Fundus Image Classification via an Integrated Deep Learning Model and Random Forest for Glaucoma Diagnostics

Haotian Zeng[1], Jiheng Cong[1], Hantong Hong[1], Zihan Yu[1], Fan Zhang[2][0000−0001−6228−940X], and Guojie Li[2,3,4][0009−0003−2642−598X]

[1] School of Advanced Technology, Xi'an Jiaotong-Liverpool University, Suzhou, China
[2] School of Robotics, Xi'an Jiaotong-Liverpool University, Suzhou, China
[3] Hebei University of Environmental Engineering, Qinhuangdao, China
[4] School of Electrical Engineering, Electronics and Computer Science, University of Liverpool, Liverpool, UK
{guojie.li}@liverpool.ac.uk

Abstract. Glaucoma is a leading cause of irreversible blindness, often exacerbated by delayed diagnosis. Traditional diagnostic methods have limitations in early detection and require significant expertise. Recent advancements in Machine Learning and Deep Learning have shown promise in enhancing the diagnostic accuracy and efficiency for glaucoma using fundus imagery. This study presents a four-stage method for glaucoma classification. Fundus images are processed using VGG16 and ViT for feature extraction, capturing both local and global features. These features are then fused and reduced in dimensionality using Principal Component Analysis. Finally, the reduced features are classified using a Random Forest classifier. The integrated feature fusion model demonstrates significant improvements in diagnostic performance, achieving higher accuracy, specificity, and sensitivity in distinguishing between non-referable glaucoma and referable glaucoma classes compared to traditional methods. Specifically, our model achieved an accuracy of 94.2%, an F1 score of 94.2%, a sensitivity of 94.44%, and a specificity of 94.0%. The use of both CNNs and ViTs for feature extraction leverages their strengths, resulting in a more effective diagnostic tool. The combination of CNNs, ViTs, and Random Forest classifiers, along with advanced data augmentation techniques, shows substantial potential for early glaucoma detection and ongoing monitoring. This approach addresses the limitations of current ML models and enhances diagnostic accuracy and efficiency, making it a promising tool for clinical settings.

Keywords: Glaucoma diagnostics · Ophthalmological Diagnostics · Fundus Image Classification· Random Forest · Deep Learning

1 Introduction

Glaucoma is the leading cause of irreversible blindness, caused by elevated intraocular pressure that damages the optic nerve[1,2]. Early detection is crucial to prevent further vision loss[3,4]. However, in developing countries, the scarcity of ophthalmologists combined with the difficulty in recognizing early symptoms hinders effective diagnosis and treatment of the disease[5,6]. Ophthalmologists typically diagnose glaucoma based on the cup-to-disc ratio (CDR). The CDR measures the ratio of the optic cup's diameter to the optic disc's diameter. In normal individuals, the CDR is usually less than 0.5. A CDR greater than 0.5 often indicates glaucoma, suggesting an enlarged optic cup[7]. Optical Coherence Tomography (OCT) and Fluorescein Fundus Angiography (FFA) are commonly used for glaucoma detection, along with fundus photographs. However, compared to OCT and FFA, fundus imaging is a non-invasive and cost-effective assessment method. This makes fundus photographs suitable for widespread screening of glaucoma, offering a practical and affordable option for early detection and management[8,9]. Fig.1 shows a comparison of fundus images from a normal individual (a) and a glaucoma patient (b). It is evident that the OD/OC ratio in (a) is smaller than that in the glaucoma patient (b). Glaucoma can be categorized into referable glaucoma (RG) and non-referable glaucoma (NRG) based on the CDR and other clinical indicators. RG typically manifests with an elevated CDR and more pronounced symptoms necessitating referral to a specialist for further assessment and intervention, whereas NRG is characterized by a reduced CDR and stable clinical presentation that can be effectively managed by primary care providers or general ophthalmologists.

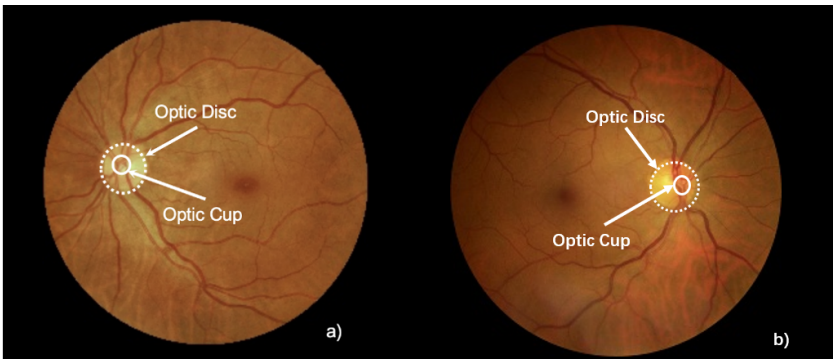

Fig. 1: Comparison of fundus images from a normal individual (a) and a glaucoma patient (b).

