Interpretable Glaucoma Classification and Severity Assessment Using Deep Learning Models

Adla Ananya,
PG Scholar,
Department of Computer Science and Engineering,
Jawaharlal Nehru Technological University Hyderabad
University College of Engineering Jagtial (Autonomous),
Nachupally (Kondagattu), Telangana, India -505501.
edlaananya222@gmail.com

Dr. B. Sateesh Kumar
Professor,
Department of Computer Science and Engineering,
Jawaharlal Nehru Technological University Hyderabad
University College of Engineering Jagtial (Autonomous),
Nachupally (kondagattu), Telangana, India - 505501.
sateeshbkumar@jntuh.ac.in

Abstract— **Glaucoma is among the leading causes of irreversible blindness worldwide and early and accurate diagnosis is essential to avoid irreversible loss of sight. Retinal fundus imaging is a non-invasive method of screening for glaucoma. Manual evaluation is time consuming, needs particular clinical skill and is subjected to inter-observer variability. We aim to develop a single web-based unified computer-aided diagnosis system for glaucoma categorisation, retinal structure segmentation, glaucoma localisation, glaucoma severity evaluation and XAI through the binary glaucoma classification on a publically available retinal fundus image dataset after image pre-processing (resizing, normalisation, contrast enhancement and data augmentation). Multiple classification models such as ResNet50, DenseNet201, ViT, ViT-CapsNet, Xception and ConvNeXtTiny have been compared. For the segmentation of optic disc, optic cup and CDR estimation, EfficientUNet was used and for glaucoma localization, different YOLO based models were used. However, visual explanations were added with the help of Grad-CAM and Grad-CAM++ so as to get a better understanding of the model and improve the clinical clarity. The experimental evaluation was done using Accuracy, Precision, Recall, F1-score, AUC, Dice Score, IoU, Pixel Accuracy and mean Average Precision (mAP). DenseNet201 obtained the highest glaucoma classification accuracy of 89.60%, while YOLOv5s6 obtained the highest localisation accuracy with mAP of 97.50% and EfficientUNet obtained the highest Dice scores of 95.00% for optic disc segmentation and 82.40% for optic cup segmentation. The integrated framework therefore provides an effective and interpretable computer-aided solution to ocular screening which can be employed to make accurate glaucoma diagnosis, structural analysis in retina, localisation of the disease, severity evaluation, explainable prediction and automatic report generation.**

Keywords*— *Glaucoma Detection, Retinal Fundus Images, Deep Learning, DenseNet201, EfficientUNet, YOLOv5s6, Explainable Artificial Intelligence, Cup-to-Disc Ratio.

I. Introduction

Glaucoma, a progressive disease that can lead to permanent vision loss if not treated, is the leading cause of blindness in the world. Diagnosis is important at an early stage to save the visual function and improve the prognosis as the disease may be asymptomatic in the early phase. Because it provides a way to observe structural changes in the optic nerve and surrounding retinal tissue, non-invasive retinal fundus imaging is a regularly used method for glaucoma screening [1]. Recent advances in DL technology has further enhanced the automated analysis of retinal images, enabling computer-aided diagnostic tools to aid the ophthalmologists in quicker and more uniform glaucoma diagnostics as well as minimizing diagnostic variability [2].

Despite the recent advances in automated glaucoma detection, most techniques are focused on disease classification, and provide little support for comprehensive clinical assessment. Some studies have demonstrated satisfactory accuracy in the classification of some retinal diseases using CNN and Transformer based architecture [3] while some studies have investigated optic disc and optic cup segmentation in order to perform structural analysis [4]. The challenge, however, is on how to integrate all these: illness classification, retinal structure analysis, glaucoma localisation and severity estimation, and XAI in a single diagnosis. Furthermore, in some cases, the lack of visual evidence to be interpreted would limit the trust towards the DL based prediction by clinicians [5].

In order to solve such problems, an integrated DL system for automated glaucoma detection from retinal fundus images is proposed. The framework incorporates glaucoma categorisation, retinal structures segmentation, glaucoma localisation, severity estimation based on CDR and XAI, to offer a comprehensive clinical evaluation of a single retinal image. To improve the reliability of diagnosis and to help in clinical interpretation, and also to minimize the manual effort involved in routine glaucoma screening, complementary diagnostic components are proposed to be provided in a single process [6].

The suggested framework adds to the intelligent ocular diagnosis by integrating numerous diagnostic capabilities in a single computer-aided system with an emphasis on prediction transparency and clinical usability. By integrating anatomical analysis, illness localisation and severity evaluation, explainable visualisations, the method provides valuable diagnostic evidence to the clinicians beyond the conventional classification results. This strategy can increase the efficiency of screening, help in clinical decision making, and promote the wider use of AI in the diagnosis of glaucoma [7][8].

