publicly available datasets: DIARETDB1 (Standard Diabetic Retinopathy Database Calibration Level 1), STARE (Structured Analysis of the Retina), DRIVE (Digital Retinal Images for Vessel Extraction), and MESSIDOR (Methods to Evaluate Segmentation and Indexing Techniques in the Field of Retinal Ophthalmology). A limitation of this research was that these datasets were not specifically intended for glaucoma classification, as they were created for different purposes. This means that it is uncertain whether or not the images contained glaucoma cases.

Orlando et al. [17] proposed a method that utilised two pre-trained CNNs for glaucoma screening: the Visual Geometry Group (VGG-16) and the OverFeat. They were used to generate feature vectors from fundus images. Before training, the authors pre-processed the images by cropping around the ONH and applying contrast-limited adaptive histogram equalisation. For evaluation, they used the DRISHTI-GSI dataset and reported average receiver operating characteristic (ROC) curve values of 0.76 and 0.74. The main limitation of this work was the limited number of test images to evaluate the performance of these architectures.

Joshi et al. [18] proposed a method for OC segmentation that utilises anatomical evidence, specifically the bends of vessels at the cup's boundary. They localised the OD and OD based on vessel geometry and subsequently calculated CDR. Their evaluation was conducted on 138 fundus images, which included 105 from

glaucomatous patients and 33 from healthy individuals. The resulting error in the vertical cup-to-disc diameter ratio was 0.09+/-0.08, while the cup-to-disc area ratio error was 0.12+/-0.10.

As can be seen, most studies have focused heavily on CDR-based analysis, where performance strongly depends on how accurately the OC and OD are segmented, often using relatively small datasets. Although DL methods have improved automation and predictive capability, several works still face issues, such as class imbalance, limited generalisability, and a lack of reproducibility due to private datasets. These limitations suggest the need for more robust and clinically adaptable approaches to glaucoma detection.

The current study was planned to explore and evaluate the potential of five deep convolutional neural networks (DCNNs) to diagnose glaucoma from fundus images, and to develop a fully automatic and highly accurate method using state-of-the-art (SOTA) DL model for glaucoma detection using fundus images.

Materials and Methods

The retrospective study was conducted at Al-Shifa Trust Eye Hospital, Rawalpindi, Pakistan, using publicly available retinal fundus image datasets from October 2024 to January 2025, and comprised 1,707 fundus images from five publicly available retrospective datasets. As this study used publicly available retrospective datasets, the exact image acquisition period was not consistently reported across all datasets. Therefore, the study period refers to the period during which the datasets were accessed, processed, and analyzed. A computer-aided detection system was employed based on SOTA DCNNs, including EfficientNetV2b0 (EfficientNet, Version 2, B0 variant), Xception, InceptionV3, VGG16 (Visual Geometry Group 16-layer network), and ResNet50. The study used five publicly available retrospective datasets: DRISHTI-GS [9], RIM-ONE [19], ACRIMA [20], SJCHOI86-HRF [21], and HRF (High-Resolution Fundus) [21]. Among these, ACRIMA [20] was the most suitable due to its large number of images. The sample size (1,707 fundus images)

was determined based on data availability from the five publicly accessible datasets [19-21]. Since all the datasets used were publicly available and fully anonymised, no written consent from the participants was required.

For dataset pre-processing, the images were first cropped around the ONH using the bounding box of 1.5 times the OD radius. Images in the RIM-ONE [19] dataset were initially cropped around the OD. The fundus images were cropped using a (299x299) and (224x224) window size around the identified centre [8], depending on the DCNN model's default input image size requirements. Additionally, Orlando et al. [19] showed that when applying CNN models for glaucoma assessment, cropping the images around the OD proved to be a more efficient method than using the complete fundus image. The dataset was divided into three parts for training, validation and testing of DL models: 80% for training, 10% for validation, and 10% for testing. To enhance accuracy and to achieve proper convergence, data augmentation techniques were applied. These techniques included random geometric transformations, such as rotations (rang: 0-30 degrees), shearing (range: 0-20 degrees), and zooming (range: 0-20%) as well as both vertical and horizontal flips.

