Online-Journals.org

iJOE International Journal of Online and Biomedical Engineering

iJOE | eISSN: 2626-8493 | Vol. 21 No. 2 (2025) | OPEN ACCESS

https://doi.org/10.3991/ijoe.v21i02.51409

PAPER

Dissecting Retinal Disease: A Multi-Modal Deep Learning Approach with Explainable AI for Disease Classification across Various Classes

Ayush Gupta, Jeya Mala D.(✉), Vishal Kumar Yadav, Mayank Arora

Vellore Institute of Technology, Chennai, Tamil Nadu, India

jeyamala.d@vit.ac.in

ABSTRACT

This study investigates the efficacy of various deep learning (DL) models in detecting retinal diseases, specifically focusing on cataract detection. Utilizing a pre-processed fundus images data set classified into normal and cataract classes, we evaluate the performance of ResNet, VGG-16 and VGG-19 models based on accuracy, sensitivity, and specificity in classifying fundus images. The primary objective of this work is to provide explanations on the predictions done by the said DL models in order to ensure the ground-truth verification. The explanation is achieved using the explainable artificial intelligence (XAI) model namely gradient-weighted class activation mapping (Grad-CAM), which helps to visualize and interpret the decision-making process of these models. Through a comprehensive exploratory data analysis (EDA), model training, and evaluation, VGG-19 emerged as the superior model, achieving the highest accuracy, precision, and recall. Grad-CAM heat maps provide insights into the models' attention in image features, highlighting the impact of cataracts on retinal structure. The study underscores the potential of DL in retinal disease detection and the pivotal role of explainable artificial intelligence (XAI) in enhancing model interpretability. Future directions include exploring more advanced DL architectures and furthering the application of XAI techniques to improve detection systems' accuracy and transparency.

KEYWORDS

retinal fundus, multi modal, deep learning (DL), explainable artificial intelligence (XAI), ResNet, VGG, gradient-weighted class activation mapping (Grad-CAM)

1 INTRODUCTION

Retinal diseases are a significant global health concern, affecting millions of individuals and leading to vision loss and blindness. Early detection and diagnosis of retinal diseases are critical for effective treatment and prevention of vision loss.

Gupta, A., Jeya Mala, D., Yadav, V.K., Arora, M. (2025). Dissecting Retinal Disease: A Multi-Modal Deep Learning Approach with Explainable AI for Disease Classification across Various Classes. *International Journal of Online and Biomedical Engineering (iJOE)*, 21(2), pp. 38–51. https://doi.org/10.3991/ijoe.v21i02.51409

Article submitted 2024-08-02. Revision uploaded 2024-09-16. Final acceptance 2024-11-12.

In recent years, deep learning (DL) models have shown promise in detecting retinal diseases, with models such as ResNet, VGG-16, and VGG-19 121 demonstrating impressive results. In this study, we have used multimodal DL models of the above as the dataset is going to be the retinal images with imaging modalities, which included the scanned images of the eyes for dissecting retinal diseases. Also, as there is a lack of comparative analysis of these models and the application of explainable artificial intelligence (XAI) to understand the differences in their performance, this study has provided all these and helps in understanding and interpreting the reasons for decisions being taken.

This study aims to address this knowledge gap by comparing the performance of different multimodal DL models for retinal disease detection and analyzing their results using gradient-weighted class activation mapping (Grad-CAM) for XAI. The study aims to answer the research question of how different DL models perform in detecting retinal diseases and what factors contribute to their performance. The hypothesis is that the performance of the models will vary depending on the type of retinal disease and the complexity of the images, and that Grad-CAM will provide insights into the factors that contribute to their performance.

Understanding the strengths and limitations of these DL models is crucial for developing effective and reliable retinal disease detection systems. By comparing the performance of ResNet, VGG-16, and VGG-19 models and analyzing their results using Grad-CAM, this study aims to contribute to the growing body of knowledge in this field and provide insights into the factors that contribute to the performance of DL models for retinal disease detection. This study's findings have the potential to inform the development of more effective and reliable retinal disease detection systems, ultimately improving patient outcomes and reducing the burden of retinal diseases worldwide.

2 LITERATURE SURVEY

The recent advancements in artificial intelligence (AI) for medical imaging in the fight against COVID-19 have been examined in [1]. The authors explored how AI can automate image acquisition procedures, improving workflow and minimizing contact with patients. Additionally, the application of AI to enhance work efficiency by accurately identifying infections in X-ray and CT scans, facilitating better quantification. Research works highlighted the application of explainable DL for efficient and robust pattern recognition [2]. The authors provided an overview of the latest advancements and representative works within each category.

The study by the authors [3] investigated the application of DL for classifying mesothelioma, an aggressive cancer. The authors proposed MesoNet, a deep convolutional neural network (CNN) that predicts patient survival from whole-slide digitized images without requiring pathologist-defined regions of interest. Interestingly, the model identified regions in the stroma, linked to inflammation and cellular diversity that contribute to patient outcome prediction. This suggested that DL can potentially reveal novel features for prognosis and biomarker discovery.

A systematic review [4] explored the application of XAI in healthcare over the past decade. While AI models have achieved human-level performance in various healthcare tasks, their lack of transparency remains a significant barrier to adoption. This paper emphasized the need for XAI research to address this challenge and build trust in AI-based healthcare systems.

Researchers [5] have presented a deep learning platform (DLP) for detecting multiple retinal diseases and conditions in retinal fundus photographs. The DLP achieved high accuracy in identifying 39 different classes using a vast dataset of fundus images with corresponding labels. Notably, the performance surpassed the average level of retina specialists.

Researchers have presented a novel approach for rapid intraoperative brain tumor diagnosis to combine stimulated Raman histology (SRH), a label-free optical imaging technique, with deep CNNs. The study demonstrated high accuracy in a multicenter clinical trial, comparable to pathologist interpretation of traditional histology images [6].

