AI-Based Glaucoma Assessment Tool with Chatbot Diagnostics

Chandradeep Reddy Edurinty, Dudekula Sai Mallesh, RamaRao, Gedala Hemanth Kumar, Sajitha Krishnan*
Department of Computer Science and Engineering
Amrita School of Computing Bengaluru
Amrita Vishwa Vidyapeetham, India.
*Corresponding Author: k_sajitha@blr.amrita.edu

***Abstract*—Glaucoma is one of the primary reasons for permanent blindness, as people usually cannot recognize its presence because it exists without indications. The system presents a deep learning-based approach to perform early diagnosis along with assessments. A deep learning model comprised of ResNet50 together with VGG16 alongside EfficientNetB0 and MobileNetV2 determines between normal eye condition and cataracts, as well as diabetic retinopathy and glaucoma. Humphrey Visual Field (HVF) analysis becomes necessary when glaucoma diagnosis occurs. The VQA system leverages clinical training to evaluate HVF results and present information about vision deficit and disease advancement to healthcare professionals. The integration of deep learning classifying algorithms with VQA systems results in a system that enables rapid diagnosis, increases clinical decision-making support and makes the normal glaucoma testing more accessible.**

Keywords— Visual Question Answering, Glaucoma, deep learning, HVF analysis.

I. Introduction

Glaucoma is a progressive eye disease that causes damage to the optic nerve, which often leads to irreversible loss of sight if not detected early. It impacts about 8.5 percent of the global population, mostly with undiagnosed cases of a significantly higher rate in low- and middle-income countries. In India, 9 out of 10 cases of glaucoma are estimated to go undiagnosed mainly due to a severe lack of awareness and inadequate access to proper diagnostics. The major obstacle to diagnosing glaucoma is its asymptomatic stages. By the time vision loss is apparent, considerable and permanent damage to the optic nerve has already occurred.

The traditional methods used in the diagnosis of glaucoma is the fundus imaging and the Humphrey Visual Field (HVF) testing. These techniques need high cost machinery and experienced eye specialist to interpret result properly. However these resources are not available freely in rural and deprived settings, (complicating whilst not rather a lifeless delay in enough remedy) The requirement for an accessible and cost-effective technique for the early glaucoma detection is of high importance for preventing the preventable blindness.

We presents a system that employ deep learning based image classification approach to diagnose glaucoma and a structured Visual Question and Answering (VQA) system to aid in interpreting HVF reports. The classification model are evaluated using Convolutional Neural Networks (CNNs) like ResNet50, VGG16, EfficientNetB0, MobileNetV2 to classify the eye conditions as normal, cataract, diabetic retinopathy, and glaucoma. if glaucoma is confirmed then system advise for further investigation with HVF and VQA module helps in extracting meaningful information from HVF reports.

The proposed system presents an efficient and simple way of glaucoma detection and evaluation through integrating of deep learning-based classification with an automatic HVF interpretation process. This method is designed to enhance early detection, support doctors in clinical decision according to day by day medical standards, and bring a gap in glaucoma detection, particularly for the medical aid of certain low resource nation.

II. Literature Survey

Some studies focused on the sophisticated methodologies for glaucoma detection, such as deep learning, federated learning and Visual Question Answering (VQA) model. Traditional methods like Optical Coherence Tomography (OCT) and fundus imaging have gone forward and integrated with machine learning-based techniques for boosting the accuracy and enhance of the accessibility.

Rasel et al. [1] for instance are compared the performance level of 2D & 3D CNN models in OCT images-based glaucoma disease recognition. Their work found that the 3D CNN models have captured volumetric features more effectively than 2D CNNs result in better performances in terms of sensitivity and specificity for early detection. Choi [2] has investigated multi-objective molecular optimization methods with self-critically sequence training which was although based on biological application nonetheless gave ideas of optimization strategies that is helpful increase a model performance in the medical image analysis.

Aljohani and Aburasain [3] recommended a mixed framework that combined federated learning with deep learning models to spot glaucoma. Through their method, they guaranteed the data privacy during the collaborative model training among the several healthcare establishments and reacted to various other issues associated with sensitive patient data. The results showed putting federated learning being accurate to a great extent and chasing the ban on risks that come with saving information in a central spot. Also, Kashyap et al. [5] an improved U-Net architecture with the help of fundus images for glaucoma detection. Their model performed better segmentation out of accuracy and showed them that deep learning architecture can do a better job than other methods at distinguishing normal from glaucomatous eyes.