Recent advancements in the field of deep learning have significantly increased its effectiveness in the automated diagnosis of glaucoma. These methods have continuously evolved to enhance the accuracy and efficiency of diagnostic pro-

cesses. The GD-CNN model, outlined in [10], exemplified the application of deep learning in medical image analysis by automatically detecting glaucomatous optic neuropathy (GON) from fundus images. This system utilizes retinal fundus images and a convolutional neural network (CNN) called GD-CNN to classify GON with high sensitivity and specificity. The system was trained on a large dataset and validated in multiple cohorts, showing strong generalization across different populations and image qualities. The system demonstrated a high degree of accuracy in several validation datasets, including an area under the receiver operating characteristic curve (AUC) of 0.823, a sensitivity of 82.2%, and a specificity of 70.4% in the local validation dataset. Similarly, [11] contributed to the field by developing a deep learning framework with four convolutional layers and two fully connected layers tailored for automated glaucoma diagnosis. The architecture is designed to extract and learn complex patterns and features from digital fundus images, which are critical for diagnosing glaucoma. To improve the performance of the model and prevent overfitting, the researchers used discard and data enhancement strategies. They tested their model on two different datasets, ORIGA, achieving an area under the curve (AUC) score of 0.831. Going a step further, another deep learning algorithm aims to identify referable glaucomatous optic neuropathy (GON) from color fundus photographs [12]. A large dataset with a large number of fundus photographs was utilized to train and validate the model. The deep learning system achieved impressive results, exhibiting an area under the curve (AUC) of 0.986, sensitivity of 95.6%, and specificity of 92.0% in the validation dataset. Moreover, the Disc-aware Integration Network (DENet) described in [13] enhanced the feature extraction process by considering both global and local characteristics of fundus images. The proposed DENet consists of four interconnected streams: a global image stream, a segmentation guidance network, a local disc region stream, and a disc polar coordinate transformation stream. These streams work synergistically to enhance the feature extraction process, with a special focus on the optic disk, which is a critical region for glaucoma detection. This approach differs from conventional techniques in that it not only relies on clinical measurements, such as the cup-to-disc ratio, but also utilizes morphological and textural features extracted directly from fundus images. The system has been validated using two different datasets, showing superior performance to existing state-of-the-art methods. For example, DENet's area under the curve (AUC) score on the SCES dataset was 0.9183, demonstrating its efficacy in accurately screening for glaucoma. The multi-branch neural network (MB-NN) model discussed in [14] increased interpretability and diagnostic capabilities by integrating clinical data and image-specific features. This approach utilizes a multi-branch neural network (MB-NN) model that enhances the capabilities of traditional deep learning systems by integrating domain knowledge directly into the learning process.The MB-NN model is unique because it integrates key medical measurements and image-specific features, such as cup-to-disc ratios, which are critical for diagnosing glaucoma. The approach allows complex patterns to be extracted from retinal images while maintaining interpretability by integrating medical domain

knowledge. The model demonstrated excellent performance with an accuracy of 0.9151, sensitivity of 0.9233 and specificity of 0.9090. Finally, the two-stage deep learning approach presented in [15] demonstrated how precise localization and classification of the optic disc could enhance model adaptability and robustness across various datasets, ensuring high levels of customization. The main framework operates in two phases: optic disc localization: this phase utilizes a convolutional neural network (RCNN) region approach to accurately localize and extract the optic disc from retinal fundus images. Glaucoma classification: After localizing and extracting the discs, a deep convolutional neural network (DCNN) is used to classify the discs into healthy or glaucomatous categories. A major advancement of the method is its ability to operate independently of dataset-specific characteristics, which enhances its applicability and robustness across different datasets. The authors have also developed a semi-automatic ground truth generation method to overcome the lack of bounding box annotations required to train RCNNs, thus further innovating in this area. The proposed framework has been evaluated on several public datasets, achieving state-of-the-art performance in both localization and classification tasks. The results show that on the ORIGA dataset, the area under the curve (AUC) score for glaucoma classification for the subject's operating characteristics is 0.868, accuracy is 0.874, sensitivity is 0.85, specificity is 0.7117, and F1 score is 0.77. These studies have progressively advanced the use of deep learning techniques in glaucoma detection, from foundational models to specialized solutions, consistently addressing technical challenges and improving diagnostic workflows.

