An interpretable diagnosis of retinal diseases using Vision Transformer and Grad-CAM

Tahsin Zaman Jilan, Mahdi Hasan Bhuiyan, Sumit Haldar, Maisha Shabnam Chowdhury,
and Nazifa Bushra
Department of Computer Science and Engineering (CSE)
School of Data and Sciences (SDS)
Brac University
66 Mohakhali, Dhaka - 1212, Bangladesh
{tahsin.zaman.jilan, mahdi.hasan.bhuiyan, sumit.haldar, maisha.shabnam.chowdhury, nazifa.bushra,}@g.bracu.ac.bd

***Abstract*—Early detection of retinal diseases is very crucial to prevent partial or complete blindness. This research presents a novel, interpretable diagnosis framework that combines VGG-16 and Swin Transformer models and then visualized through Grad-CAM to address the challenge of multi-label classification in retinal disease detection. Using the OCT images, we have developed a unique hybrid model that integrates the power of both Convolutional Neural Networks and Vision Transformers. This model does not only classifies images but also gives a clear visual explanations of the network's decisions through gradient-weighted class activation mapping which is also known as GRAD-CAM. The VGG-16 achieved an accuracy of 88.88% whereas the Vision Transformer reached 91.39%. On the other side, Our fine-tuned hybrid model significantly outperformed individual components. Achieving an accuracy of 98.8%, this model demonstrated its potential as a powerful tool for retinal disease detection.**

***Index Terms*—Deep Learning, Vision Transformers, Ocular diseases screening, Grad-CAM, Detection, Diagnosis, classification**

I. Introduction

Retinal disease is a phenomenon that affects the retina, the light-sensitive membrane at the back of the eye that sends visual signals to the brain. Conditions such as age-related macular degeneration (AMD), diabetic retinopathy and retinal detachment constitute one of the most frequently found top health problems around the world; affecting multiple millions of people of all ages. It is estimated that the number of people affected by retinal disorders will exceed 42 million in developed countries and 82 million in others by 2030 [1]. Age-related macular degeneration (AMD) represents a primary cause of blindness in the elderly and has estimated global prevalence rates of 288 million in 2010 and a projected increase to 600 million by 2020 [2].

A. Integrating AI for efficiency and precision in Analysis

The early detection and treatment of retinal diseases are essential in maintaining vision and preventing vision loss [1], [2]. However, diagnosing retinal diseases with AI poses significant challenges due to their subtle symptoms and complex image patterns.

1) Comprehension in diagnosis should not be Compromised: Although conventional deep learning approaches have shown promise in handling retinal images [3], [4], they often lack the interpretability needed for clinical deployment in medicine. Recent advances such as Grad-CAM (Gradient-weighted Class Activation Mapping) and Vision Transformers (ViTs) have proven to mitigate these interpretability challenges. Grad-CAM creates a heatmap indicating important areas with respect to the model's predictions [5], while Vision Transformers apply an attention mechanism to localize relevant regions in an image [6], [7]. This research employs a hybrid methodology that amalgamates VGG-16, a convolutional neural network, with the Swin Transformer, a variation of Vision Transformer, to create a highly precise diagnostic system for retinal disorders. The model attains an accuracy of 98.8%, representing a substantial enhancement over current methodologies. Furthermore, Grad-CAM is utilised to elucidate significant regions of the image pertaining to conditions such as Normal, CNV, DRUSEN, and DME, hence augmenting the model's interpretability for healthcare practitioners.

B. Primary Contributions of the Study

This study presents a comprehensive and elucidative framework for diagnosing retinal diseases, demonstrating considerable advancements compared to current methodologies. The key contributions are outlined below.

1) Employing Vision Transformers for Disease Classification: This study utilizes Vision Transformers for effective picture classification and the detection of specific retinal diseases.

2) Enhancing Interpretability with Grad-CAM: To tackle the interpretability challenge, Grad-CAM is incorporated to highlight significant image regions. This allows for better identification of focal areas in retinal images through attention mechanisms, improving classification accuracy.

3) Developing a Hybrid Model: A hybrid model integrating VGG-16 and Swin Transformer is proposed, achieving an accuracy of 98.8%. This approach leverages the strengths of both architectures to enhance performance.

4) Improving Clinical Confidence through Visualization: Grad-CAM is used to provide visual explanations of model predictions, increasing clinical confidence and comprehension of the diagnostic process.