II. Related Work

DL is now a powerful technique for automated glaucoma diagnosis to achieve accurate interpretation of retinal fundus pictures. The first computer-aided methods were based on classifying glaucoma using convolutional neural networks trained to differentiate normal and glaucomatous eyes based on fundus pictures. With the growing availability of large retinal imaging datasets and transfer learning algorithms, there has been significant improvement in the diagnosis performance of these methods, which depend less on handcrafted picture characteristics. Many, however, only take the glaucoma diagnosis as a single-label prediction problem and don't give any clues about the underlying structural abnormalities that led to the prediction and so do not help for clinical decision-making [9]. In addition, in some cases the output of the technologies is not interpretable, limiting the comfort level of clinical use during routine ophthalmic care [10].

Anatomical examination of structures of the retina have been added to the diagnosis of glaucoma recently and it has more relevance in the clinical practice. High accuracy in segmentation of the optic disc and optic cup allows quantitative assessment of structural changes associated with glaucoma, including the computation of the CDR, an important clinical parameter used for glaucoma screening and follow-up. DL based segmentation methods have shown promising performance by learning complicated retinal structures directly from fundus pictures, minimising variability related to manual measurements [11]. Although these techniques are useful in providing anatomical information, they typically deal with segmentation as a stand-alone activity and fail to provide a single diagnostic framework for structural analysis in the classification of illnesses, localisation and severity assessment [12].

Another important research direction is the interpretability of DL predictions and localisation of the areas of the retina associated with the disease. YOLO family of models have been proved to be very accurate and retain the real-time nature of detection which makes them interesting for medical image localisation applications [13]. More recently, solutions have been developed to improve the usability, to allow for more efficient computing and to be deployed flexibly, making computer vision applications in healthcare more practical [14]. At the same time, the XAI methods, including the Grad-CAM and the Grad-CAM++ have been extensively used to obtain a visual explanation of the regions of the image that contributed to the CNN prediction, thereby enhancing the transparency of CNN and giving visual evidence to support the clinician in diagnosis [15] [16]. But, localisation and explainability are usually given separately, rather than a part of a whole glaucoma diagnosis system.

Research in automated glaucoma screening has made tremendous progress with the availability of publicly available retinal imaging datasets, which help standardise training and evaluation of DL algorithms for glaucoma diagnosis. The availability of extensive fundus image databases with annotations has led to objective benchmarking and has facilitated the development of more comprehensive diagnostic algorithms that encompass a range of retinal features [17]. Moreover, a number of survey studies have examined how glaucoma detection has progressed from handcrafted feature engineering to DL-based detection methods and that the methods need to be robust, interpretable and clinically validated for real-world deployment [18]. These reviews generally reveal that although significant progress has been made in the diagnosis of certain diagnostic tasks, very little research has addressed an overall solution to aid the whole glaucoma screening process [19].

The available literature has shown great development in glaucoma classification, retinal structural segmentation, object localisation and XAI. However, much of the current research only focuses on a single or two diagnostic tasks in isolation, resulting in incomplete solutions with limited clinical input. Despite this, a need still exists for a single workflow of CAD frameworks that encompass the correct categorisation of glaucoma, the analysis of retinal structure, the location of the disease, the severity of the disease, and the visual explanation. The filling of this gap can improve the transparency of diagnosis, clinical usability and everyday use of intelligent glaucoma screening technologies [20].

III. Materials And Methods

A. Overview

We provide a method that offers an integrated DL pipeline for automated glaucoma detection using retinal fundus pictures. The methodology involves preparing the dataset, preprocessing the images, glaucoma classification, optic disc and optic cup segmentation, glaucoma region localisation and severity assessment using CDR, and XAI for interpreting the forecast. We perform a comparative analysis of multiple DL models for every diagnostic task to find the best suited architecture for good performance in classification, segmentation and localisation. The segmented retinal structures are used to estimate CDR which helps in automated glaucoma severity evaluation. To increase the transparency of prediction, explanatory approaches are used to give visual evidence. Finally, the developed framework is implemented as a web-based clinical decision support system, which enables the automated analysis of retinal images, the visualisation of the results, the management of predictions and the generation of reports, providing an accurate, interpretable and practical solution for glaucoma screening.