However, the aforementioned studies have some limitations. First, they are all based on CNNs and lack research on attention mechanisms in this field. CNNs are excellent at handling spatial data, but their local receptive fields mean they must rely on deep layers to capture global information. In contrast, attention mechanisms can rely on global dependencies and support parallel computation. Second, these studies use classifiers based on fully connected layers. While fully connected layers can automatically update weights, they lack interpretability compared to machine learning methods like SVM, Bayesian networks, or Random Forests.

Therefore, our research aims to design an automated glaucoma detection model. This model will combine the ability of CNNs to extract local features, such as retinal blood vessels and the optic disc/cup structures, with the global dependency extraction capability of attention mechanisms. For example, since structures like blood vessels are distributed throughout the entire fundus image, there may be dependencies between differently distributed vessels. To enhance the interpretability of the entire model, we will attempt to replace the fully connected layer "black box" with machine learning algorithms.

The rest of this paper is organized as follows: Section 2 details the methods used in our proposed hybrid model for glaucoma diagnosis. Section 3 describes the implementation and results of the framework. Section 4 presents conclusions and future work.

2 Methods

2.1 Datasets

For our research, we are utilizing the EyePACS AIROGS Light dataset[16]. This dataset consists of a balanced subset of standardized fundus images. As shown in Table 1, the dataset is organized into training, validation, and test folders, containing 2500, 270, and 500 fundus images per class, respectively. Each training set includes folders for referable glaucoma (RG) and non-referable glaucoma (NRG) classes, supporting the effective development and evaluation of our glaucoma detection models.

Table 1: EyePACS AIROGS Light Dataset Distribution[16]

Class	**Training**	**Validation**	**Test**
Referable Glaucoma (RG)	2500	270	500
Non-Referable Glaucoma (NRG)	2500	270	500

2.2 Preprocessing & Data Augmentation

One of the most critical steps in glaucoma detection using deep learning models is effective data preprocessing. The first step of this is resizing all images to a size of 224x224 pixels for homogeneousness and compatibility amongst multiple models. Normalization is equally important since it not only makes the process of learning more stable but also strongly convergent by standardizing pixel values. Given the limited quantity of glaucoma fundus images, even moderate overfitting within this dataset is most likely. Data augmentation techniques should be applied to mitigate these: random horizontal flips, random vertical flips, random rotations, and color jitter. These methods create a wide variety of different changes in original images, increasing diversity in the dataset, hence combating overfitting. Increasing the contrast of images is also essential to make the important features more visible. For this, contrast limited adaptive histogram equalization (CLAHE) can be used because it enhances local contrast and makes important features more distinguishable. Fig.2 demonstrates a fundus image from the EyePACS AIROGS Light dataset. Image a) is the original fundus image, while image b) shows the same image after being processed with CLAHE. It can be observed that the CLAHE processed image enhances the visibility of the blood vessels and overall contrast in the fundus image. These preprocessing and augmentation methods increase the variability and quality of the dataset incredibly well.

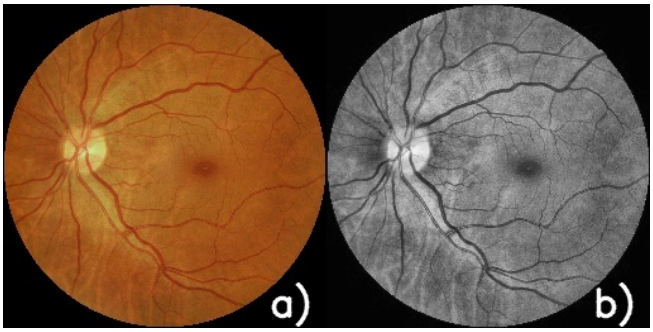

Fig. 2: An example of a fundus image from the EyePACS AIROGS Light dataset[16] processed with CLAHE. The image labeled a) is the original image, and the image labeled b) is the CLAHE processed image.

2.3 DL Model for Glaucoma Classification

Overall, we aim to design a model that consists of four components, as depicted in Fig.3. Firstly, the feature extractor employs various deep learning models to extract features from fundus images. Next, these extracted features are transformed into matrices and then fused. Following this, the fused high-dimensional features undergo dimensionality reduction. Finally, a Random Forest classifier is utilized to classify the reduced-dimensional data.