A survey [7] explored the integration of Internet of Things (IoT) and machine learning (ML) for developing healthcare systems. The authors proposed a unique taxonomy for IoT-ML healthcare systems, highlighting key development steps and research challenges, including exploration of DL models, data acquisition and handling, and privacy and security concerns.

The authors applied CNNs for classifying and localizing wheat diseases in images [8]. They acknowledged the limitations of DL models as "black boxes" and proposed Grad-CAM, a visualization method, to pinpoint disease locations within wheat spike images. This approach has offered valuable insights into the decision-making process of the CNN model.

The application of DL for colorectal cancer (CRC) diagnosis during colonoscopy was investigated in [9]. The authors introduced CRCNet, a DL model trained on a massive dataset of colonoscopy images from over 12,000 patients. They evaluated the model on independent datasets and demonstrated high accuracy in identifying CRC at the patient level, exceeding the average performance of endoscopists in recall rate for two test sets and precision for one set.

The authors in [10] compared three pre-trained CNN architectures (VGG16, ResNet50, and SE-ResNet50) for recognizing seven basic emotions: anger, contempt, disgust, fear, happiness, and sadness. The study demonstrated that all three CNNs achieve high validation accuracy, with ResNet50 surpassing the others. This suggests the effectiveness of CNNs for affect detection and the potential of transfer learning for improving performance in such tasks.

The research work [11] addressed the challenge of identifying plant diseases using DL. The authors proposed a transfer learning-based CNN model built on the ResNet50 architecture. The model achieved a remarkable training accuracy of 99.80%, indicating its effectiveness in classifying plant diseases.

The authors in [12] investigated the potential application of transformers for image retrieval tasks. Building upon their success in natural language processing and image classification, the authors proposed a transformer-based approach. Notably, their method achieves state-of-the-art performance on benchmark datasets for category-level retrieval. Additionally, the study suggested that transformers are competitive for object retrieval tasks, particularly when dealing with short vector representations and low-resolution images.

The growing trend of using 3D medical image data for tasks like brain tumor segmentation was analyzed in [13]. This paper proposed a novel '3D brain tumor segmentation' model that leverages several advancements. Overall, this work has contributed to the development of more effective 3D brain tumor segmentation techniques.

The research presented in [14] addressed brain tumor (BT) classification, a crucial step in early cancer detection. The paper proposed a novel hybrid

transformer-enhanced convolutional neural network (TECNN) model. This model combines the strengths of both approaches: CNNs for local feature extraction and the vision transformer's (ViT's) self-attention mechanism for capturing long-range dependencies.

The development of an automated system for early skin cancer detection using medical images utilized a hybrid feature set that combines deep features extracted from the AlexNet architecture with local optimal-oriented patterns [15]. This approach aimed for accurate skin lesion prediction. The results demonstrated the effectiveness of the hybrid features in conjunction with ML, achieving an accuracy of 94.7% using an ensemble classifier. This suggested the potential of the proposed method for aiding physicians in skin cancer diagnosis.

3 BACKGROUND

The retina plays a crucial role in vision as it is a thin layer of tissue located at the back of the eye that contains millions of light-sensitive cells, including rods and cones. These cells are responsible for receiving, organizing, and sending visual information to the brain through the optic nerve, enabling individuals to see. Retinal diseases encompass a range of conditions that cause damage to different parts of the retina. Common retinal diseases such as diabetic retinopathy, age-related macular degeneration, cataract, and glaucoma have a significant impact on vision loss and pose a substantial burden on public health due to their prevalence and potential for severe visual impairment or blindness [17].

In the realm of retinal imaging, two key modalities play a vital role in capturing retinal structures. Fundus photography is a technique used to photograph the back of the eye, providing detailed images of the retina. On the other hand, optical coherence tomography (OCT) is a non-invasive imaging technique that allows for the visualization of retinal layers with high resolution, aiding in the diagnosis and monitoring of retinal diseases. Each imaging modality has its advantages and limitations, contributing to the comprehensive assessment of retinal health.

Deep learning has emerged as a powerful tool for several real-time applications [1] [2] [18]. These leverage artificial neural networks and CNNs to process and interpret complex medical images. The application of DL in retinal disease diagnosis presents numerous benefits, including improved accuracy, speed, and scalability in processing large volumes of medical images.

Explainable artificial intelligence is a critical concept in healthcare applications, including retinal disease diagnosis, as it focuses on making the decision-making process of AI models transparent and interpretable. Given the inherent complexity of "black-box" DL models, XAI techniques such as Grad-CAM to provide insights into how these models arrive at their predictions, enhancing trust, understanding, and acceptance of AI-driven diagnostic systems in the medical field.

4 PROPOSED METHODOLOGY

4.1 Merits and contributions of this research work

The proposed work has performed an extensive exploratory data analysis and has applied the multi-modal DL models to identify the retinal disease. The work is

novel as it has integrated the XAI part to ensure the fidelity of the DL models. As the problem domain is related to the eye, the decisions should be trustworthy in order to deploy in the target environment. Hence, this study has applied Grad-CAM model to get the explanations from the model's prediction result. This explanation is a visualization in the form of a heat map, which helps the doctors and other stakeholders to easily understand the decision and validate its trustworthiness. Further, the work has compared the results of various multi-modal DL models and concluded that the VGG-19 model outperforms the other models.

4.2 Exploratory data analysis

The dataset used in this study consists of pre-processed fundus images from the Kaggle repository [16]. It includes eight different types of eye conditions: normal, diabetes, glaucoma, cataract, age-related macular degeneration, hypertension, pathological myopia, and other abnormalities and the data distribution, snapshot of the dataset are given in Figures 1(a), (b), (c), and (d). The dataset comprises approximately 1000 images for each class, with the cataract class containing 1038 images and the normal class containing 1098 images. The dataset was originally sourced from Kaggle, where it includes fundus image data for normal eyes, cataracts, glaucoma, and retinal diseases, which are used in some of the works.