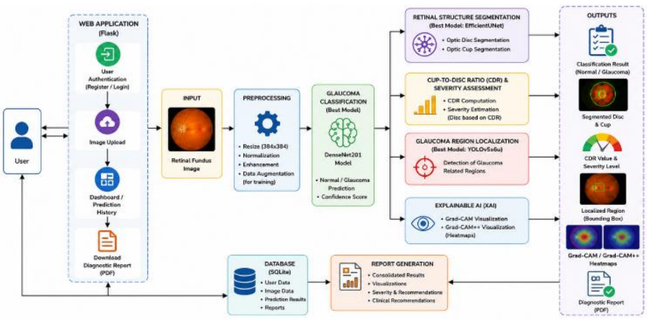

Fig. 1. System Architecture

The whole process of the proposed glaucoma diagnosis system is shown in Fig. 1. The method starts with acquisition and preprocessing of retinal fundus image, followed by glaucoma classification, optic disc and optic cup segmentation, glaucoma region localisation and severity assessment based on CDR. The resulting diagnostic results are then further processed by XAI methods and visualized and output via a web-based application to be used for automated analysis, visualization, prediction management and clinical report generation.

B. Dataset

The proposed system is based on a public benchmark dataset called the SMDG-19, which is created by combining and standardising the retinal fundus images from 19 datasets of glaucoma patients. The dataset comprises colour fundus images with clinically relevant metadata such as glaucoma labels, optic disc segmentation masks, optic cup segmentation masks, blood vessel segmentation masks and patient attributes (where available) including age and sex. SMDG-19 is a comprehensive and structured dataset, which can be used as a benchmark for building and testing glaucoma diagnosis algorithms using AI.

The training set for the glaucoma classification contains 3,920 glaucoma and 6,039 normal photos, and the testing set has 980 glaucoma and 1,510 normal images . There are also several types of annotations, allowing for retinal structure segmentation and glaucoma localisation, thus providing a complete assessment of various diagnostic tasks in a single framework. The dataset provides a diversity of retinal pictures and their annotations which makes it suitable for robust glaucoma diagnosis and comparison of performance.

C. Data Preprocessing

To have a consistent input format for all the diagnostic tasks and to have improved data quality, the retinal fundus images are preprocessed in a series of steps before model training. First, the photos are downsized to a fixed resolution, so that there is uniformity during model training and inference. The intensities of the images obtained by different imaging devices and at different settings are different, and the pixel values of these images are normalised to facilitate the learning of robust features throughout the image set. These pre-processing procedures improve the performance of the computation processes and make sure the input into the DL models are standardised.

In order to enhance the diversity of training data, and thus to improve the model's generalisation, a few data augmentation techniques are used, including random horizontal flipping, rotation, brightness manipulation and scaling. The changes simulate fluctuations commonly encountered in real retinal imaging, and greatly reduce the danger of overfitting. These are organised into annotation resources for each diagnostic task like binary labels for glaucoma classification, segmentation masks for optic disc and OC extraction, and bounding box annotations for glaucoma region localisation. Finally, the created data set is divided into training, validation, and testing data sets to build a reliable model, to tune the parameters, and to evaluate the model in an unbiased manner in an unexpected clinical scenario.

D. Classification Models

This paper proposes to comparatively analyze six state-of-the-art image classification models of retinal fundus images to select the most effective DL model for glaucoma diagnosis. The models are chosen based on their reported performance in medical imaging analysis, and their capacity to learn the discriminative properties of retina. All the models are trained and tested in the same experimental environment, so as to have a fair comparison. The accuracy, precision, recall, F1-score and AUC are used to measure the performance. The comparison analysis helps to choose the best appropriate classification model which can be incorporated in the proposed framework for glaucoma diagnosis.

ResNet50: ResNet50 implements residual learning with skip connections which helps to train deep neural networks better by relieving the vanishing gradient problem. It has an architecture that allows the extraction of much of the characteristics in retina and will be used as a good baseline in categorising glaucoma.

DenseNet201: DenseNet201 is based on Dense Connectivity between the successive layers, thus enabling efficient reusing of features and enhancing gradient flow. Optimized redundant learning and features representation can be obtained by the architecture to get better classification performance for retinal fundus image analysis.

Vision Transformer (ViT-Base): The Vision Transformer (ViT-Base) is a picture classification model that partitions retina photographs into patches of a hardcoded dimension and teaches the worldwide contextual connections between these patches by way of self-attention. This enables modelling of long range dependency efficiently and also rich feature representation.

Vision Transformer–Capsule Network (ViT-CapsNet): ViT-CapsNet is a fusion of Vision Transformers' global feature learning capability and Capsule Networks' ability to model spatial relationships. This hybrid architecture is trying to preserve the anatomical information and improve the discrimination of features of the retina related to glaucoma.

Xception: The depthwise separable convolutions are the backbone of Xception, enabling features to be efficiently extracted with lower computing footprint. It manages to learn fine-grained retinal features without sacrificing good classification accuracy over a vast range of medical imaging applications.