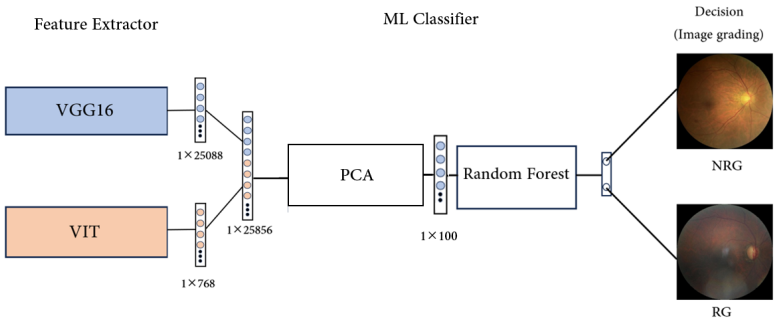

Fig. 3: Architecture of the proposed model. Image classification framework incorporating deep learning-based feature extraction, feature fusion, dimensionality reduction, and random forest classification.

Feature Extraction. In the process of feature extraction, we have selected VGG16 and ViT to extract features from fundus images, respectively. VGG16 excels at spatial feature extraction, while ViT relies on capturing global features.

CNN. Convolutional Neural Networks (CNNs) are a type of deep learning model particularly well-suited for processing two-dimensional data, such as images and videos. By combining convolutional layers, pooling layers, and fully connected layers, CNNs can automatically extract features, reduce the number of parameters, and enhance computational efficiency. Due to their spatial invariance, CNNs excel in tasks such as image recognition and classification[17]. VGG16 is a kind of CNNs proposed by the Visual Geometry Group (VGG) team at the University of Oxford[18]. VGG16 achieves efficient image recognition by stacking multiple 3x3 convolutional layers and max-pooling layers, followed by three fully connected layers for classification. The main advantage of this architecture is that it enhances the model's representational power and accuracy by using smaller convolutional kernels to achieve a deeper network.

ViT. The Vision Transformer (ViT) is a deep learning architecture that applies the Transformer model to computer vision tasks. ViT divides the input image into fixed-size patches and uses the sequence of linear embeddings of these patches as input to the Transformer, thereby handling image classification tasks. By avoiding the complex convolution operations of Convolutional Neural Networks and relying on the self-attention mechanism to capture global features in images, ViT performs exceptionally well in many vision tasks[19].

Feature Fusion. Feature fusion refers to the combination of features from different sources or representations to enhance the representational capacity of a model or improve task performance. In deep learning, features can come from different layers of a neural network, different models, or different data representations. By merging these features, models can better understand and interpret data, thereby improving their prediction or classification capabilities.

Fusing features from VGG16 and Vision Transformer (ViT) can leverage both local and global information, combining VGG16's local feature learning with ViT's global feature capturing ability. VGG16 extracts spatial features, resulting in a 7x7 feature map with 512 channels, which we then flatten into a 1x25088 matrix. Similarly, ViT's features are processed into a 1x768 matrix. We merge these two matrices into a single 1x25856 matrix. This allows the model to consider both local details and global structures more comprehensively when processing images. Additionally, feature fusion can integrate the rich information extracted by the two different models, thereby enhancing the accuracy and robustness of image recognition or classification tasks. Most importantly, feature fusion can improve the model's generalization ability and reduce the risk of overfitting, especially in scenarios with limited data or significant data variability.

Dimensionality Reduction. The next step involves dimensionality reduction of the combined 1x25856 matrix, as its high dimensionality can lead to issues such as increased computational complexity, overfitting, and reduced model performance. We utilized Principal Component Analysis (PCA) for dimensionality reduction, highlighting several key benefits.

First, we standardized the extracted features using StandardScaler to ensure they have similar scales. This step is crucial for PCA to function correctly, as it relies on variance which can be affected by the scale of the features. Next, we used SelectKBest to select the top 100 features, ensuring that only the most useful information was used for classification. This selection helps in retaining the most relevant features while discarding noise and redundant data.

Finally, we applied PCA to further reduce the feature dimensions to 100. This reduction significantly lowers computational complexity, making the model more efficient and faster to train and predict. Additionally, it improves the model's generalization capability by preventing overfitting, as it reduces the likelihood of the model learning noise from the high-dimensional data. Overall, PCA enhances the robustness and accuracy of the image classification task.