The use of fundus imaging-based AI techniques has been demonstrated to possess a considerable promise in automated glaucoma detection. Sidhu and Mansoori et al. [4] evaluated the effectiveness of AI systems for diagnosis of glaucoma by analysis of colour fundus photographs. Their work highlighted the need for dataset diversity and feature extraction for good overfitting in different populations. Fong et al. [10] undertook review the long term visual outcomes of primary congenital glaucoma in china and find out the critical factors in progression of disease. Their conclusion highlights the importance of early treatment and continuous follow-up to safeguard vision in the diseased patients. Lin et al. [6] practised a model to synthesis several vision measures to forecast visual disable in glaucoma patients. Their research contributed significant information on how integrated assessment methods can enhance prognosis prediction, and therefore, guides management of patients.

Zhang et al. [9] built a big medical VQA dataset which supports the training of question-answering models for clinical decision support. Their work emphasized the importance of structured data gathering in making VQA model better. Chen et al. [7] propose an R-LLaVA model, that boosts Med-VQA comprehension by searching expeditious to some particular visual districts of involvement in medical camera. This method considerably improves interpretability and validity of the diagnosis. Qi et al. [8] gave a comprehensive overview towards medical vision-and-language applications, describing new tendencies and challenges in automation of medical evaluation. Their research highlighted the benefits of multi-modal learning for enhancing the diagnostic decision-making.

These studies alone demonstrated that deep learning and VQA models are effective in achieving glaucoma detection and disease monitoring. The proposed system further leverages these developments to, for the first time within the scope of VQA, integrate deep learning classified and structured VQA to make high assistivity, enable the doctor to enhance in the clinical decision-making and provide early glaucoma detection possible in resource-limited environment.

III. Dataset

The study employs HVF-VQA as a newly developed visual question answering dataset that specializes in the interpretation of Humphrey Visual Field (HVF) tests essential for glaucoma progression evaluations. HVF-VQA functions as the initial documented VQA database which targets ophthalmology interpretation while avoiding natural scene or radiological scan content. The development of HVF-VQA emerged because medical professionals require AI systems to assist ophthalmologists through explainable visual field interpretation delivered as question-answer pairs.

	A	B	C	D
1	Image Name	Image Location	Question	Answer
2	Patient_1.png	C:\Users\hemanth\Documents\	Are there any areas with moderate visual	No
3	Patient_1.png	C:\Users\hemanth\Documents\	Are there any areas with severe visual fiel	Yes
4	Patient_1.png	C:\Users\hemanth\Documents\	Is the percentage of overall loss high?	No
5	Patient_1.png	C:\Users\hemanth\Documents\	Is there any 0 dB in the image?	Yes
6	Patient_1.png	C:\Users\hemanth\Documents\	Is the loss clustered?	No
7	Patient_1.png	C:\Users\hemanth\Documents\	Is there a severe loss?	Yes
8	Patient_1.png	C:\Users\hemanth\Documents\	Does the vision loss indicate early glaucor	No
9	Patient_1.png	C:\Users\hemanth\Documents\	Does the vision loss indicate advanced gla	No
10	Patient_1.png	C:\Users\hemanth\Documents\	Does the patient have glaucoma?	Yes
11	Patient_1.png	C:\Users\hemanth\Documents\	Is any value lower than 20 present in the f	Yes
12	Patient_1.png	C:\Users\hemanth\Documents\	Is there any moderate loss?	No
13	Patient_10.png	C:\Users\hemanth\Documents\	Are there any areas with moderate visual	Yes
14	Patient_10.png	C:\Users\hemanth\Documents\	Are there any areas with severe visual fiel	Yes
15	Patient_10.png	C:\Users\hemanth\Documents\	Is the percentage of overall loss high?	No
16	Patient_10.png	C:\Users\hemanth\Documents\	Is there any 0 dB in the image?	No
17	Patient_10.png	C:\Users\hemanth\Documents\	Is the loss clustered?	Yes
18	Patient_10.png	C:\Users\hemanth\Documents\	Is there a severe loss?	Yes
19	Patient_10.png	C:\Users\hemanth\Documents\	Does the vision loss indicate early glaucor	Yes
20	Patient_10.png	C:\Users\hemanth\Documents\	Does the vision loss indicate advanced gla	No
21	Patient_10.png	C:\Users\hemanth\Documents\	Does the patient have glaucoma?	No
22	Patient_10.png	C:\Users\hemanth\Documents\	Is any value lower than 20 present in the f	Yes
23	Patient_10.png	C:\Users\hemanth\Documents\	Is there any moderate loss?	Yes
24	Patient_100.png	C:\Users\hemanth\Documents\	Are there any areas with moderate visual	No
25	Patient_100.png	C:\Users\hemanth\Documents\	Are there any areas with severe visual fiel	Yes
26	Patient_100.png	C:\Users\hemanth\Documents\	Is the percentage of overall loss high?	No
27	Patient_100.png	C:\Users\hemanth\Documents\	Is there any 0 dB in the image?	Yes
28	Patient_100.png	C:\Users\hemanth\Documents\	Is the loss clustered?	No