ConvNeXtTiny: The context of retinal fundus analysis is ConvNeXtTiny, an architectural variant of vanilla convolutional neural networks with the latest architectural design concepts that facilitate feature learning, speed up the computation, and achieve competitive picture classification performance.

E. Optic Disc and Cup Segmentation

It is essential to segment the optic disc and optic cup accurately to make quantitative assessments of glaucoma, where they are used to estimate the CDR, a key clinical indicator of glaucomatous damage. In the proposed structure, three DL-based segmentation models, UNet++, ResUNet and EfficientUNet are comparatively evaluated to determine the best performing DL model for the retinal structure segmentation. The annotated optic disc and optic cup masks are used to train the models and the performance of the models in the segmentation task is tested using DSC, IoU and Pixel Accuracy.

UNet++: To reduce the semantic gap between encoder feature maps and decoder feature maps, UNet++ employs a layered skip connection, which enables more features of retinal anatomy to be extracted. The architecture is able to extract fine boundary information effectively, and serves as a good baseline system for optic disc and optic cup segmentation.

ResUNet: ResUNet is a U-Net based network with a residual learning approach which improves the propagation and optimisation of the deep network. The residual connections enable the stable learning of the model and retain the structural information, and thus result in better segmentation accuracy.

EfficientUNet: The segmentation is produced in an efficient manner by taking advantage of EfficientNet's lightweight feature extraction and U-Net's encoder-decoder structure, thanks to its EfficientUNet. It has the capability of learning local and global properties of the retina and thus it is the best performing segmentation model in the proposed framework.

F. Glaucoma Localization

The process of identifying and marking the glaucoma associated regions in the retina is termed as glaucoma localisation, and it gives the spatial information to the glaucoma classification. The proposed framework is based on four models of object detection using the YOLO concept: YOLOv5s6, YOLOv8, YOLOv11, and YOLO26 for the localization of glaucoma-related regions in retinal fundus pictures. The models are trained using annotated bounding box labels and evaluated by the metrics of Precision, Recall and mAP. A comparative evaluation helps to choose the best detector for accurate and reliable localisation of the glaucoma.

YOLOv5s6: YOLOv5s6 is a compact, one-stage object detection model with a moderate level of accuracy and performance. Thanks to its optimised architecture, it is able to localise precisely the retinal regions relevant to glaucoma, while providing a quick inference. It is the optimum localization model of the proposed framework.

YOLOv8: YOLOv8 offers architectural changes to boost feature extraction and object localisation capabilities. It has an efficient detection pipeline, which results in high localization accuracy and real-time inference for retinal image processing.

YOLOv11: YOLOv11 takes advantage of advanced feature representation and optimisation methods to improve object detection performance in challenging imaging scenarios. The model can also localise the retinopathy areas of interest to glaucoma in various anatomical and image conditions.

YOLO26: YOLO26 is suggested to enhance the resilience of multi-scale feature learning and localisation by logging the ocular features at a number of spatial resolutions. The proposed architecture is used as an additional baseline for comparative evaluation within the proposed framework and is also able to provide competitive detection performance.

G. CDR-Based Severity Assessment

The CDR is a therapeutically useful measure to quantify glaucomatous damage by evaluating its relative size of the optic cup to the optic disc. The proposed framework automatically calculates the vertical CDR based on the segmented retinal structures after segmenting the optic disc and optic cup. The computed CDR is an objective indicator of optic nerve head changes and allows automated assessment of the severity of glaucoma, helping to assess the development of the disease.

The CDR is calculated as:

$$CDR = \frac{Vertical\ Cup\ Diameter}{Vertical\ Disc\ Diameter} \tag{1}$$

The classification of the retinal pictures to the different degrees of glaucoma severity is done by this system and relies on the CDR value calculated and clinical thresholds determined. A CDR value is generally a reflection of the normal retinal structure and a more progressive optic nerve damage, indicative of glaucoma, is reflected by larger values. The proposed system provides a more reliable automated diagnosis and provides a comprehensive and clinically-relevant evaluation of glaucoma severity by combining quantitative CDR analysis with classification, segmentation and localisation results.

H. Explainable Artificial Intelligence

To enhance the transparency and interpretability of glaucoma diagnosis, the suggested system combines the XAI methods, namely, Grad-CAM and Grad-CAM++. These methods present maps of class activation that highlight regions of the retina that most influenced the class decision and thus allow a visual interpretation of the model's predictions. The generated heatmaps enable doctors to determine if the model focuses on clinically significant anatomical structures, enhancing the trust in the automated diagnostic process.