C. Challenges in Existing Methods

Designing reliable retinal disease detection systems poses a range of critical challenges to overcome:

- Lack of Explainability: Conventional deep learning methods lack interpretability for a medical case, compromising medical professionals' trust and understanding.
- Constraints in Datasets: Smaller datasets, unbalanced datasets, and variable picture quality affect model generalizability and performance.
- Architectural Challenges: Conventional architectures of CNNs suffer with unbalanced datasets and overall and localized information extraction.

  They demand new, sophisticated techniques with capabilities to counteract medical and technical requirements with high accuracy and interpretability

II. Literature Review

Most of the weight in the deep neural network architecture, which was specifically developed for natural language processing tasks, comes from the self-attention mechanism that forms the foundation of the Transformer. As shown by K. Han et al. [8] applications for transformer-based models have also been extended from NLP to computer vision, showing superior performance over conventional convolutional and recurrent networks on several visual benchmarks. These models have achieved state-of-the-art performance on a number of vision tasks compared to CNN-based models, owing to their ability to capture long-range relations.

S. S. M. Sheet et al. The CLAHE technique was found to effectively enhance retinal contrast in medical images in [9]. While good accuracy was obtained using data augmentation with the RESNET50 model, it suggested limitations on training datasets. However, one important operation in a medical setting is visible in the microscopic structure of the tissue, so we need to help it: This is where CLAHE comes in. However, they were only able to classify into five classes, suggesting that more sophisticated models are required to potentially meet a wider array of retinal diseases.

As a comparative, VTGAN which is a semi-supervised GAN proposed by S. A. Kamran et al. [10], also for demarcating healthy versus sick retinas and generating joints fluorescence angiography FA images. Their superior performance is validated using widely accepted quantitative metrics such as Frechet Inception Distance (FID) and Kernel Inception Distance (KID), which demonstrates the importance of quantifying medical image synthesis quality.

On this note, A. K. Bitto et al. [11] employed CNN based transfer learning techniques to classify normal eyes, cataract and conjunctivitis by achieving an accuracy of 97.86

M. Subramanian et al. In a different paper, [12] also built their models with transfer learning approaches. That meant a big reduction in the computation and the volume of training sets required. These observations highlighted the importance of finding light-weight deep learning architectures with high performance and low computational cost. These models will be even more beneficial for therapeutics if they are generalised to incorporate real time applications.

As pointed out by V. Annavarjula [13], OCT has become the accepted forerunner in noninvasive, high-resolution imaging of eye internal microstructures. Various researchers have been able to classify retinal diseases such as AMD and DME more accurately than CNNs by using OCT and Vision Transformers in conjunction. Early detection and treatment of such diseases is essential for maintaining vision, more so in the senior population.

Moreover, Deep learning algorithms have revolutionized imaging based diagnosis of systemic diseases. For instance, R. Fan et al. Still, recent work has shown that medical imaging techniques (such as CT scans and X-rays) can be used to detect COVID-19 [14]. Combined with YOLO-based object detection methods, these techniques yielded accuracies above 90

Z. Ma, et al. [15] proposed a new transformer-based architecture tailored for CXR data in the image analysis of chest X-rays that achieved better accuracy in detecting pneumonia than traditional CNNs. The second work also said their model's interpretability was further corroborated by their use of Grad-CAM for visualisation which could indicate how it could be used across a wider range of vision tasks in the future.

Vision Transformers also show potential for the diagnosis of cardiovascular disease. K. Przystalski et al. [16] also stressed the importance of the dataset quality in obtaining accurate results. Despite their potential, convolutional networks often created superior scores on some tasks than transformer-based models, suggesting that transformer designs could be further optimized.

Vision Transformers have ever been applied to infrastructure assessments outside of the medical field. The authors used LeViT, a transformer-based asphalt pavement image classification model [17]. By employing visualisation tools like Grad-CAM and Attention Rollout, their model eclipsed six state-of-the-art DL models in both accuracy and interpretable.

G. Cai et al. (cai2022multimodal) designed a transformer network using clinical and image metadata for the classification of skin diseases in multimodal medical data analysis. The addition of mutual attention mechanisms significantly improved performance on benchmark datasets, thus demonstrating the potential benefit of multimodal approaches in medical diagnosis.