Fig. 1. HVF-VQA dataset

The University of Washington Humphrey Visual Fields (UWHVF) serves as the data source for constructing the publicly available dataset. The procedure carefully handle each HVF-VQA data sample to establish its standardization. The visual field data gets processed into grayscale images which simulate standard printout formats to allow AI models training under proper medical execution contexts. The JSON metadata contains structured information about patient characteristics together with test setup parameters as well as essential clinical assessment scores.

The unique aspect of HVF-VQA is in the fact that it uses structured QA pairs that are focused on clinical use of visual field interpretation. The QA pairs are designed to resemble a set of real clinical situations, in which the task is made up of delineating vision-affected zones, measuring the degree of injury, and analyzing how visual fields are outlined. The QA pairs we construct come from real clinical diagnostic questions used to analyze eye health, test dependability, and assess glaucoma severity. All the generated QA pairs take the form of binary Yes/No questions, providing a consistent approach and simplifying the VQA task.

The HVF-VQA platform provides two benefits because it preserves assessment clarity from clinical HVF testing while enabling the creation of AI models which answer textual inquiries using natural language. The goal to create trustworthy explainable decision support systems in ophthalmology finds support through this development. The dataset serves two primary objectives that lead to automated visual field analysis and create a connection between raw perimetry data and diagnostic processes.

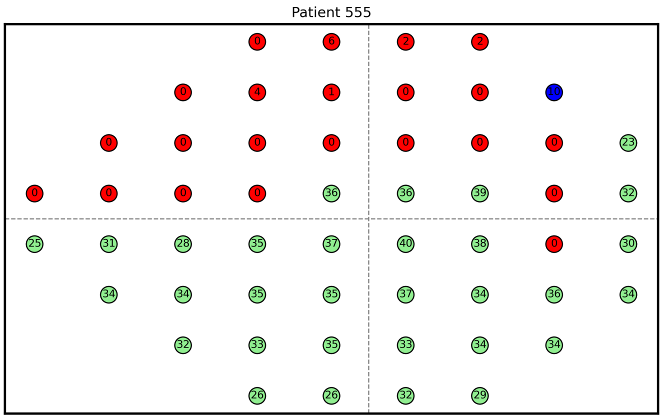

Fig. 2. Heatmap image showing retinal sensitivity for patient 555

As shown in Fig. 2, a grayscale heatmap image has been created based on single-patient HVF sensitivity data. Each circle is the outcome of a measurement from the 24-2 visual field grid; the hue and gradation of the circle represent the patient's retinal response at the very test point being measured. Regions with low values of sensitivity are coloured in shades of red or blue and represent potential eyesight impairment, and green colours reflect a healthy retinal health condition.

This visual display is very important in clinical findings interpretation and in teaching the system of Visual Question Answering (VQA). The image is a good representation of spatial distribution of vision loss as expounded in the corresponding QA pairs presented for each plot. The visual-text information combination used in these plots forms the foundation for the development of the HVF-VQA dataset while shaping the development of the system to mimic clinical diagnosis processes.

In the proposed system, the HVF-VQA dataset to conduct graphical analysis of HVF testing through its visual question answering framework. Model training depends on the essential QA pairs to identify spatial deficits while assessing their severity level and creating simulated clinical diagnostic processes. The system accesses structured data from the University of Washington HVF repository which provides dependable representation of glaucoma-affected and normal visual patterns throughout the dataset.