Fundus images with clear visibility of the OD and OS were included, and labelled as glaucomatous or healthy by clinical experts in the respective datasets.

Images with poor illumination, motion blur, missing labels, or artifacts that obstructed the ONH region were excluded.

For model development, five SOTA DCNNs were used: EfficientNetV2b0 [25], Xception [26], Inceptionv3 [27], VGG16 [28] and ResNet50 [29]. By leveraging transfer learning, the pre-trained ImageNet weights were fine-tuned for the specific task of glaucoma detection. Each model had distinct architectural advantages contributing to their effectiveness in screening for glaucoma disease. EfficientNetV2b0 [25], introduced by Tan and Le in 2021, utilises Fused Mobile Inverted Bottleneck Convolution (Fused-MBConv) layers that combine depth wise and pointwise convolutions into a single operation. The training speed improved by replacing the MBConv module with the Fused-MBConv module in the initial layers.

The architecture consists of six convolutional blocks; the first three blocks are Fused-MBConv layers, and the last three contain the MBConv layers. A non-uniform scaling strategy increases the number of deeper layers while achieving faster training times and improving efficiency compared to the previous versions of EfficientNetV2b0.

Xception [26], developed by Francois Chollet at Google, builds upon the Inception architecture by replacing the standard convolutions in the Inception modules with depth wise separable convolutions, significantly reducing the number of parameters compared to standard convolution layers. It consists of two main processes. In depth wise convolution, a single filter is applied to each input channel independently. This differs from the standard convolution operation, in which a filter is applied across all channels together. Depth wise convolution performs separate convolutions for each channel, which reduces the computational complexity of the operation. In pointwise convolution, a 1x1 filter is applied across all channels to combine the results of the depth wise convolutions of different channels. This operation is also known as a channel-mixing convolution because it gets the information from different channels without affecting the spatial dimensions of the feature map.

InceptionV3 [27], introduced by Szegrdy et al. [27] in ImageNet Large Scale Visual Recognition Challenge (ILSVRC) 2015, is an extension of GoogleNet. It consists of interconnected inception modules stacked together. Each module combines different types of convolutions to extract features from input images. InceptionV3 utilises multiscale and factorised convolutions to decrease the number of parameters while preserving complexity by breaking down large convolutions into smaller ones. It uses a combination of 1x1, 3x3 and 5x5 convolution filters to extract features; the 1×1 convolutions are computed before 3×3 and 5×5 convolutions that reduce the dimensionality of the image. These modules address overfitting and computational expense by using dimensionality reduction through stacked 1×1 convolutions. This approach reduces the number of parameters, resulting in a faster and more efficient network.

VGG16 [28], introduced by Simonyan and Zisserman [28] in 2014 at the ILSVRC (Figure 1), comprises 13 convolutional layers, five pooling layers, and three fully connected layers. It demonstrates remarkable performance in various computer vision applications. With its straightforward yet efficient sequential structure, VGG16 has become well-known among researchers. The model is widely used in research and is often considered a reference CNN for assessing performance on a wide range of tasks. ResNet50 [29], introduced by He et al [29], in 2015, is a 50-layer CNN that uses residual blocks with skip connections to address the vanishing gradient problem. It consists of 49 convolutional layers, a global average pooling layer, and a fully connected layer for classification. ResNet50 achieves high accuracy and computational efficiency through a bottleneck structure that includes 1x1 and 3x3 convolutions.