To conclude, the field of medical diagnostics has been revolutionized by the introduction of deep learning and Vision Transformers, enabling unprecedented degrees of accuracy and efficiency in clinical applications. But there remains scope for additional work on aspects such as real-time deployment, model interpretability and dataset constraints. Future research on integrating these advanced models into clinical practice should focus on improving patient outcomes and optimizing resource use.

III. Data Collection

A. Data Description

The dataset we used is publicly available — ”Large Dataset of Labeled Optical Coherence Tomography (OCT) and Chest X-Ray Images” [18] and contains machine learning retinal images. It released in Jun 2018 with has 88416 training images, 1000 test images and 32 validation images for CNV, DME, DRUSEN and NORMAL. The dataset is highly imbalanced, as shown in Fig. 1. To address this, we applied random oversampling to balance the class proportions (Fig. 2).

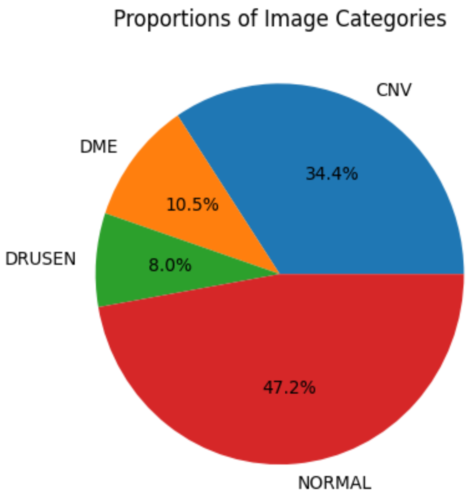


Fig. 1: Dataset class proportions

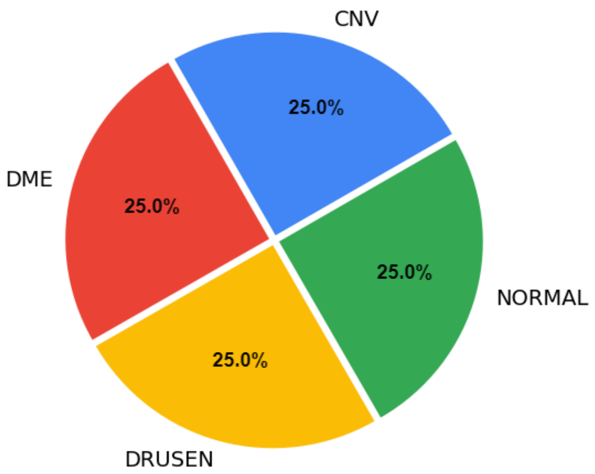


Fig. 2: Dataset class proportions after balancing

1) The dataset consists of four distinct classes, each representing a specific retinal condition:

a) CNV (Choroidal Neovascularization): This class refers to abnormal blood vessel growth in the retina, commonly associated with wet age-related macular degeneration (AMD) [19]. CNV can lead to severe vision impairment if not detected early.

b) DRUSEN: Drusen are small yellow or white deposits that form under the retina, often indicative of age-related macular degeneration (AMD) [20]. While they do not directly affect vision, their presence can increase the risk of further retinal damage.

c) DME (Diabetic Macular Edema): DME is characterized by retinal swelling caused by diabetes, leading to fluid leakage into the macula and potential vision loss [21]. Early detection and treatment are crucial to prevent severe outcomes.

d) NORMAL: This class represents healthy retinal images, with no visible signs of disease or damage. It serves as a baseline for comparing abnormalities in the other classes.

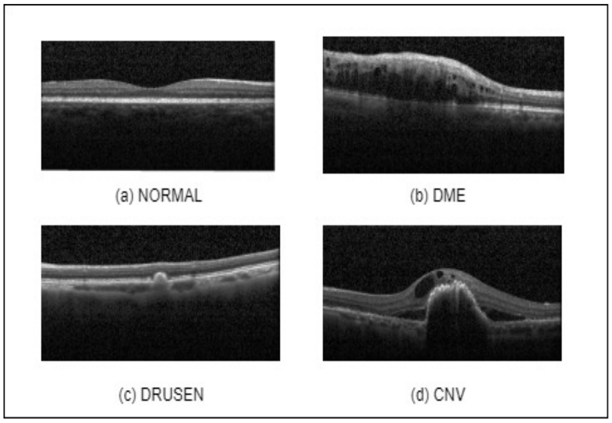

(a) NORMAL (b) DME (c) DRUSEN (d) CNV

Fig. 3: Examples of retinal image classes

B. Data Pre-Processing

The data were processed in Jupyter Notebook. We then carried out random oversampling on the data to balance the class proportions. The images were resized at78×78 pixels and data augmentation techniques (rotation, flipping, zooming and translation) were applied. Then we split our data into 90 % training and 5 % testing and validation. Fig. 4 illustrates the pre-processing workflow.