These datasets with QA annotations that focus on the spatial aspects of visual plots enable the model to discover contextual response methods which produce explainable solutions for detailed clinical inquiries. The structured database enables optimal deep learning architecture training which ensures the system provides immediate assistance for practitioners as well as potential expansion of healthcare diagnostics in underserved communities.

IV. Methodology

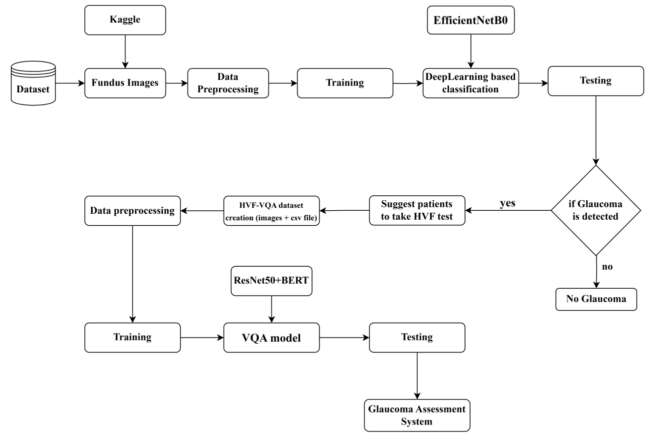

Fig. 3. Workflow of the Glaucoma-VQA system

The proposed system combines deep learning and VQA infrastructure functions to recognize glaucoma cases and help medical personnel with interpretation through visual question answering databases. A sequential approach controls the methodology which includes data preprocessing followed by deep learning image classification and question-answering with Humphrey Visual Field data.

The first step requires pre-processing several fundus images sourced from different datasets. Fundus images undergo processing with different pre-trained Convolutional Neural Networks (CNNs) consisting of ResNet50 and VGG16 and EfficientNetB0 and MobileNetV2. The models receive their evaluation based on accuracy measurements in classifying both eye diseases and healthy states.

The implementation includes development of a specialized CNN model that optimizes glaucoma detection specifically for this kind of medical issue. All CNN models rely on a Softmax classifier as their output layer for multi-class prediction. Patients detected with glaucoma must take a Humphrey Visual Field (HVF) test as a next step. Data from HVF is obtained from the UWHVF dataset which is freely accessible.

The visual field records from HVF testing undergo a conversion process which creates grayscale images that link to clinical QA pairs to establish the HVF-VQA dataset. The established dataset helps doctors detect vision loss patterns to support diagnostic assessment in ophthalmology practice. After the models are trained, the various components of the pipeline receive the inputs serially, therefore, they aim to VQA models for eye disease identification and then they are capable of reading the clinical test results.

The approach is of a simplified two-stage workflow of glaucoma screening incorporating structural imaging and functional vision examination as shown in Fig 4. Users start by clicking on any of two options, “Structural” or “Functional”. During structural phase deep learning model performs analysis of retinal fundus images for glaucomatous damage and gives easy yes/no result. In the event of observation of signs of glaucoma, the users move along with the functional stage, at which HVF images are examined with a VQA model that has been trained on clinical questions. This model determines vision loss severity as well as field symmetry. An embedded chatbot provides interactive, explainable feedback, which makes the system clinically robust and user-friendly with the flexibility for future improvements in diagnostics.