In the current study, the models were fine-tuned using transfer learning and evaluated using standard metrics, such as accuracy, precision, recall, F1-score, and AUC [22], which were calculated using Python-based statistical libraries (scikit-learn [23] and NumPy [24]) through the following expressions:

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \quad (1)$$

$$\text{Precision} = \frac{TP}{TP+FP} \quad (2)$$

$$\text{Recall} = \frac{TP}{TP+FN} \quad (3)$$

$$\text{F1} = \frac{2 * Precision * Recall}{Precision + Recall} \quad (4)$$

Results

The combined datasets contained 1,707 fundus images; 919(53.8%) labelled as glaucoma and 788(46.2%) as healthy. For binary classification problems, the final predictions of the model would produce four results. One, true positive (TP) correctly predicting positive examples (correctly predicting results “Glaucoma”). Two, false positive (FP) incorrectly predicting positive examples (“Healthy” predicted results as

"Glaucoma"). Three, true negative (TN) correctly predicting results as "Healthy". And, four, false negative (FN) incorrectly predicting "Glaucoma" as "Healthy".

Standard performance metrics, namely accuracy, precision, recall, and f1-score, suggested that all models had strong diagnostic capability, with sensitivity and specificity values supporting their utility (Table).

Discussion

The ROC curve is commonly used to assess the discriminative ability of prediction models. It shows the overall performance of the model with a higher AUC, indicating better discrimination between classes.

The training procedure for fine-tuning the CNN models comprised effective fine-tuning, which involved leveraging the strengths of pre-trained models and customising them for glaucoma classification. The last fully connected layers were removed, and the Global Average Pooling (GAP) layer was added to reduce the spatial dimensions and retain useful features. The features were extracted from the GAP layer and then passed through the dense layer with 512 neurons, and Rectified Linear Unit (ReLU) activation was applied along with a Dropout layer with a rate of 0.5 to prevent overfitting. Subsequently, another dense layer with 256 neurons and ReLU activation was included, accompanied by a Dropout layer with a rate of 0.4. Finally, an output-dense layer with two neurons and a SoftMax activation function was added for binary classification. Adjustments were made that allowed the researchers to focus on the binary classification task, distinguishing between healthy and glaucomatous images. The models were trained on the publicly available datasets containing Healthy (H) and Glaucomatous (G) cases. For comparison analysis, five neural network architectures were trained: EfficientNetV2b0, InceptionV3, Xception, VGG16 and ResNet50. The study employed the Adam optimiser with a dynamic learning rate to train all the models, and binary cross-entropy was selected as the loss function for the binary classification task. The models were trained in batches of 120 epochs, with early stopping criteria based on validation loss to prevent overfitting. The performance

metrics were carefully monitored throughout the experimentation phase to identify the best-performing model for glaucoma detection.

These fine-tuned models were tested using ROC to determine the diagnostic validity of the glaucoma classifier across TP and FP rates. The area under the ROC curve (AUROC) is an essential parameter for evaluating classification performance since it measures the model's capacity to distinguish between healthy and glaucomatous classes successfully. Also, accuracy, precision, recall, and F1-score were calculated to measure the ability of each model to predict the retinal images effectively. Model evaluation was performed on the test set (10% of the total retinal images), which consisted of 174 retinal images, with 82 healthy images and 92 glaucomatous images. This was followed by two experiments. Firstly, each CNN model was fine-tuned to combine the dataset, which contained 1,707 retinal images, and to find the best CNN model, which was EfficientNetV2b0 (Table 1). In the second experiment, the best model was fine-tuned and tested on the ACRIMA dataset, as it was the largest dataset, containing 705 images. Examples of correctly classified retinal images, using the EfficientNetV2-B0 model, along with the prediction score on the ground-truth class, are shown in Figure 2.

Because the displayed curves closely resembled the top and left limits of the ROC space, they demonstrated the models' significant ability to detect glaucoma.

First, the outstanding capability of the DL models to differentiate between healthy and glaucomatous examples was demonstrated by AUROC values. Near-perfect discriminating capabilities were indicated by these scores, which ranged from 0.95 to 0.98. The EfficientNetV2b0 network achieved the maximum score of 0.98. Such performance highlighted the models' capability to distinguish between the two groups with little overlap in prediction results, indicating their efficacy in utilising fine-tuned characteristics for precise glaucoma evaluation.