Grad-CAM: Grad-CAM uses the gradient details of the final convolutional layers to produce coarse localization maps indicating the discriminative areas that are pivotal for the categorisation of glaucoma. The visualisation offers an intuitive evidence to support the model's prediction and helps in the clinical interpretation.

Grad-CAM++: In addition to the 1st order gradient information, they use higher order gradient information, which yields more accurate and detailed localisation maps than Grad-CAM. This improvement results in better quality visualisation particularly for images with many discriminative retinal regions, and consequently better diagnostic explanations that are reliable and interpretative.

I. Web-Based Deployment

The proposed glaucoma diagnosis system is realized as a web platform on the Flask framework, for easy and user-friendly access to automated retinal image processing. The application includes a safe user registration, upload of the retinal fundus image, automatic prediction and visualisation of the diagnostic results with an interactive user interface. Effective data storage and retrieval is achieved with a light-weight SQLite database, which is used for storing user information, prediction record and analysis history.

After uploading a retinal fundus image, the application processes the whole diagnostics pipeline, including glaucoma categorisation, optic disc and optic cup segmentation, glaucoma localisation, severity assessment from the CDR and explainable visualisation using Grad-CAM and Grad-CAM++. Results are presented in a structured format and can be printed out in a clinical report. The web-based version offers efficient clinical decision support by incorporating many of the diagnostic tools into one convenient platform for standard glaucoma screening.

IV. Experimental Results

A. Performance Metrics

To evaluate the performance of the proposed glaucoma diagnosis framework, commonly recognised classification criteria are used to completely analyse the prediction capabilities and the model dependability. The accuracy is the proportion of all correctly classified samples, Precision is the proportion of positive samples correctly classified, Recall is the proportion of positive samples correctly classified, F1-Score is the harmonic mean of Precision and Recall and Area Under the ROC-AUC is a measurement of the model's ability to differentiate between the classes of glaucoma and normal at varying decision thresholds. These measures provide a complete assessment of categorisation performance all in all.

Accuracy:

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

Precision:

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

Recall:

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

F1-Score:

$$F1\ Score = \frac{2 \times (Precision \times Recall)}{Precision + Recall} \quad (5)$$

AUC-ROC Curve:

$$AUC = \sum_{i=1}^{n-1} (FPR_{i+1} - FPR_i) \times \frac{TPR_{i+1} + TPR_i}{2} \quad (6)$$

The performance of the various DL models is analyzed using these assessment metrics to determine the best suited model for automated glaucoma diagnosis.

B. Classification Results

Six DL models were trained and tested under the same experimental conditions to evaluate the performance of the proposed glaucoma classification framework. The models are evaluated with Accuracy, Precision, Recall, F1-Score and ROC-AUC to find the most appropriate architecture for automated glaucoma diagnosis.

Table 1. Classification Performance Comparison

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	ROC-AUC (%)
ResNet50	89.20	89.20	89.20	89.10	95.20
DenseNet201	89.60	89.80	89.60	89.50	96.10
ViT-Base	87.60	87.70	87.60	87.50	94.10
ViT-CapsNet	60.80	58.70	60.80	47.70	57.90
Xception	88.80	88.80	88.80	88.80	95.80
ConvNeXtTiny	88.40	88.40	88.40	88.20	94.80

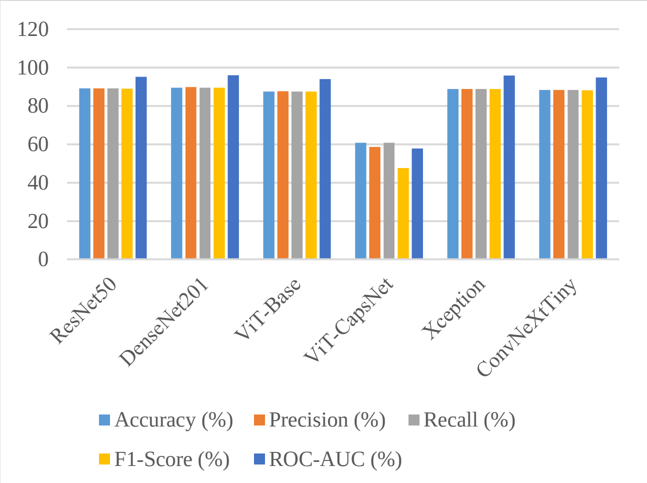

Fig. 2. Comparison of Classification Performance

TABLE 1 AND FIGURE. This figure shows the comparative study conducted using six DL models for glaucoma classification. Overall, the DenseNet201 model achieved the highest scores, demonstrating strong classification performance with an Accuracy of 89.60%, Precision of 89.80%, Recall of 89.60%, F1-Score of 89.50%, and the highest ROC-AUC score of 96.10%, indicating its superior discriminative ability. Other models like ResNet50, Xception and ConvNeXtTiny were able to obtain somewhat competitive results with a score of greater than 88%, while ViT-Base was able to get a slightly lower score. In comparison, ViT-CapsNet showed much lower performance on all metrics, which means it is not very useful for this dataset. The experiment results reveal that DenseNet201 provides the most powerful and reliable representation of features and is, therefore, the most suitable classification model for the proposed automated glaucoma diagnosis system.