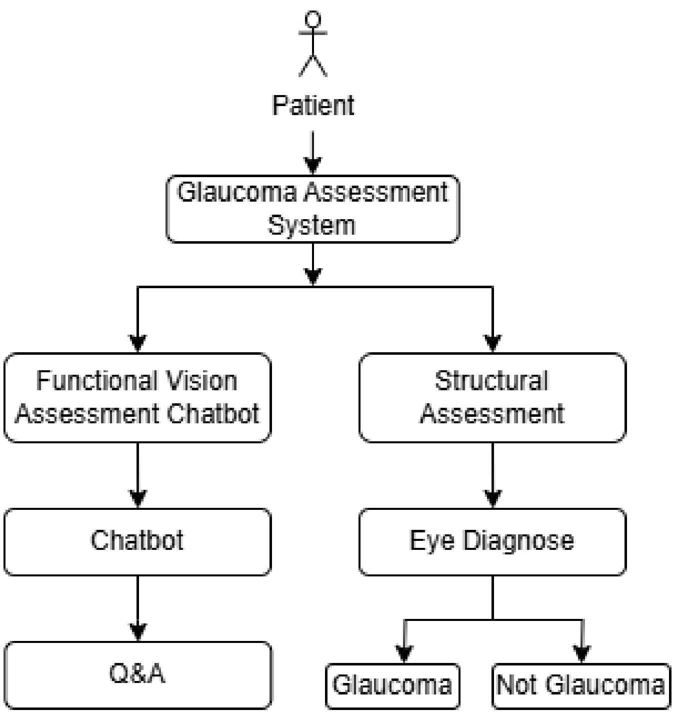

Fig. 4. Flowchart of the Glaucoma Assessment System

To address the task of Visual Question Answering (VQA) in the context of eye condition diagnosis, the proposed system focuses on six different methods in which deep learning models are mixed in order to complete each other. These methods deal with special things, and they are used for particular purposes in the diagnostic pipeline.

A. ResNet50 Model

The ResNet50 model functions as the platform for deriving deep features from fundus images. The framework allows the network to process intricate variations across eye images originating from multiple clinical situations. A classifier functions at the model’s last layer to achieve multi-class classification between normal and cataract and diabetic retinopathy and glaucoma conditions.

B. VGG16 Model

The widely known CNN architecture, VGG16 processes retinal images to detect vital spatial features that aid in eye condition early warning. The use of softmax and this network architecture provides both precise and efficient classification performance for all four types of eye conditions.

C. EfficientNetB0 Model for Optimized Feature Learning

EfficientNetB0 optimizes depth and width scaling to offer an efficient model solution that achieves high performance in image classification tasks. This model shows great utility for real-time processing applications that require limited computation power because it delivers accurate disease identification through softmax-based multi-classification systems.

D. MobileNetV2 Model for Lightweight Classification

MobileNetV2 is integrated to offer a lightweight and fast solution for classifying eye conditions. It uses inverted residual blocks and depthwise separable convolutions, making it ideal for deployment on mobile or edge devices. The final classification is handled by a softmax layer trained on labeled datasets of eye images.

E. Custom CNN Model

The CNN features a specialized design structure to detect Glaucoma and uses adjusted hyperparameters and architecture configuration. For enhanced sensitivity to early vision loss patterns. A customized convolutional and pooling layer system forms part of this model design. The model includes dedicated fully connected layers that lead to a softmax output to achieve precise glaucoma diagnosis.

F. Visual Question Answering (VQA) with HVF Analysis

A deep learning-based VQA model functions to analyze HVF test results. The system helps doctors evaluate vision impairment severity through AI-powered assessments by analyzing images alongside doctor questions specifically in glaucoma situations.

V. Results

The glaucoma detection system was tested using actual fundus images to measure how efficient the system is able to detect indications of glaucoma. Users can upload images using a simple web interface, with the help of trained deep learning models the image is classified in “glaucoma” or “not glaucoma”. The system was able to analyze Normal and diseased eye images successfully, which offered prompt and clear feedback and can be used for first screening.

The system not only gives predictions, but also provides recommendations for next step, such as suggesting a follow up Humphrey Visual Field (HVF) test, in case of diagnosis of glaucoma. Such functionality will assist in early screenings and in avoiding delays in diagnosis, an aspect that is very useful in remote areas where medical assistance is not readily available. On the whole, the system is efficient, convenient and handy for early diagnosis of eye diseases.

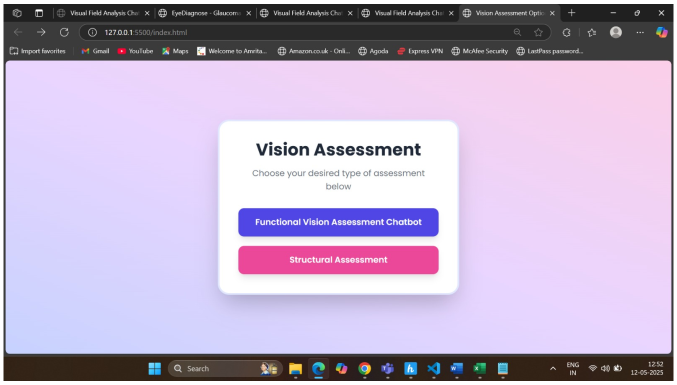

Fig. 5. Structural and Functional options for glaucoma detection and HVF analysis

A web interface for screening glaucoma that is simple and user-friendly was designed to make the process easier. It has two distinguishable buttons “Structural” and “Functional” as shown in Fig 5 With a click on the “Structural” button, the users can upload the retinal fundus images for the analysis by a deep learning model to see the signs of glaucoma. If in case, there are signs, the system states taking the “Functional” test. By clicking this, the users can upload Humphrey Visual Field (HVF) results that are then analyzed through a VQA model to evaluate the vision loss and the affected areas. This two-step process aims at achieving a comprehensive and accurate glaucoma check.