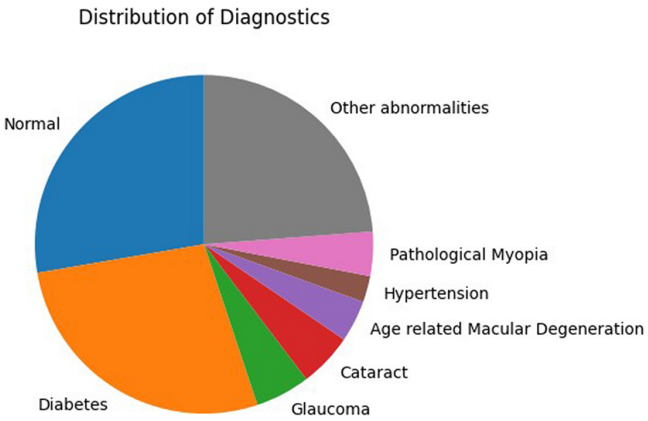

Fig. 1a. Pie chart of the distribution

```
RangeIndex: 3500 entries, 0 to 3499
Data columns (total 15 columns):
 #   Column                    Non-Null Count  Dtype
---  ------                    --------------  -----
 0   ID                        3500 non-null   int64
 1   Patient Age               3500 non-null   int64
 2   Patient Sex               3500 non-null   object
 3   Left-Fundus               3500 non-null   object
 4   Right-Fundus              3500 non-null   object
 5   Left-Diagnostic Keywords  3500 non-null   object
 6   Right-Diagnostic Keywords 3500 non-null   object
 7   N                         3500 non-null   int64
 8   D                         3500 non-null   int64
 9   G                         3500 non-null   int64
 10  C                         3500 non-null   int64
 11  A                         3500 non-null   int64
 12  H                         3500 non-null   int64
 13  M                         3500 non-null   int64
 14  O                         3500 non-null   int64
dtypes: int64(10), object(5)
memory usage: 410.3+ KB
```

Fig. 1b. Snapshot of dataset

For the purpose of this study, we work specifically with the cataract and normal classes. The distribution of these two classes is nearly balanced, with a slightly larger number of normal images. This dataset will provide a solid foundation for analyzing and classifying cataract and normal fundus images using DL techniques, with the potential for further expansion and refinement in future studies.

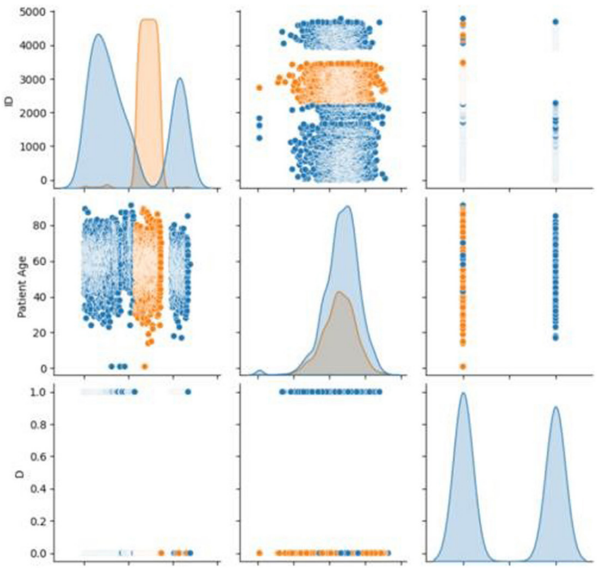

Fig. 1c. Pair plot comparison and KDE distribution

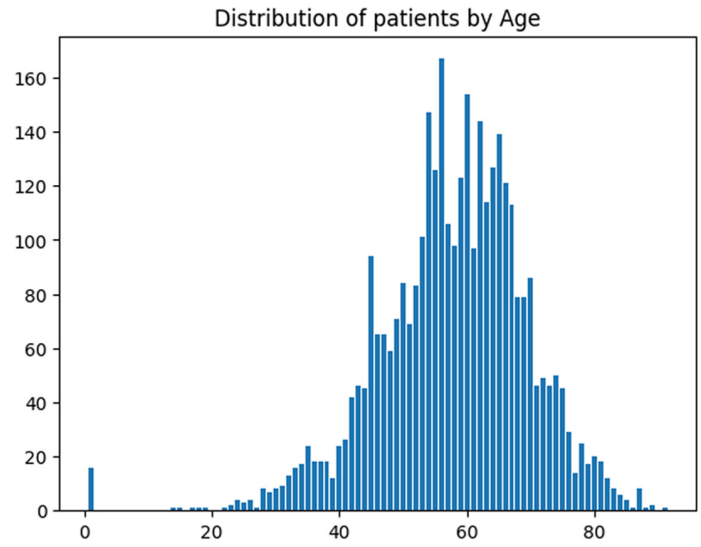

Fig. 1d. Patients by age distribution

4.3 Experimentation of multi modal deep learning approach

In this study, we explore the performance of three state-of-the-art DL models, namely ResNet, VGG-16, and VGG-19, in classifying cataract and normal fundus images. The models were trained using a batch size of 32, with each image resized to 264 × 264 pixels. The training process was carried out for 100 epochs to ensure the models converged to optimal performance. The dataset was used to train all three models, providing a comprehensive comparison of their classification capabilities.

ResNet. ResNet, or residual network, is a DL architecture that introduces skip connections to mitigate the vanishing gradient problem, allowing for the training of very deep networks. The ResNet-50 architecture was used in this study, with 50 layers and 23 million parameters. It is given in Figure 2.

Layer (type)	Output Shape	Param #
rescaling_1 (Rescaling)	(None, 264, 264, 3)	0
resnet50 (Functional)	(None, None, None, 2048)	23587712
flatten_1 (Flatten)	(None, 165888)	0
dense_3 (Dense)	(None, 256)	42467584
dense_4 (Dense)	(None, 256)	65792
dense_5 (Dense)	(None, 1)	257
activation_1 (Activation)	(None, 1)	0

Total params: 66121345 (252.23 MB)
Trainable params: 66068225 (252.03 MB)
Non-trainable params: 53120 (207.50 KB)

Fig. 2. Architecture or flow of the ResNet

VGG-16. VGG-16 is a deep CNN that gained popularity due to its simplicity and high performance. The model as shown in Figure 3 consists of 16 layers, including 13 convolutional layers, five max pooling layers, and three fully connected layers.