Random Forest Classifier. The 1x100 dimensional matrix of features obtained after feature selection is then fed into a Random Forest (RF) classifier to ultimately classify the images into NRG and RG categories. Random Forest is an ensemble learning method that constructs multiple decision trees to perform classification and regression tasks. During the construction of each decision tree, a random subset of the dataset and a random subset of features are selected for training, which enhances the model's diversity and generalization capability[20]. In the prediction phase, Random Forest integrates the predictions from all the decision trees through voting (for classification) or averaging (for regression) to produce the final output.

Compared to a fully connected layer, this approach offers better interpretability and faster training speed. The use of Random Forest not only leverages the ensemble learning technique for improved performance but also provides insights into feature importance, making it easier to understand which features contribute most to the classification. Furthermore, the training process is accelerated due to the efficient nature of the Random Forest algorithm. This combination of feature selection and Random Forest classification results in a robust, interpretable, and efficient model for image classification.

3 Implementation & Results

The overall approach involves training the VGG16 and ViT models separately, utilizing pre-trained weights from ImageNet. The dataset used for this purpose is the EyePACS AIROGS Light dataset. In the context of transfer learning, only the weights of the fully connected layers are updated, while the feature extraction layers remain entirely frozen. This is done because freezing the feature

extraction layers prevents overfitting and leverages the robust feature representations already learned from the large-scale ImageNet dataset. Subsequently, we also train and evaluate the VGG16 model, using the same pre-trained ImageNet weights, and incorporate the feature extraction part of ImageNet into the ViT model for the same dataset. This is because leveraging pre-trained weights helps in achieving better performance with limited data by utilizing the extensive knowledge embedded in these weights from large-scale datasets. Finally, we compare the performance of the standalone VGG16, standalone ViT, and our proposed model on the same dataset to evaluate their respective performance. This comparative analysis allows us to determine the efficacy of our model in integrating both local and global feature extraction capabilities, thus providing a comprehensive understanding of its advantages over individual models.

3.1 VGG16

Using the VGG16 model from torchvision, we performed feature freezing to prevent the weights of the feature extraction layers from being updated during training and replaced the final fully connected layer with a new linear layer having two output units. The training parameters were set with a learning rate of 1×10^{-4}, a batch size of 64, and 50 epochs, using the Adam optimizer and cross-entropy loss function. To enhance performance, we employed PyTorch's GradScaler for mixed precision training and recorded the training loss and images every 10 training steps using TensorBoard, as shown in Fig.**??**. In model evaluation, The random forest classifier achieved an accuracy of 0.828 on the validation set. As presented in Fig.4, the confusion matrix showed the following results: [[420 80] [92 408]]. By this matrix, the model's precision was 0.84, recall was 0.816, and F1 score was 0.828.

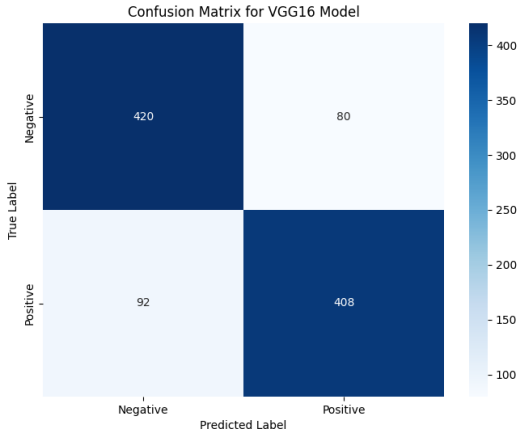

Fig. 4: Confusion matrix results of VGG16 model on the AIROGS dataset.

3.2 ViT

Using the ViT model from torchvision, we performed feature freezing to prevent all parameters from being updated during training and replaced the head of the ViT model with a new linear layer having two output units. The training parameters were set with a learning rate of 1×10^{-3}, a batch size of 64, and 1000 epochs, using the Adam optimizer and cross-entropy loss function. Images and loss were recorded every 100 training steps using TensorBoard, as shown in Fig. **??**. In model evaluation, the random forest classifier achieved an accuracy of 0.7860 on the validation set. As presented in Fig.5, the confusion matrix showed the following results: [[450, 50], [60, 440]]. By this matrix, the model's precision was 0.8980, recall was 0.8800, and F1 score was 0.8880.

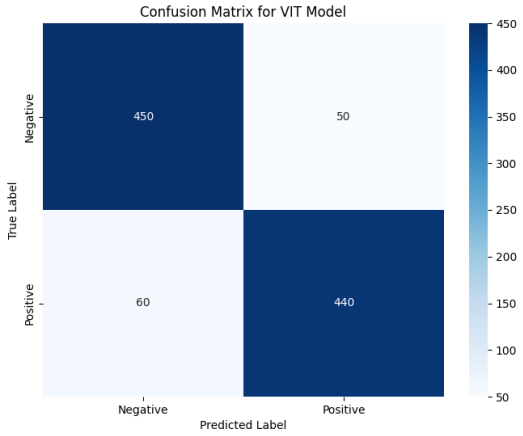

Fig. 5: Confusion matrix results of Vit model on the AIROGS dataset.