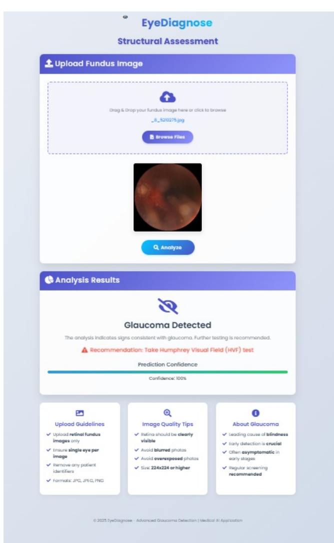

Fig. 6. Glaucoma Detection showing classification result: Glaucoma

The Glaucoma Detection system consists of a web-friendly, adaptive graphical user interface that gives a user the ability to upload retinal fundus images, and get instant diagnosis without delay. As can be seen in Fig 6, the interface contains a straightforward design, and the users can click “Choose File” to upload an image and then “Check” to initiate the analysis. A trained deep learning model processes the uploaded image and estimates the presence of signs of glaucoma.

If glaucoma is diagnosed, the system will immediately show the result and suggest the user to continue with further analysis using the Humphrey Visual Field (HVF) test. The integration of the front-end and back-end functionality guarantees the seamless, real-time screening experience that is operationally practical for clinical use and remote evaluations.

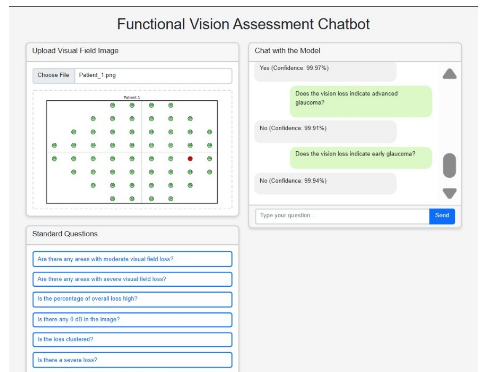

Fig. 7. Functional Vision Assessment Chatbot for HVF-based Q&A

Functional Vision Assessment Chatbot enables its users to upload a visual field image from a Humphrey device click tests for automatic clinical input. After the upload, the chatbot responds to commonly used and user-defined questions concerning the image as shown in Fig 7.

Along with the responses, the chatbot includes confidence scores, which demonstrate how sure the model was about its answers. The users can also select preset questions concerning vision loss patterns, severity as well as early signs of glaucoma. This helps to understand complex HVF results in a clear and interactive way.

Model	Accuracy	Precision	Recall	F1 score
VQA-CNN	89.93	80.30	80.33	80.21
ResNet50	89.57	89.71	89.57	89.38
VGG16	26.78	80.39	26.78	11.31
EfficientNetB0	**93.01**	**93.09**	**93.00**	**92.89**
MobileNetV2	90.52	90.87	90.52	90.54

TABLE I
RESULTS OF DEEP LEARNING MODELS

We presents a comparison of performance for 5 deep learning models i.e. VQA-CNN, ResNet50, VGG16, EfficientNet-B0, and MobileNetV2 used on glaucoma detection as shown in

Table 1. All models were tested on the same dataset according to the metrics such as, Accuracy, Precision, Recall, and F1 Score to predict fundus images into “glaucoma” or “non-glaucoma”. EfficientNet-B0 demonstrated the best result at 93.01% accuracy, as well as balanced precision, recall, and F1-score, which makes it a very safe option. MobileNetV2 and ResNet50 also performed well, whereas VQA-CNN had the slightly lower performance, but yet acceptable. VGG16

performed poorly in all the metrics which would suggest its lack of suitability for this task, possibly resulting from its old architecture. On the other hand, the newer models such as EfficientNet-B0 and MobileNetV2 were more effective at detecting faint patterns in medical images. Due to its consistent and high performance, EfficientNet-B0 was chosen for inclusion in the web-based glaucoma detection tool for fast and accurate real-time predictions for users.