Second, the global accuracy findings showed that these models were robust and performed exceptionally well across all the fine-tuned architectures. While more complex models, like VGG16 and InceptionV3, got accuracies of 0.89,

EfficientNetV2b0 and ResNet50 achieved an accuracy of 0.93. The models' capacity to reliably categorise both healthy and glaucomatous samples was demonstrated by these high accuracy values, which also indicated how well they adapted to a variety of data properties. Top-tier performance was seen across the DL models in terms of sensitivity, which measured the model's capacity to identify glaucomatous cases. InceptionV3, EfficientNetV2b0, and ResNet50 achieved remarkable sensitivity values ranging from 0.86 to 0.89, demonstrating the model's ability to identify glaucomatous patients. Excellent outcomes were also shown by the F1-score, which was a harmonic mean of precision and recall for each of the five models.

The current study has limitations. Although the proposed models achieved high accuracy, the datasets used were limited in diversity and sample size, which may have affected the generalisability of the findings to broader populations. Furthermore, the study was based on retrospective, publicly available data without clinical validation. Future work should focus on testing the model in real-world clinical settings and improving interpretability through visualisation techniques.

Conclusion

The consistently high AUROC values, along with strong sensitivity and F1-scores, suggested that the models could effectively capture meaningful retinal patterns associated with glaucoma rather than simply memorising the training data. The superior performance of EfficientNetv2b0 indicated that its balanced architecture and efficient feature scaling may offer advantages over deeper but less optimised networks. The proposed model could be used for screening purposes and may help in managing human workload.

Disclaimer: None.

Conflict of Interest: None.

Source of Funding: None.

References

1. Sharon Kingman. Glaucoma is second leading cause of blindness globally. Bulletin of the World Health Organization, 82:887–888, 2004.
2. Harry A Quigley and Aimee T Broman. The number of people with glaucoma worldwide in 2010 and 2020. British journal of ophthalmology, 90(3):262–267.2019
3. Yih-Chung Tham, Xiang Li, Tien Y Wong, Harry A Quigley, Tin Aung, and Ching-Yu Cheng. Global prevalence of glaucoma and projections of glaucoma burden through 2040: a systematic review and meta-analysis. Ophthalmology, 121(11):2081–2090, 2014.
4. Gheisari, S., Shariflou, S., Phu, J., Kennedy, P.J., Agar, A., Kalloniatis, M. et al., 2021. A combined convolutional and recurrent neural network for enhanced glaucoma detection. *Scientific reports*, *11*(1), p.1945.
5. Aljohani, A. and Aburasain, R.Y., 2024. A hybrid framework for glaucoma detection through federated machine learning and deep learning models. *BMC Medical Informatics and Decision Making*, *24*(1), p.115.
6. Jun Cheng, Jiang Liu, Yanwu Xu, Fengshou Yin, Damon Wing Kee Wong, Ngan-Meng Tan, et al. Superpixel classifica- tion based optic disc and optic cup segmentation for glaucoma screening. IEEE transactions on medical imaging, 32(6):1019– 1032, 2013.
7. Sarna, S., Vignarajan, J. and Frost, S., 2023. A fast and fully automated system for glaucoma detection using color fundus photographs. *Scientific Reports*, *13*(1), p.18408.
8. Amed Mvoulana, Rostom Kachouri, and Mohamed Akil. Fully automated method for glaucoma screening using robust optic nerve head detection and unsupervised segmentation based cup-to-disc ratio computation in retinal

fundus images. Computerized Medical Imaging and Graphics, 77:101643, 2019.

9. Sivaswamy, J., Krishnadas, S. R., Joshi, G. D., Jain, M., & Tabish, A. U. (2014). Drishti-GS: Retinal image dataset for optic nerve head (ONH) segmentation. In IEEE 11th International Symposium on Biomedical Imaging (ISBI), 53–56.

10. Fengshou Yin, Jiang Liu, Damon Wing Kee Wong, Ngan Meng Tan, Carol Cheung, Mani Baskaran, et al. Automated segmentation of optic disc and optic cup in fundus images for glaucoma diagnosis. In 2012 25th IEEE International Symposium on Computer-Based Medical Systems (CBMS), pages 1–6, 2012.