C. Segmentation Results

We tested the segmentation ability of the proposed framework for the optic cup and optic disc, utilising three DL architectures: UNet++, ResUNet and EfficientUNet. The models were evaluated using Dice Score, IoU and Pixel Accuracy which are used to measure the degree of segmentation overlap, region similarity and pixelwise prediction accuracy, respectively.

Table 2. Segmentation Performance Comparison

Structure	Model	Dice Score	IoU	Pixel Accuracy
Optic Cup	UNet++	0.786	0.674	0.998
	ResUNet	0.793	0.685	0.998
	EfficientUNet	0.824	0.723	0.999
Optic Disc	UNet++	0.936	0.890	0.998
	ResUNet	0.932	0.882	0.998
	EfficientUNet	0.950	0.908	0.998

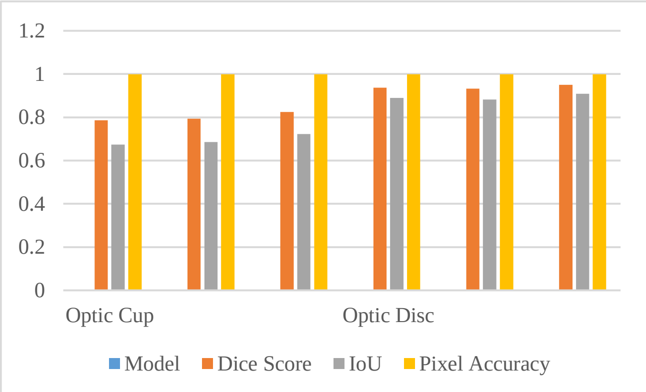

Fig. 3. Comparison of Segmentation Performance

Table 2 and Fig. . 3 compare the performance of the tested models for optic disc/optic cup extraction . Constant best segmentation accuracy for both anatomical structures was achieved by EfficientUNet. For optic cup segmentation, it achieved Dice Score 0.824, IoU 0.723 and Pixel Accuracy 0.999 surpassing UNet++ and ResUNet by generating more accurate cup borders. Likewise, in the optic disc segmentation, EfficientUNet demonstrated the highest results with Dice Score of 0.950 and IoU of 0.908, indicating better ability to delineate the optic disc area while retaining high pixel accuracy of 0.998. The results indicated that EfficientUNet can effectively learn fine-grained boundary details and global retinal context, and thus is best suited to compute the CDR in the proposed framework and to provide accurate glaucoma severity grading.

D. Localization Results

The proposed glaucoma detection system's localisation performance was compared against four state-of-the-art object detection models namely YOLOv8, YOLOv5s6, YOLOv11 and YOLO26. Precision, Recall, and mean Average Precision (mAP) were used to assess the localisation accuracy, the detection completeness and the overall performance of the object detection for glaucoma-related retinal structures.

Table 3. Localization Performance Comparison

Model	Precision	Recall	mAP
YOLOv8	0.877	0.910	0.966
YOLOv5s6	0.943	0.927	0.975
YOLOv11	0.915	0.866	0.920
YOLO26	0.845	0.908	0.974

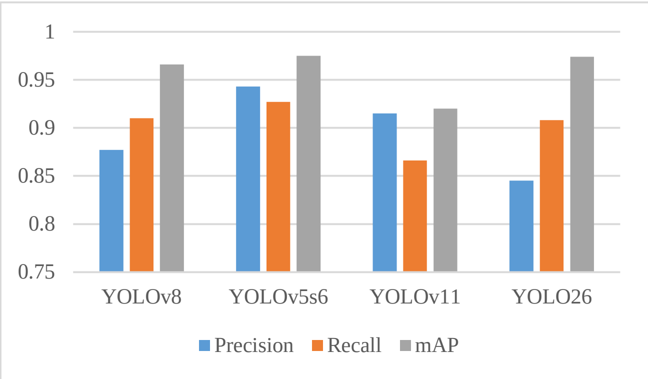

Fig. 4. Comparison of Localization Performance

The localisation performance of the object detection models studied is depicted in Table 3 and Fig. 4. YOLOv5s6 showed the best holistic performance with the highest Precision (0.943), Recall (0.927), and mAP (0.975), showing higher accuracy in identifying glaucoma-related retinal areas while keeping a good detection ability. YOLO26 obtained a competitive mAP of 0.974, though the precision was slightly lower but the localisation was good. YOLOv8 had good results, achieving mAP=0.966. YOLOv11 had the lowest recall and mAP. Experimental results show that the YOLOv5s6 model achieves the best localization accuracy and reliability, and thus is selected as the object detection model for integration into the proposed automated glaucoma diagnosis framework.