Layer (type)	Output Shape	Param #
rescaling_1 (Rescaling)	(None, 264, 264, 3)	0
resnet50 (Functional)	(None, None, None, 2048)	23587712
flatten_1 (Flatten)	(None, 165888)	0
dense_3 (Dense)	(None, 256)	42467584
dense_4 (Dense)	(None, 256)	65792
dense_5 (Dense)	(None, 1)	257
activation_1 (Activation)	(None, 1)	0

Total params: 66121345 (252.23 MB)
Trainable params: 66068225 (252.03 MB)
Non-trainable params: 53120 (207.50 KB)

Fig. 3. Architecture or flow of VGG-16

VGG-19. VGG-19 is an extension of the VGG-16 architecture as shown in Figure 3 has 19 layers and approximately 144 million parameters. The additional layers in VGG-19 provide a higher capacity for learning complex features, potentially improving its performance in image classification tasks. It is given in Figure 4.

Layer (type)	Output Shape	Param #
rescaling_2 (Rescaling)	(None, 264, 264, 3)	0
vgg16 (Functional)	(None, None, None, 512)	14714688
flatten_2 (Flatten)	(None, 32768)	0
dense_6 (Dense)	(None, 256)	8388864
dense_7 (Dense)	(None, 256)	65792
dense_8 (Dense)	(None, 1)	257
activation_2 (Activation)	(None, 1)	0

Total params: 23169601 (88.39 MB)
Trainable params: 23169601 (88.39 MB)
Non-trainable params: 0 (0.00 Byte)

Fig. 4. Architecture or flow of VGG-19

5 COMPARATIVE ANALYSIS

In this study, we conducted a comparative analysis of the three DL models, namely ResNet, VGG-16, and VGG-19 to evaluate their performance in classifying cataract and normal fundus images. To ensure compatibility with real-time scenarios, we considered various factors such as computational efficiency, model size, and inference time.

The confusion matrices for each model, as presented in Figure 5, demonstrates their classification performance in terms of true positives, true negatives, false positives, and false negatives. These matrices allowed us to calculate key performance metrics such as accuracy, precision, recall, and F1-score. The VGG-19 outperformed the other models in terms of accuracy, precision, and recall. This model's ability to maintain strong feature propagation, as shown in Figure 5, reduces the number of parameters contributed to its robust performance in image classification tasks. Table 1 shows the training and testing accuracy and loss of the different models.

Table 1. Comparison of accuracy and loss of different models

Model	#Epochs	Training Accuracy	Testing Accuracy	Model Loss (Training)	Model Loss (Testing)
VGG16	16	0.783	0.825	1.758	1.623
VGG19	16	0.972	0.981	0.186	0.189
Resnet50	16	0.872	0.851	0.124	0.126 (higher fluctuation)