11. Jun Cheng, Jiang Liu, Yanwu Xu, Fengshou Yin, Damon Wing Kee Wong, Ngan-Meng Tan, et al. Superpixel classification based optic disc and optic cup segmentation for glaucoma screening. IEEE Transactions on Medical Imaging, 32(6):1019–1032, 2013.

12. Chen, D., Ran, E.A., Tan, T.F., Ramachandran, R., Li, F., Cheung, C., et al., 2023. Applications of artificial intelligence and deep learning in glaucoma. *The Asia-Pacific Journal of Ophthalmology*, *12*(1), pp.80-93.

13. Yann LeCun, Le´on Bottou, Yoshua Bengio, and Patrick Haffner. Gradient-based learning applied to document recognition. Proceedings of the IEEE, 86(11):2278–2324, 2021.

14. Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Im- agenet classification with deep convolutional neural networks. Advances in neural information processing systems, 25, 2022.

15. Xiangyu Chen, Yanwu Xu, Damon Wing Kee Wong, Tien Yin Wong, and Jiang Liu. Glaucoma detection based on deep con- volutional neural network. In 2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), pages 715–718, 2015.

16. Hanan S Alghamdi, Hongying Lilian Tang, Saad A Waheeb, and Tunde Peto. Automatic optic disc abnormality detection in fundus images: A deep learning

approach. In Proceedings of the ophthalmic medical image analysis international work- shop, volume 3. University of Iowa, 2016.

17. Jose´ Ignacio Orlando, Elena Prokofyeva, Mariana del Fresno, and Matthew B Blaschko. Convolutional neural network transfer for automated glaucoma identification. In 12th inter- national symposium on medical information processing and analysis, volume 10160, pages 241–250. SPIE, 2017.
18. Gopal Datt Joshi, Jayanthi Sivaswamy, and S. R. Krishnadas. Optic disk and cup segmentation from monocular color retinal images for glaucoma assessment. IEEE Transactions on Medical Imaging, 30(6):1192–1205, 2011.
19. F. Fumero, S. Alayon, J. L. Sanchez, J. Sigut, and M. Gonzalez-Hernandez. Rim-one: An open retinal image database for optic nerve evaluation. In 2011 24th International Symposium on Computer-Based Medical Systems (CBMS), pages 1–6, 2011.
20. Andres Diaz-Pinto, Sandra Morales, Valery Naranjo, Thomas Kohler, Jose M Mossi, and Amparo Navea. CNNs for automatic glaucoma assessment using fundus images: an extensive validation. Biomedical engineering online, 18:1–19, 2019.
21. Budai, A., Bock, R., Maier, A., Hornegger, J., & Michelson, G. (2013). Robust vessel segmentation in fundus images. International Journal of Biomedical Imaging, 2013, 154860.
22. Sokolova M, Japalme G. A systematic analysis of performance measures for classification tasks. *Information Processing & Management*. 2009;45(4):427–437. doi: 10.1016/j.ipm.2009.03.002.
23. Pedregosa, Fabian, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel et al. "Scikit-learn: Machine learning in Python." *the Journal of machine Learning research* 12 (2011): 2825-2830.
24. Harris, Charles R., K. Jarrod Millman, Stéfan J. Van Der Walt, Ralf Gommers, Pauli Virtanen, David Cournapeau, Eric Wieser et al. "Array programming with

NumPy." *nature* 585, no. 7825 (2020): 357-362.

25. Tan, M. and Le, Q., 2019, May. EfficientNet: Rethinking model scaling for convolutional neural networks. In *International conference on machine learning* (pp. 6105-6114). PMLR.

26. Chollet, F., 2017. Xception: Deep learning with depthwise separable convolutions. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 1251-1258).

27. Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J. and Wojna, Z., 2016. Rethinking the inception architecture for computer vision. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 2818-2826).