3.3 Proposed Model

We implemented our approach following the model design outlined in Section 2.3, the confusion matrix (as displayed in Fig.6) showed the following results:[[470 30][28 472]. By this matrix, our proposed model achieved an accuracy of 0.942 on the validation set, with a precision of 0.942, recall of 0.942, and an F1 score of 0.942.

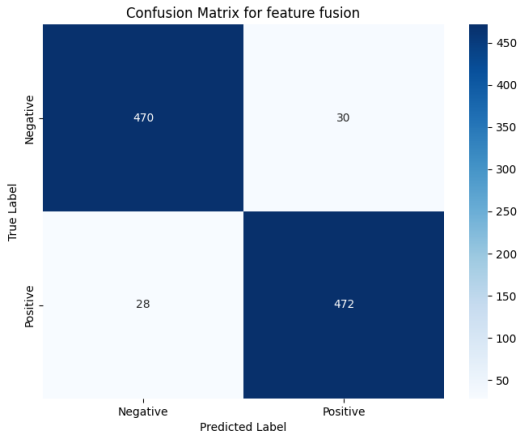

Fig. 6: Confusion matrix results of our proposed model on the AIROGS dataset.

3.4 Comparison and Analysis

As shown in the Fig.7 and Table 2, in terms of accuracy, our method significantly outperforms other models on the AIROGS dataset, achieving 94.2%, while VGG16 and ViT achieved 82.8% and 78.6% respectively. This indicates that our method has a higher accuracy in distinguishing between RG and NRG classes, thereby reducing classification errors. Regarding the F1 score, our method also performed the best, reaching 94.2%. In comparison, the F1 scores for ViT and VGG16 were 88.8% and 82.8% respectively. The F1 score reflects the balance between precision and recall, and our model excels in both aspects, effectively identifying RG samples while minimizing false positives. In terms of sensitivity, our method achieved 94.44%, significantly higher than ViT's 88.0% and VGG16's 81.6%. High sensitivity means our method is very effective in detecting RG (positive samples), capturing as many RG cases as possible and thus reducing the rate of missed detections, which is crucial for clinical diagnosis. For specificity, our method also performed excellently, achieving 94.0%, compared to 90.0% and 84.0% for ViT and VGG16, respectively. High specificity indicates that our method can accurately identify NRG (negative samples), reducing the false positive rate and unnecessary further examinations or treatments. In the Table 2, our method outperforms existing methods across multiple key metrics. Compared to the highest accuracy (91.83%) on the SCES Dataset[13], our method demonstrates higher accuracy (94.2%) and overall performance improvement on the AIROGS dataset. However, it is important to note that, compared to other studies, while our method shows superior performance, there may be some limitations due to the differences in databases and tasks. Other studies use different datasets and tasks, which may affect the direct comparability of

results. Nonetheless, our method excels in multiple key metrics, demonstrating its potential and advantages in the RG and NRG classification task.

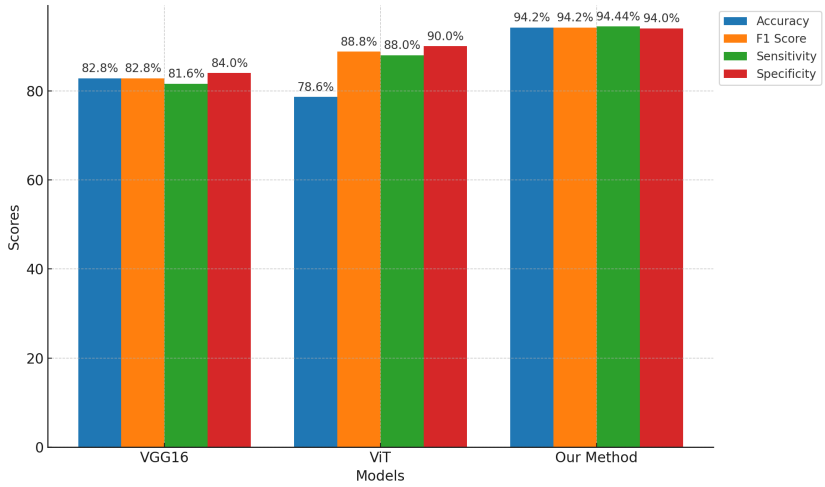

Fig. 7: Performance metrics in the experimental results of VGG16, ViT, and our method.