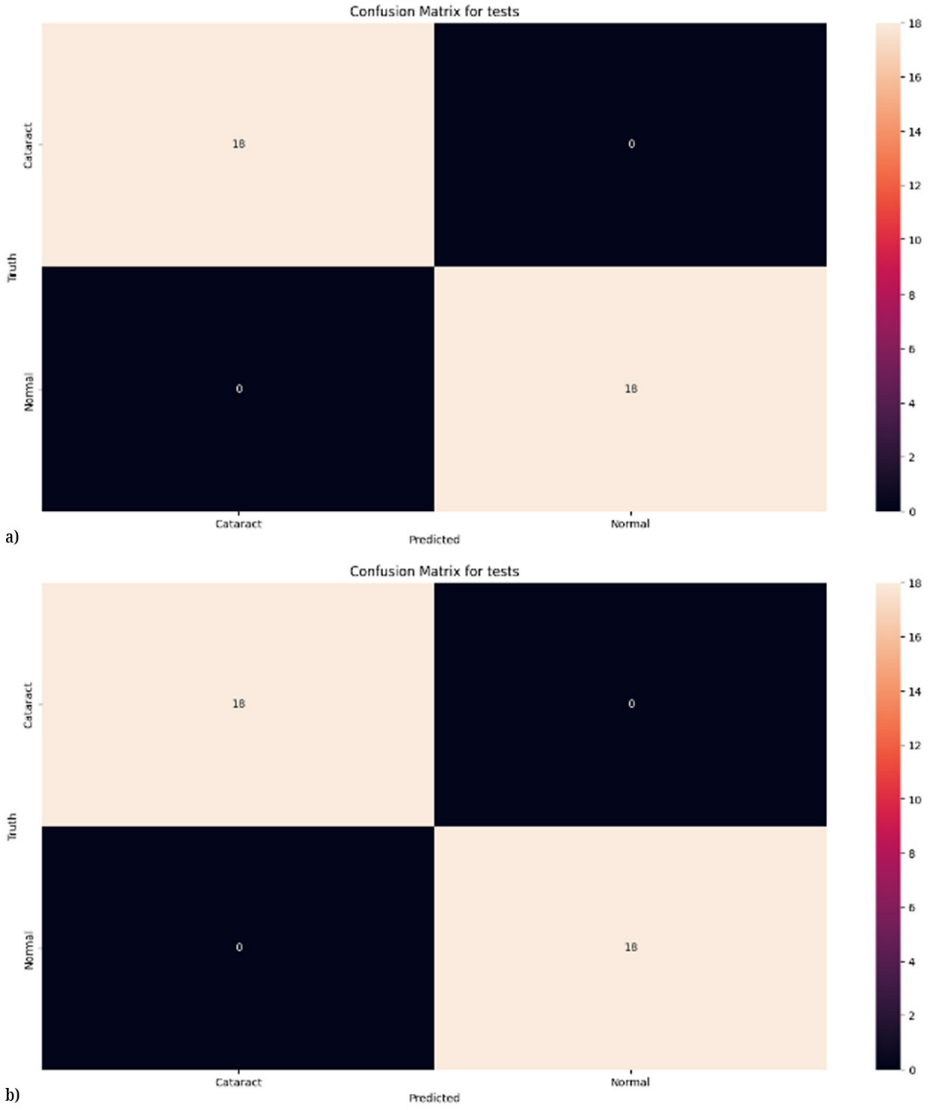

Fig. 5. *(Continued)*

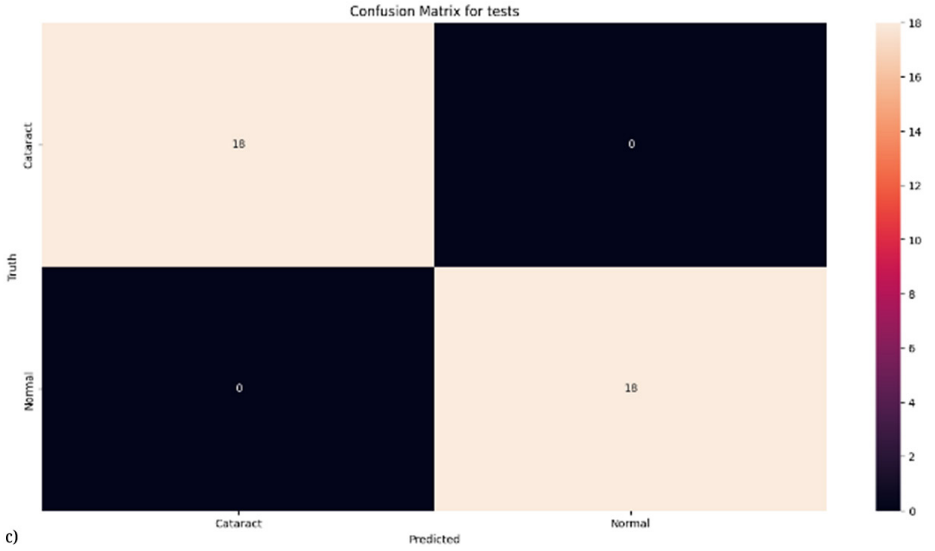

Fig. 5. (a) Confusion matrix for ResNet50, (b) Confusion matrix for VGG-16, (c) Confusion matrix for VGG-19

The comparative analysis as given in Table 1 and Figure 6 provided valuable insights into the performance of each model in classifying cataract and normal fundus images. These insights can inform the selection of appropriate models for real-time medical applications, ensuring accurate diagnosis of eye conditions.

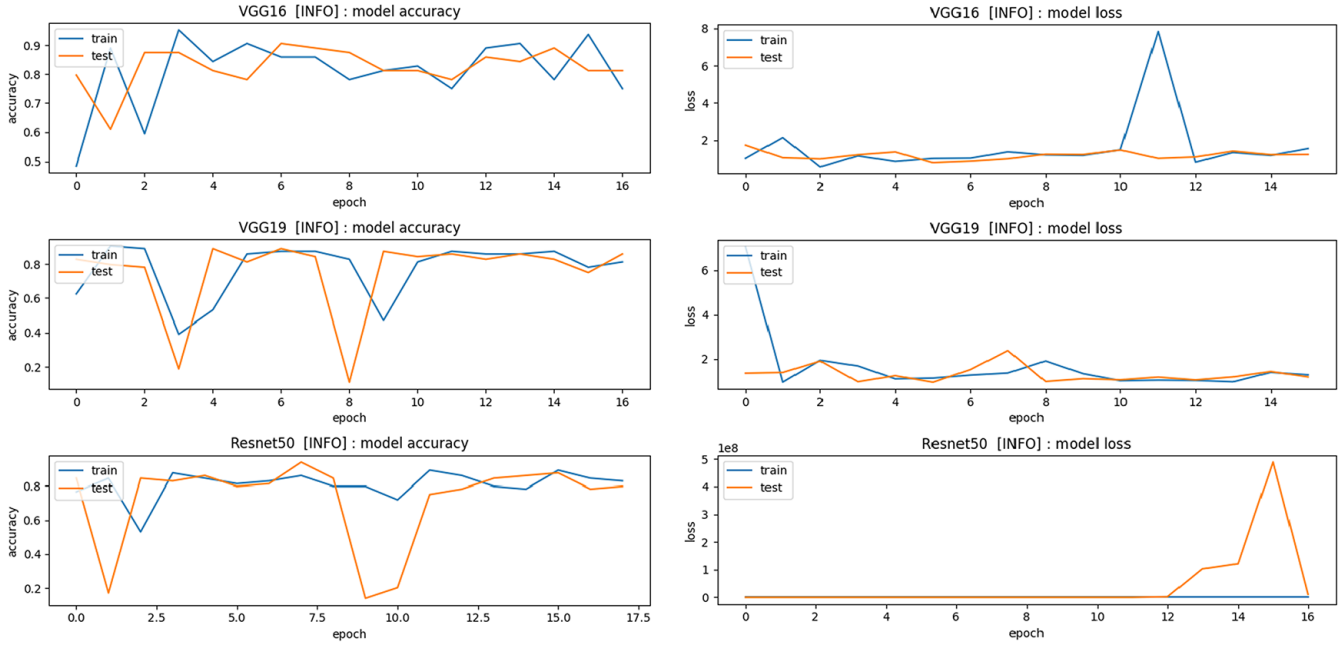

Fig. 6. Graph of accuracy and loss with epochs of models VGG-16, VGG-19 and Resnet50

6 PRE-PROCESSING OF RESULT

Due to the limitations of Grad-CAM in TensorFlow, we performed an inter-model conversion from TensorFlow models to open neural network exchange (ONNX) models using the ONNX runtime. The ONNX models were then converted to PyTorch

runtime, enabling us to extract layer-wise activation maps for each model. These activation maps were essential for the Grad-CAM analysis, as they allowed us to back-propagate the gradients and generate heat-maps that highlighted the regions of the input images that most influenced the models' predictions.