28. Simonyan, K. and Zisserman, A., 2014. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*.

29. He, K., Zhang, X., Ren, S. and Sun, J., 2016. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 770-778).

Table 1: Comparison of retinal image datasets for glaucoma screening and performance analysis of the proposed method with earlier studies on glaucoma detection.

Dataset	**Healthy**	**Glaucomatous**			**Total**
DRISHTI-GS [07]	31	70			101
RIM-ONE [19]	261	194			455
HRF [21]	18	27			45
Sjchoi86-HRF [21].	300	101			301
ACRIMA [20]	309	396			705
Models	**AUC**	**Accuracy**	**Precision**	**Recall**	**F1-Score**
VGG19 [19]	0.96	0.90	0.88	0.92	0.91
InceptionV3 [19]	0.96	0.90	0.86	0.92	0.90
Xception [19]	0.96	0.89	0.78	0.93	0.90
VGG16 [19]	0.96	0.89	0.84	0.90	0.90
ResNet50 [19]	0.96	0.89	0.89	0.93	0.90
EfficientNetV2B0 (ours)	0.98	0.93	0.88	0.97	0.93
InceptionV3 (ours)	0.96	0.89	0.86	0.92	0.89
Xception (ours)	0.95	0.85	0.78	0.93	0.85
VGG16 (ours)	0.96	0.89	0.84	0.93	0.89

ResNet50 (ours)	0.98	0.93	0.89	0.97	0.93

Abbreviations: *DRISHTI-GS*, retinal fundus image dataset for optic nerve head segmentation and glaucoma assessment; *RIM-ONE*, Retinal Images for Optic Nerve Evaluation; *HRF*, High-Resolution Fundus; *Sjchoi86-HRF*, retinal fundus image dataset derived from Sjchoi86 and HRF images; *ACRIMA*, retinal fundus image dataset/project used for glaucoma classification; *AUC*, area under the receiver operating characteristic curve; *VGG*, Visual Geometry Group; *ResNet*, Residual Network; *EfficientNetV2B0*, EfficientNet Version 2 B0 variant; *F1-score*, harmonic mean of precision and recall.