E. Qualitative Analysis

For visually evaluating the performance of the proposed glaucoma diagnosis paradigm, representative cases were selected. Figure 5 shows a mild glaucoma case while Figure 6 shows a moderate glaucoma case. We display the original retinal fundus image, the illness category (and the associated confidence score) predicted by the model, and the model's Grad-CAM visualisation for each case. The reliable and interpretable classification results are complemented by the Grad-CAM heatmaps that visually demonstrate the model's attention to clinically significant areas of the ONH. The qualitative results confirm the quantitative results presented in the preceding sections which shows the effectiveness of the proposed framework for automated glaucoma diagnosis.

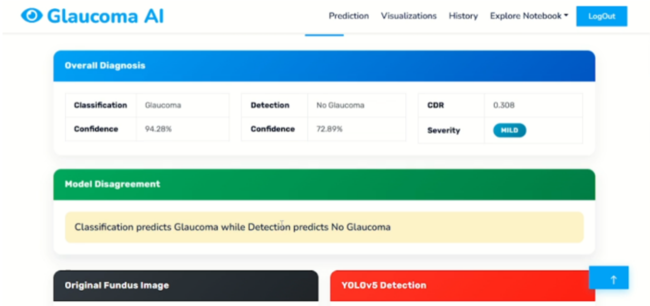

Fig. 5. Mild Glaucoma Case

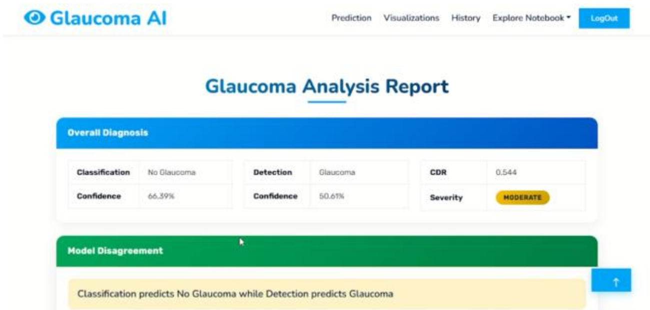

Fig. 6. Moderate Glaucoma Case

F. Discussion

The experimental findings indicate that the proposed framework has the potential to deliver competitive performance when compared to the existing glaucoma diagnosis approaches which were integrated with classification, localisation, segmentation, severity evaluation and XAI in a single pipeline. DenseNet201, EfficientUNet and YOLOv5s6 showed superior performance to other models studied and provided accurate diagnosis and reliable analysis of the anatomy of the retina. The regions highlighted by using the Grad-CAM aid in transparency of the model, which is useful for interpretability to healthcare experts. Nevertheless, the system was tested with only one benchmark data set, which would not be generalizable to other clinical situations. Future study will include multi-center validation, more heterogeneous datasets and real-time deployment to further increase the resilience and clinical usefulness.

V. Conclusion

An integrated DL architecture of automated glaucoma detection using retinal fundus pictures was proposed in this paper. The proposed method is a hybrid technique of glaucoma classification, localisation of retinal structure, segmentation of optic disc and cup, severity evaluation based on CDR, and XAI, as it provides a complete CAD solution. As per the experimental results, DenseNet201 has been observed to have the best classification results with an accuracy of 89.60% and a ROC-AUC of 96.10%. EfficientUNet produces the most accurate optic disc and cup segmentation, and YOLOv5s6 has the best localisation performance with a mAP of 97.50%. Grad-CAM visualisations to identify clinically significant retinal areas improved model interpretability. To summarise, the presented approach offers an accurate, interpretable and practical alternative to support ophthalmologists in glaucoma screening and facilitate quick clinical decision making.

For future study, this system will be tested on larger multi-center datasets of retinal images to further improve the generalisability of the system to other clinical situations. The system can be generalized to multi-stage glaucoma classification, multi-modal imaging of the eye (such as OCT)), and be deployed on the cloud or mobile devices for real-time glaucoma screening applications. In addition, the combination of advanced XAI approaches with continuous model adaption using newly acquired clinical data can significantly boost the diagnostic reliability and enable greater deployment in teleophthalmology and ordinary clinical practice.