7 DISCUSSION AND GROUND TRUTH VERIFICATION

After evaluating the performance of the ResNet, VGG-16, and VGG-19 in classifying cataract and normal fundus images, we further analyzed the results using the XAI techniques, specifically Grad-CAM. The equation (1) shows the Grad-CAM activation map generation.

$$a_k^c = \frac{1}{z}\sum_i\sum_j \eth y^c / \eth A ij^k \tag{1}$$

The heat-maps generated through Grad-CAM allowed us to interpret and understand the decision-making process of each model. By visualizing the regions of interest using XAI technique namely GradCAM as shown in Figure 7 (i), (ii) and (iii) influenced the classification results, we gained valuable insights into the features and patterns that the models used to differentiate between Cataract and Normal fundus images. This detailed analysis enriched our interpretation of the models' performance and provided a more comprehensive understanding of their classification capabilities in the context of 'Cataract and Normal fundus image' classification.

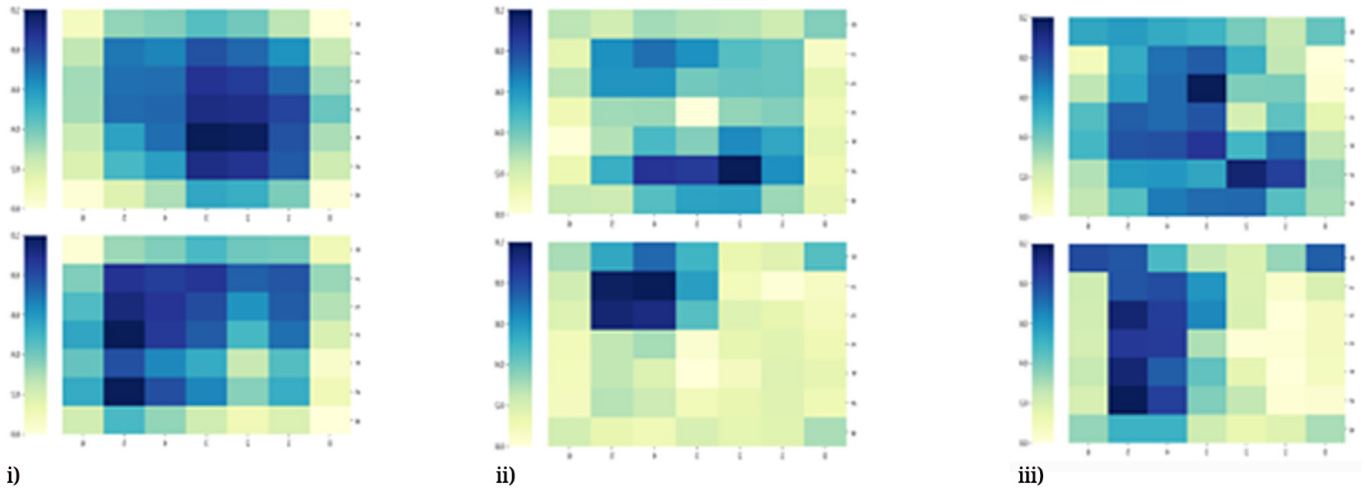

Fig. 7. Grad-CAM activation maps for (i). VGG-19, (ii), VGG-16, (iii). ResNet

After obtaining the heatmaps through Grad-CAM analysis, we conducted a detailed vision analysis of the results to gain deeper insights into the models' decision-making processes. The activation maps revealed distinct patterns in the fundus images, highlighting regions with varying intensities of nerve cells and structural features.

Upon analysis, we observed that the activation maps as given in Figure 8 exhibited pronounced gradients in areas with high intensity of nerve cells, indicating the models' focus on these regions for classification. In contrast, for cataract images, the activation maps displayed a diffuse and widespread distribution, covering the entire image. This observation aligns with the characteristic clouding of the retina

in cataract cases, where the presence of dominant nerve cells is obscured, leading to a lack of distinct features in the fundus scan. The distinct patterns observed in the activation maps for normal and cataract images reflected the underlying structural changes in the retina associated with each condition. The comprehensive coverage of the retina in cataract images by the activation map further emphasized the impact of the disease on the visibility of nerve cells and structural details in the fundus scan.

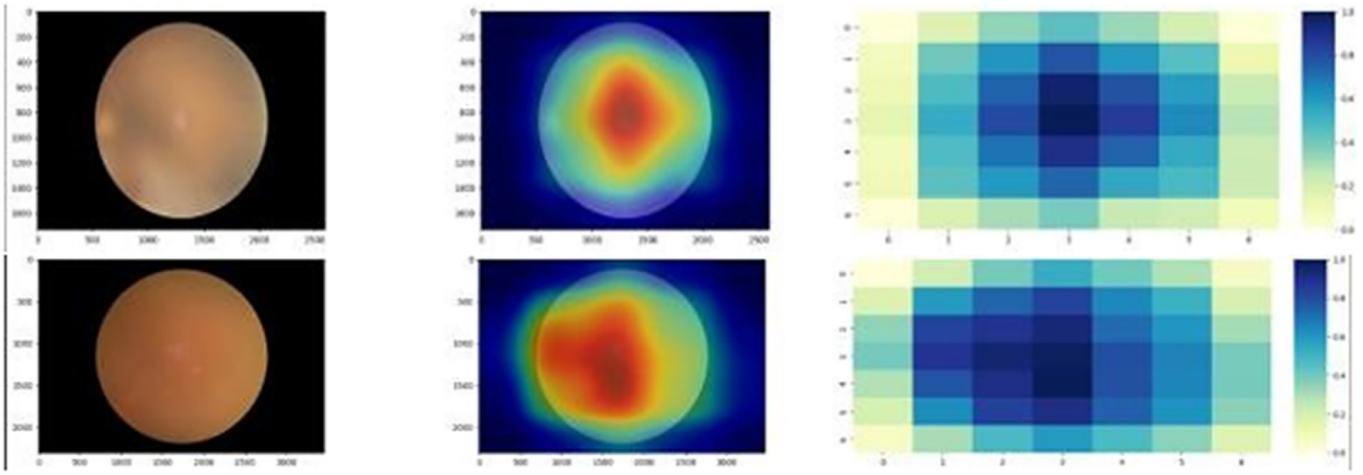

Fig. 8. Cataract heat map and Grad-CAM activation images in VGG-19

The Figure 8 also illustrates the activation map overlay on a cataract fundus image, highlighting the widespread distribution of the heat-map across the entire image. This visualization highlights the model's attention to regions with reduced nerve cell visibility, providing a clear indication of the disease's impact on the retinal structure.