Table 2: Performance metrics comparison among VGG16, ViT, our method, and other studies.

Authors	**Datasets**	**Accuracy**	**AUC**	**F1 Score**	**Sensitivity**	**Specificity**
[10]	Web Dataset		82.3 %		82.2 %	70.4 %
[11]	ORIGA		83.1 %			
[12]	Web Dataset		98.6 %		95.6 %	92.0 %
[13]	SCES Dataset	91.83 %	95.3 %		83.8 %	83.8 %
[14]	Tongren Hospital	91.51 %			92.33 %	90.9 %
[15]	ORIGA	87.4 %	86.8 %	77 %	85 %	71.17 %
VIT	AIROGS	78.6 %		88.8 %	88.0%	90.0%
VGG16	AIROGS	82.8 %		82.8 %	81.6%	84.0%
Our method	**AIROGS**	**94.2 %**		**94.2 %**	**94.44%**	**94.0%**

4 Conclusion

Glaucoma is a very common cause of irreversible blindness due to an elevated intraocular pressure, which has resulted in optic nerve damage. It can be diagnosed early and further deterioration prevented, but there are challenges to effective diagnosis in developing regions, the major ones being access to eye specialists and difficulty in recognizing symptoms at an early stage. Recent deep learning developments have made a huge improvement in the automated identification of glaucoma; it improves accessibility and reliability for the detection at an early stage of this disease. Our model has four key stages. First, we extract the features from fundus images through VGG16 for local feature extraction and ViT for global features. Then, fuse them into one matrix with both pieces of information. After that, PCA will be applied to this matrix for reducing the dimensionality to enhance efficiency and avoid overfitting. These reduced features are then classified using a Random Forest classifier that improves on diversity and generalization toward the building of a robust image classification model efficiently. Our method outperformed these models on the AIROGS dataset to the tune of 94.2%, against 82.8% for VGG16 and 78.6% for ViT, and returned good results for the F1 score, sensitivity, and specificity. However, this study has some limitations. First, the dataset's size and diversity are relatively small, which may limit the model's generalization ability and affect its application in real clinical settings. Second, training deep learning models requires substantial computational resources, especially when processing high-resolution fundus images, which might limit the widespread application of these methods. Future research can improve the model's generalization ability by increasing the dataset's size and diversity and explore more efficient computational methods to reduce computational costs. Additionally, studying the interpretability of the models can enhance the understanding and trust in diagnostic results. These improvements will help further enhance the accuracy and practicality of glaucoma diagnosis.

Acknowledgements. This project is supported by XJTLU Summer Undergraduate Research Fellowships SURF-2024-0399. We thank Zhongying Zhu, Ceran Huang, Zhiqin Yang, and Yuchen Li for their supports.

Open Practices Statement. Datasets generated and analyzed during the current study could be found available in the OSF repository https://osf.io/ha4gm/?view_only=aaee0f1466394c4e8072376e4d94b7a5 upon publication. None of the experiment reported in the present study was preregistered.

References

1. Singh, P.B., Singh, P., Dev, H., Tiwari, A., Batra, D., Chaurasia, B.K.: *Glaucoma Classification using Light Vision Transformer. EAI Endorsed Transactions on Pervasive Health and Technology* **9** (2023)