REFERENCES

[1] V. Vigneshwaran, S. Kumar E., and R. Shoba, "Automated Glaucoma Detection Using Vision Transformers and Capsule Networks: A Deep Learning Approach for Early Diagnosis from Retinal Fundus Images," in *Proceedings of the 2025 International Conference on Data Science, Agents and Artificial Intelligence (ICDSAAI)*, Chennai, India, 2025.

[2] H. Wang, H. Sun, Y. Fang, S. Li, M. Feng, and R. Wang, "A Workflow for Computer-Aided Diagnosis of Glaucoma," in *Proceedings of the 2022 IEEE International Symposium on Biomedical Imaging Challenges (ISBIC)*, pp. 1–4, 2022.

[3] F. Khader, C. Haarburger, J.-C. Kirr, M. Menke, J. N. Kather, J. Stegmaier, C. Kuhl, S. Nebelung, and D. Truhn, "Elevating Fundoscopic Evaluation to Expert Level: Automatic Glaucoma Detection Using Data from the AIROGS Challenge," in *Proceedings of the 2022 IEEE International Symposium on Biomedical Imaging Challenges (ISBIC)*, pp. 1–4, 2022.

[4] Dr.B.Sateesh Kumar, Detection and Classification of Objects in Satellite Images using Custom CNN, ZKG International, ISBN No.2366-1313, Vol No.VIII, Issue No.II, pp.78-86, Web of Science Group, JULY, 2023.

[5] Y. Madadi, H. Abu-Serhan, and S. Yousefi, "Domain Adaptation-Based Deep Learning Model for Forecasting and Diagnosis of Glaucoma Disease," *Biomedical Signal Processing and Control*, vol. 92, p. 106061, 2024.

[6] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, J. Uszkoreit, and N. Houlsby, "An Image Is Worth 16×16 Words: Transformers for Image Recognition at Scale," in *Proceedings of the International Conference on Learning Representations (ICLR)*, 2021.

[7] A. Sevastopolsky, "Optic Disc and Cup Segmentation Methods for Glaucoma Detection with Modification of U-Net Convolutional Neural Network," *Pattern Recognition and Image Analysis*, vol. 27, no. 3, pp. 618–624, 2017.

[8] R. Hussain and H. Basak, "UT-Net: Combining U-Net and Transformer for Joint Optic Disc and Cup Segmentation and Glaucoma Detection," *arXiv preprint arXiv:2303.04939*, 2023.

[9] Dr.B.Sateesh Kumar, A Benchmark Survey on High Resolution Remote Sensing Images Learning Techniques, International Journal of Techno Engineering, IJTE, ISBN No.2057-5688, Vol No.XV, Issue No.III, pp.23-32, IJTE, JULY, 2023

[10] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, "Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization," in *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, pp. 618–626, 2017.

[11] A. Chattopadhay, A. Sarkar, P. Howlader, and V. N. Balasubramanian, "Grad-CAM++: Improved Visual Explanations for Deep Convolutional Networks," in *Proceedings of the IEEE Winter Conference on Applications of Computer Vision (WACV)*, pp. 839–847, 2018.

[12] G. Jocher, A. Chaurasia, and J. Qiu, "Ultralytics YOLO," Ultralytics, 2023.

[13] J. Redmon and A. Farhadi, "YOLO9000: Better, Faster, Stronger," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 7263–7271, 2017.

[14] H. G. Lemij, C. de Vente, C. I. Sánchez, and K. A. Vermeer, "Characteristics of a Large, Labeled Dataset for the Training of Artificial Intelligence for Glaucoma Screening with Fundus Photographs," *Ophthalmology Science*, vol. 3, no. 3, p. 100300, 2023.

[15] A. Septiarini and A. Harjoko, "Automatic Glaucoma Detection Based on the Type of Features Used: A Review," *Journal of Theoretical and Applied Information Technology*, vol. 72, no. 3, pp. 314–328, 2015.

[16] E. Arrieta-Rodriguez, D. Morales, J. M. Molina, and J. Hernández, "A Comprehensive Review on Fundus Imaging Approaches for Glaucoma Detection," *IEEE Reviews in Biomedical Engineering*, vol. 18, pp. 154–178, 2025.

[17] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 770–778, 2016.

[18] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely Connected Convolutional Networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 4700–4708, 2017.

[19] Z. Liu, H. Mao, C.-Y. Wu, C. Feichtenhofer, T. Darrell, and S. Xie, "A ConvNet for the 2020s," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 11976–11986, 2022.

[20] F. Chollet, "Xception: Deep Learning with Depthwise Separable Convolutions," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1251–1258, 2017.