This vision analysis of the activation maps enhances our understanding of how the DL models interpret and classify cataract and normal fundus images. The alignment between the activation map patterns as shown in Figure 9 and the actual characteristics of the diseases underscores the models' ability to capture relevant features for accurate classification, demonstrating the potential of XAI techniques such as Grad-CAM in medical image analysis.

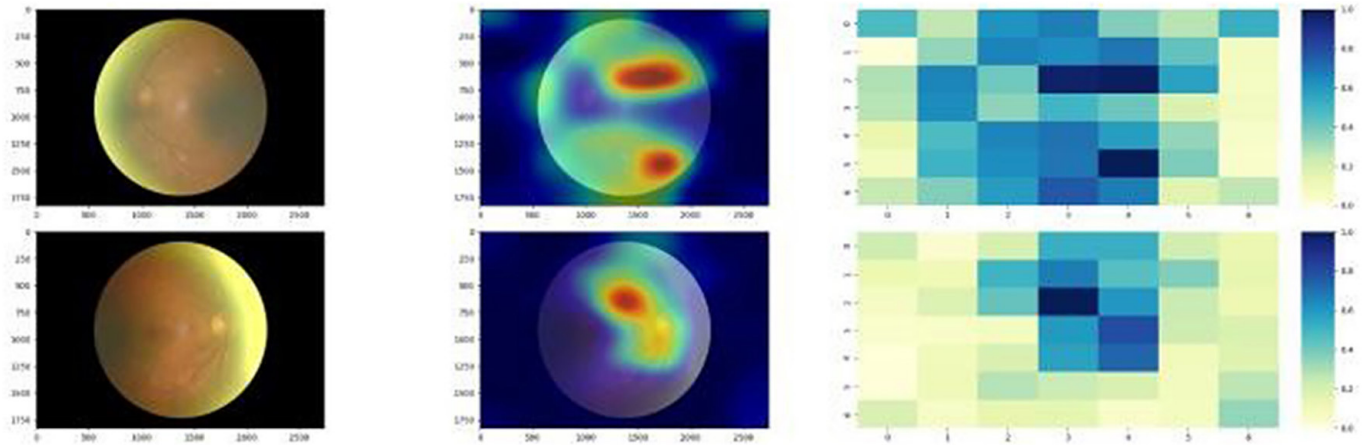

Fig. 9. Normal fundus image, showing better recognition to certain areas of nerve cells in VGG-19

8 CONCLUSION AND FUTURE WORK

This research paper presents a significant advancement in the field of medical imaging, particularly in the diagnosis of retinal diseases such as cataracts. It underscores the efficacy of DL models, with VGG-19 emerging as the most promising model in terms of accuracy, precision, and recall for classifying cataract and normal fundus images. Additionally, the application of XAI techniques, notably Grad-CAM, enriches the understanding of the decision-making processes of these models, offering a layer of transparency and trust that is crucial in clinical settings. The paper not only highlights the potential of AI in enhancing diagnostic processes but also calls for further exploration into the application of DL models for a broader range of eye conditions, aiming to leverage technological advancements for better healthcare outcomes. The paper suggests directions for future research, including the exploration of additional DL models, expanding the dataset for a broader range of disease categories, and developing XAI methods that are more sophisticated to enhance the explainability of AI-driven diagnostic systems. Further, an adaptive control approach-based neural approximation for a category of uncertain fractional-order systems with actuator nonlinearities and output constraints as described in [19] can be investigated for complex fundus images.

9 ACKNOWLEDGEMENT

The authors thank the host institution Vellore Institute of Technology, Chennai, Tamil Nadu, India for providing necessary support to carry out this research work.

10 REFERENCES

[1] P. Courtiol *et al.*, "Deep learning-based classification of mesothelioma improves prediction of patient outcome," *Nature Medicine*, vol. 25, pp. 1519–1525, 2019. https://doi.org/10.1038/s41591-019-0583-3

[2] L-P. Cen, J. Ji, J.-W. Lin, and S.-T. Ju, "Automatic detection of 39 fundus diseases and conditions in retinal photographs using deep neural networks," *Computer Methods and Programs in Biomedicine*, vol. 226, p. 116961, 2020.

[3] T. C. Hollon *et al.*, "Near real-time intraoperative brain tumor diagnosis using stimulated Raman histology and deep neural networks," *Nature Medicine*, vol. 26, pp. 52–58, 2020. https://doi.org/10.1038/s41591-019-0715-9

[4] H. W. Loh, C. Ping Ooi, S. Seoni, P. Datta Barua, F. Molinari, and U. Rajendra Acharya, "Application of explainable artificial intelligence for healthcare: A systematic review of the last decade (2011–2022)," *Computer Methods and Programs in Biomedicine*, vol. 226, p. 107161, 2022. https://doi.org/10.1016/j.cmpb.2022.107161

[5] E. Ennadifi, S. Laraba, D. Vincke, B. Mercatoris, and B. Gosselin, "Wheat diseases classification and localization using convolutional neural networks and gradCAM visualization," in *2020 International Conference on Intelligent Systems and Computer Vision (ISCV)*, Fez, Morocco, 2020, pp. 1–5. https://doi.org/10.1109/ISCV49265.2020.9204258