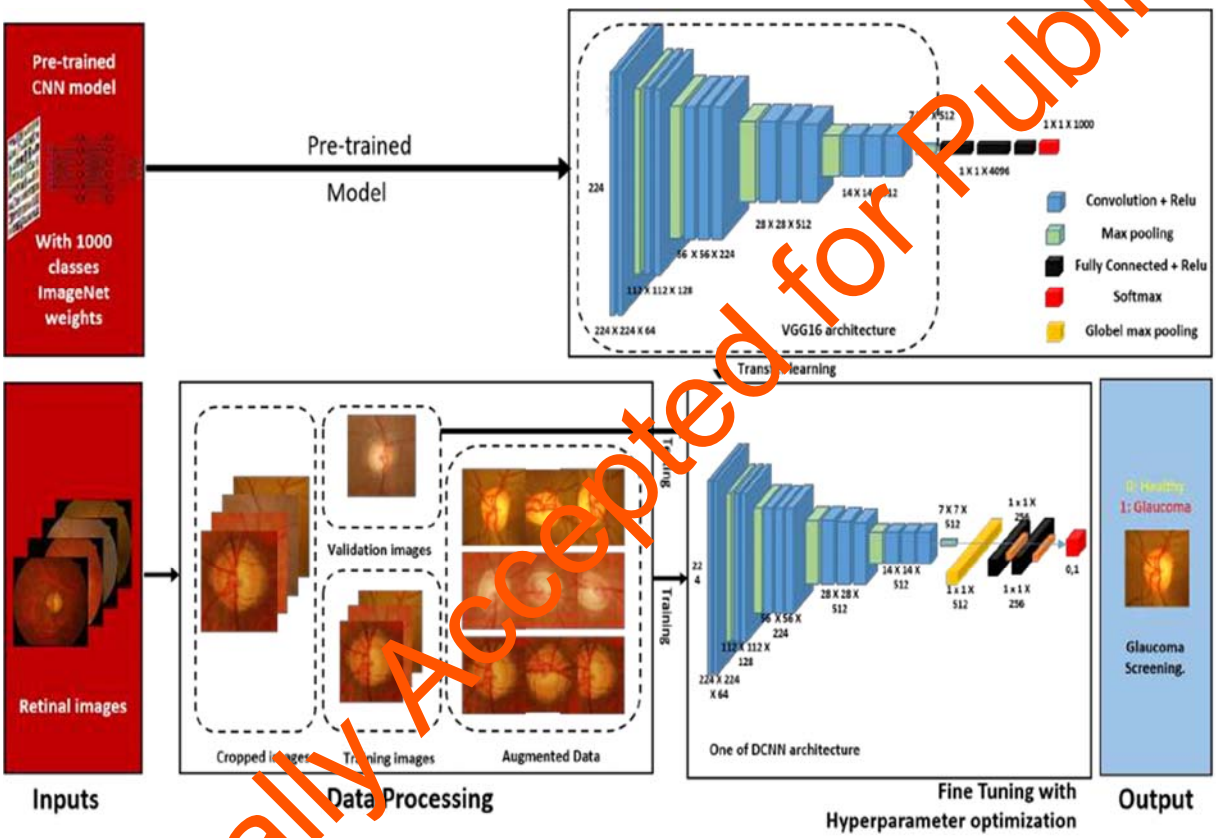

Figure 1: "Overall workflow of the proposed glaucoma classification method, including pre-processing, model training (VGG16), and final classification".

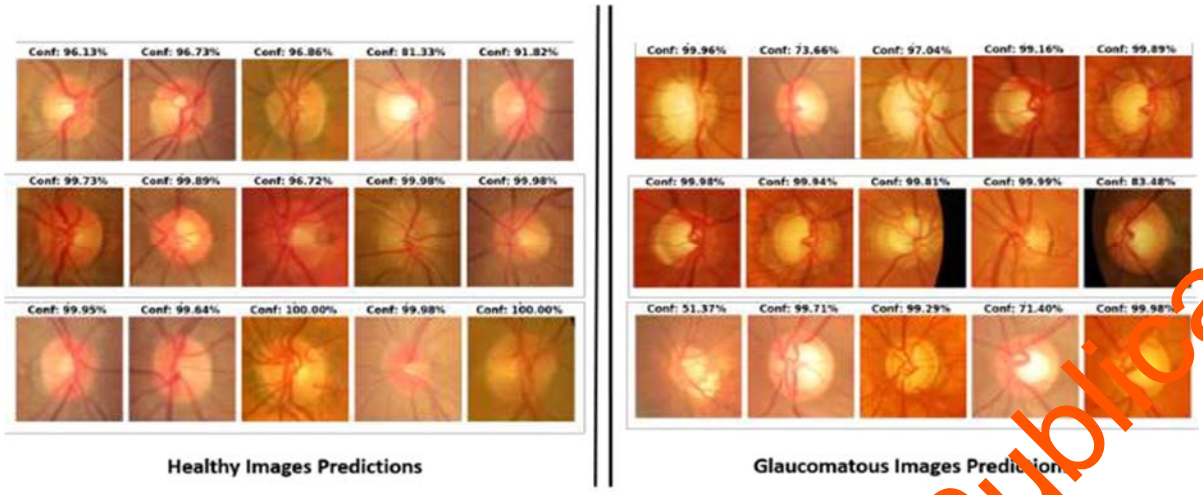

Figure 2: Examples of correctly classified retinal images, using EfficientNetV2-B0 model, along with the prediction score on the ground-truth class.

Submission completed	18-06-2025
1st Revision received	24-10-2025
Last Revision received	29-05-2026
Acceptance	30-05-2026

AUTHORS' CONTRIBUTIONS:

NR & ZS: Drafting, data acquisition, analysis and interpretation.

FS: Concept, drafting and revision.

SFB: Data acquisition, analysis and interpretation.

MA & FA: Concept, design, revision and final approval.