Model	Accuracy	Train Loss	Train Acc	Val Loss	Val Acc
ResNet50+BERT	**98.90**	**0.032**	**98.97**	**0.045**	**98.34**
ResNet18+BERT	98.84	0.029	99.08	0.032	98.84

TABLE II
PERFORMANCE COMPARISON BETWEEN DEEP LEARNING MODELS

The performance evaluation demonstrates the level of accuracy attained by two deep learning models: ResNet50+BERT and ResNet18+BERT. The ResNet50+BERT performed slightly better with an accuracy of 98.90%, whereas, the ResNet18+BERT recorded closely with 98.84%. While both models excelled both models performed, the slightly better accuracy for ResNet50+BERT reveals that this model has a better ability to capture and learn complex patterns, which makes it a better choice for deployment in glaucoma classification systems.

ResNet18+BERT had a lower training accuracy of 98.04% and higher loss but performed well and lagged on final validation accuracy. These results show that ResNet50+BERT is able to capture more complex features needed in accurate macular glaucoma classification. Considering its better accuracy and well-balanced performance, ResNet50+BERT was chosen for the system deployment. This comparison seeks to highlight the need to consider balancing the model complexity against training efficiency for effective real-world application in medical AI.

VI. CONCLUSION

This study introduces a glaucoma screening system that includes several CNN-based models, including ResNet50, VGG16, EfficientNetB0, and MobileNetV2, and also the Visual Question Answering (VQA) module to analyze the Humphrey Visual Field (HVF) results. The system performs very well in identifying eye conditions, and identifying important findings from HVF report data. Although the classification models had a high accuracy, they had some difficulties while processing noisy or incomplete visual field data.

Finally, the comparison of the deep learning models reveals the good performance of ResNet50+BERT in the glaucoma detection with that being more accurate, precise and more recalling. This combination is most suitable for real time role in glaucoma screening systems. The study emphasizes the need to balance the model performance and efficiency in medical applications; therefore, quick, reliable diagnosis for early detection and treatment of eye diseases.

ACKNOWLEDGMENT

We extend our gratitude to Amrita Vishwa Vidyapeetham for providing the necessary resources and infrastructure which helped us finish our paper. We acknowledge this text includes ChatGPT-assisted paraphrasing.

REFERENCES

[1] Rasel, Rafiul Karim, et al. ”Assessing the efficacy of 2D and 3D CNN algorithms in OCT-based glaucoma detection.” *Scientific Reports*, vol. 14, no. 1, 2024, p. 11758.
[2] Choi, Jonghwan. ”Rebadd-se: Multi-objective molecular optimisation using selfies fragment and off-policy self-critical sequence training.” *Computers in Biology and Medicine*, vol. 157, 2023, p. 106721.
[3] Aljohani, Abeer, and Rua Y. Aburasain. ”A hybrid framework for glaucoma detection through federated machine learning and deep learning models.” *BMC Medical Informatics and Decision Making*, vol. 24, no. 1, 2024, p. 115.
[4] Sidhu, Zubin, and Tarannum Mansoori. ”Artificial intelligence in glaucoma detection using color fundus photographs.” *Indian Journal of Ophthalmology*, vol. 72, no. 3, 2024, pp. 408-411.
[5] Kashyap, Ramgopal, et al. ”Glaucoma detection and classification using improved U-Net Deep Learning Model.” *Healthcare*, vol. 10, no. 12, MDPI, 2022.
[6] Lin, Stephanie, et al. ”Predicting visual disability in glaucoma with combinations of vision measures.” *Translational Vision Science & Technology*, vol. 7, no. 2, 2018, p. 22.
[7] Chen, Xupeng, et al. ”R-LLaVA: Improving Med-VQA Understanding through Visual Region of Interest.” *arXiv preprint arXiv:2410.20327*, 2024.
[8] Chen, Qi, et al. ”A survey of medical vision-and-language applications and their techniques.” *arXiv preprint arXiv:2411.12195*, 2024.
[9] Zhang, Xiaoman, et al. ”Development of a large-scale medical visual question-answering dataset.” *Communications Medicine*, vol. 4, no. 1, 2024, p. 277.
[10] Fang, Lei, et al. ”Long-term visual outcomes of primary congenital glaucoma in China.” *Ophthalmic Research*, vol. 65, no. 3, 2022, pp. 342-350.