[6] S. N. Sabrin, A. K. M. M. Islam, S. Swakkhar, and I. Salekul, "Towards development of IoT-ML driven healthcare systems: A survey," *Journal of Network and Computer Applications*, vol. 196, p. 103244, 2021. https://doi.org/10.1016/j. jnca.2021.103244

[7] F. Shi *et al.*, "Review of artificial intelligence techniques in imaging data acquisition, segmentation, and diagnosis for COVID-19," *IEEE Reviews in Biomedical Engineering*, vol. 14, pp. 4–15, 2021. https://doi.org/10.1109/RBME.2020.2987975

[8] X. Bai *et al.*, "Explainable deep learning for efficient and robust pattern recognition: A survey of recent developments," *Pattern Recognition*, vol. 120, p. 108102, 2021. https://doi.org/10.1016/j.patcog.2021.108102

[9] H. Lin, X. Cheng, X. Wu, and D. Shen, "CAT: Cross attention in vision transformer," in *IEEE International Conference on Multimedia and Expo (ICME)*, Taipei, Taiwan, 2022, pp. 1–6. https://doi.org/10.1109/ICME52920.2022.9859720

[10] W. Dong, "Low-rank rescaled vision transformer fine – tuning: A residual design approach," in *2024 IEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Seattle, WA, USA, 2024, pp. 16101–16110. https://doi.org/10.1109/CVPR52733.2024.01524

[11] A. El-Nouby, N. Neverova, I. Laptev, and H. Jégou, "Training vision transformers for image retrieval," *arXiv preprint arXiv:2102.05644*, 2021.

[12] J. Nodirov, A. B. Abdusalomov, and T. K. Whangbo, "Attention 3D U-Net with multiple skip connections for segmentation of brain tumor images," *Sensors*, vol. 22, no. 17, p. 6501, 2022. https://doi.org/10.3390/s22176501

[13] M. Aloraini, A. Khan, S. Aladhadh, S. Habib, M. F. Alsharekh, and M. Islam, "Combining the transformer and convolution for effective brain tumor classification using MRI images," *Applied Sciences*, vol. 13, no. 6, p. 3680, 2023. https://doi.org/10.3390/app13063680

[14] J. Alyami, A. Rehman, T. Sadad, M. Alruwaythi, T. Saba, and S. A. Bahaj, "Automatic skin lesions detection from images through microscopic hybrid features set and machine learning classifiers," *Microscopy Research and Technique*, vol. 85, no. 11, pp. 3600–3607, 2022. https://doi.org/10.1002/jemt.24211

[15] P. Courtiol, M. Sinanu, F. Mosqueira, A. L. Moreira, S. Azevedo, and G. Campanella, "Deep learning-based classification of mesothelioma improves prediction of patient outcome," *Nature Medicine*, vol. 25, pp. 1519–1525, 2019. https://doi.org/10.1038/s41591-019-0583-3

[16] Kaggle, https://www.kaggle.com/datasets/gunavenkatdoddi/preprocessed-eye-diseases-fundus-images.

[17] Halit Çetiner, "Cataract disease classification from fundus images with transfer learning based deep learning model on two ocular disease datasets," *Gümüşhane Üniversitesi Fen Bilimleri Dergi (GUFBD)*, vol. 13, no. 2, pp. 256–269, 2023. https://doi.org/10.17714/gumusfenbil.1168842

[18] L. Murugan, "Analysis of ANN routing method on integrated IOT with WSN," *International Journal of Interactive Mobile Technologies (iJIM)*, vol. 18, no. 16, pp. 197–210, 2024. https://doi.org/10.3991/ijim.v18i16.48983

[19] F. Zouari and A. Boubellouta, "Neural approximation-based adaptive control for pure-feedback fractional-order systems with output constraints and actuator nonlinearities," in *Advanced Synchronization Control and Bifurcation of Chaotic Fractional-Order Systems*, A. Boulkroune and S. Ladaci, Eds., 2018, pp. 468–495. https://doi.org/10.4018/978-1-5225-5418-9.ch015

11 AUTHORS

Ayush Gupta is a undergraduate student of School of Computer Science and Engineering, Vellore Institute of Technology, Chennai, Tamil Nadu, India. He has done several industry internships in the area of machine learning and deep learning. He has also completed several industrial consultancy projects. His research interests include machine learning, deep learning and image analytics (E-mail: ayushgupta6194@gmail.com).

Dr. D. Jeya Mala is currently working as Professor in the School of Computer Science and Engineering (SCOPE), Vellore Institute of Technology, Chennai, Tamil

Nadu, India. She has received her doctorate degree in the area of "Artificial Intelligence in Software Engineering" from Anna University, Chennai, Tamil Nadu, India. She has more than 20 years of teaching and research and 4 years of industrial experience. She was granted with a design patent and published a utility patent from IP, Govt. of India. She has published more than 60 papers in reputed, refereed; SCI and Scopus indexed journals and conferences and book chapters. She has published one text book and three edited books under international publishers. She has published a MOOC course for Udemy. She has formed the reviewer board of several international journals and conferences. She is a member of professional bodies, editorial boards and technical programme committees of several reputed journals and conferences. Her research interests include artificial intelligence, explainable AI, quantum AI, software engineering, machine learning, cyber security and block chain (E-mail: jeyamala.d@vit.ac.in).

Vishal Kumar Yadav is a undergraduate student of School of Computer Science and Engineering, Vellore Institute of Technology, Chennai, Tamil Nadu, India. He has completed several certifications and academic projects. His research interests include machine learning and deep learning (E-mail: vishal100403@gmail.com).

Mayank Arora is a undergraduate student of School of Computer Science and Engineering, Vellore Institute of Technology, Chennai, Tamil Nadu, India. He has completed several certifications and research projects. His research interests include machine learning and deep learning (E-mail: aroramayank829@gmail.com).