2. Coan, L.J., Williams, B.M., Adithya, V.K., Upadhyaya, S., Alkafri, A., Czanner, S., Venkatesh, R., Willoughby, C.E., Kavitha, S., Czanner, G.: *Automatic detection of glaucoma via fundus imaging and artificial intelligence: A review. Survey of ophthalmology* **68**(1), 17–41 (2023)
3. Geetha, A., Prakash, N.: *Classification of Glaucoma in Retinal Images Using EfficientnetB4 Deep Learning Model. Comput. Syst. Sci. Eng.* **43**(3), 1041–1055 (2022)
4. Sewal, N., Virmani, C.: *Predicting Glaucoma Progression Using Machine Learning.* In: *2023 5th International Conference on Advances in Computing, Communication Control and Networking (ICAC3N).* pp. 856–861. IEEE (2023)
5. Sangchocanonta, S., Ingpochai, S., Puangarom, S., Munthuli, A., Phienphanich, P., Itthipanichpong, R., Chansangpetch, S., Manassakorn, A., Ratanawongphaibul, K., Tantisevi, V., et al.: *Donut: Augmentation Technique for Enhancing The Efficacy of Glaucoma Suspect Screening.* In: *2023 45th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC).* pp. 1–5. IEEE (2023)
6. Puangarom, S., Twinvitoo, A., Sangchocanonta, S., Munthuli, A., Phienphanich, P., Itthipanichpong, R., Ratanawongphaibul, K., Chansangpetch, S., Manassakorn, A., Tantisevi, V., et al.: *3-LbNets: Tri-Labeling Deep Convolutional Neural Network for the Automated Screening of Glaucoma, Glaucoma Suspect, and No Glaucoma in Fundus Images.* In: *2023 45th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC).* pp. 1–5. IEEE (2023)
7. Garway-Heath, D., Ruben, S., Viswanathan, A., Hitchings, R.: Vertical cup/disc ratio in relation to optic disc size: its value in the assessment of the glaucoma suspect. British Journal of Ophthalmology **82**, 1118 – 1124 (1998). https://doi.org/10.1136/bjo.82.10.1118
8. Khan, M.Z., Lee, Y.: *Stacked Ensemble Network to Assess the Structural Variations in Retina: A Bio-marker for Early Disease Diagnosis.* In: *2022 44th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC).* pp. 3222–3226. IEEE (2022)
9. Elangovan, P., Nath, M.K.: *En-ConvNet: A novel approach for glaucoma detection from color fundus images using ensemble of deep convolutional neural networks. International Journal of Imaging Systems and Technology* **32**(6), 2034–2048 (2022)
10. Liu, H., Li, L., Wormstone, I.M., Qiao, C., Zhang, C., Liu, P., Li, S., Wang, H., Mou, D., Pang, R., Yang, D., Zangwill, L.M., Moghimi, S., Hou, H., Bowd, C., Jiang, L., Chen, Y., Hu, M., Xu, Y., Kang, H., Ji, X., Chang, R., Tham, C., Cheung, C., Ting, D.S.W., Wong, T.Y., Wang, Z., Weinreb, R.N., Xu, M., Wang, N.: *Development and Validation of a Deep Learning System to Detect Glaucomatous Optic Neuropathy Using Fundus Photographs. JAMA Oph thalmology* **137**, 1353–1360 (2019)
11. Chen, X., Xu, Y., Wong, D.W.K., Wong, T.Y., Liu, J.: *Glaucoma Detection based on Deep Convolutional Neural Network. IEEE Transactions on Medical Imaging* **34**, 513–521 (2015)
12. Author, F., Author, S., Author, T.: *Efficacy of a Deep Learning System for Detecting Glaucomatous Optic. Journal Name* **Vol**, Page Range (2019)
13. Fu, H., Cheng, J., Xu, Y., Zhang, C., Wong, D.W.K., Liu, J., Cao, X.: *Disc-Aware Ensemble Network for Glaucoma Screening From Fundus Image. IEEE Transactions on Medical Imaging* **37**(11), 2493–2501 (2018). https://doi.org/10.1109/TMI.2018.2837012
14. Chai, Y., Liu, H., Xu, J.: *Glaucoma diagnosis based on both hidden features and domain knowledge through deep learning models. Knowledge-Based Systems* **161**, 147–156 (2018). https://doi.org/10.1016/j.knosys.2018.07.043

15. Chen, X., Huang, Y., Huang, L.: *Two-stage framework for optic disc segmentation and glaucoma classification using deep learning. BioMedical Engineering OnLine* **18**(1), 23 (2019). https://doi.org/10.1186/s12938-019-0649-z
16. Kiefer, R.: Glaucoma dataset: Eyepacs airogs - light (2023). https://doi.org/10.34740/KAGGLE/DS/3222646, https://www.kaggle.com/ds/3222646
17. Krizhevsky, A., Sutskever, I., Hinton, G.E.: *ImageNet classification with deep convolutional neural networks. Advances in Neural Information Processing Systems* **25**, 1097–1105 (2012). https://doi.org/10.1145/3065386
18. Simonyan, K., Zisserman, A.: *Very Deep Convolutional Networks for Large-Scale Image Recognition. arXiv preprint arXiv:1409.1556* (2014). https://doi.org/10.1145/3065386
19. Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., Uszkoreit, J., Houlsby, N.: *An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale. arXiv preprint arXiv:2010.11929* (2020). https://doi.org/10.1145/3065386
20. Breiman, L.: *Random forests. Machine Learning* **45**(1), 5–32 (Oct 2001). https://doi.org/10.1023/A:1010933404324