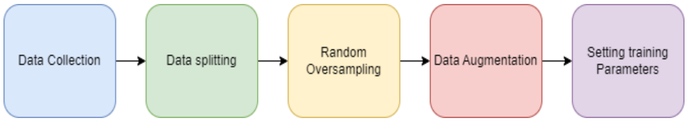


Fig. 4: Data pre-processing flowchart

1) Model Training Parameters: The parameters used for training the model are summarized in Table I.

IV. Proposed Model

A. Overview and Motivation

We present a new hybrid model, VGG-16 and Swin Transformer, in an effort to counteract weaknesses in conventional CNN architectures. VGG-16’s local feature extraction, capabilities of Swin Transformer in processing contextual information, and an inbuilt mechanism for handling imbalanced datasets for datasets of retinal disease will be utilized together.

TABLE I: Model Parameters

Parameters	**Value**
Image Size	78
Patch Size	6
Weight Decay	0.0001
Projection Dimension	64
Number of Patches	$\left(\frac{78}{6}\right)^2 = 169$
Number of Encoders	8
Transformer Unit	(128, 64)
Heads	4
MLP Head	(2048, 1024)

B. *Architectural Framework*

Our model consists of three primary components:

1) Input Processing:

- Parallel processing through VGG-16 and Swin Transformer
- Resolution: 78×78×3 RGB format

2) Dual Processing Branches: **VGG-16 Branch:**

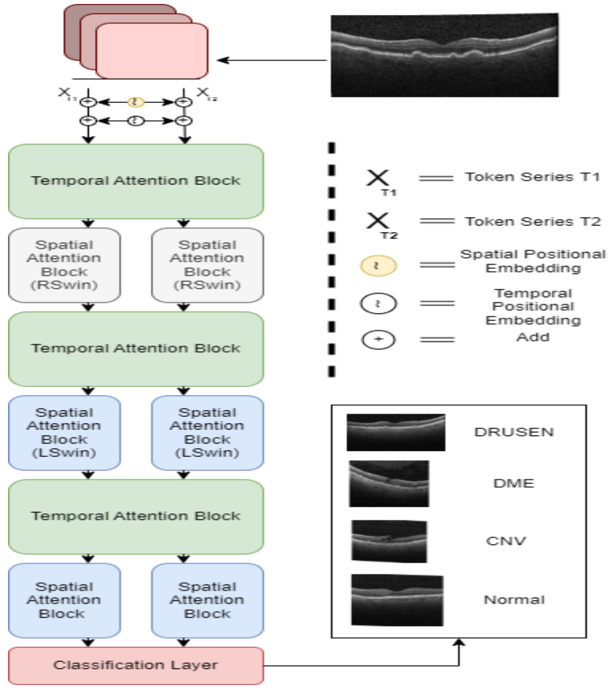
- Five blocks of convolutional layers (64 to 512 filters)
- Output: 512-dimensional feature vectors

Swin Transformer Branch:

- Patch size: 6×6, Embedding dimension: 64
- 8 transformer blocks with 4-head attention
- Output: 169-dimensional feature vectors

3) Feature Fusion:

- Concatenation of VGG-16 and Transformer features
- 1×1 convolution with batch normalization
- Softmax classification



X_{T1}
X_{T2}
Temporal Attention Block
Spatial Attention Block (RSwin)
Spatial Attention Block (RSwin)
Temporal Attention Block
Spatial Attention Block (LSwin)
Spatial Attention Block (LSwin)
Temporal Attention Block
Spatial Attention Block
Spatial Attention Block
Classification Layer
X_{T1} = Token Series T1
X_{T2} = Token Series T2
= Spatial Positional Embedding
= Temporal Positional Embedding
+ = Add
DRUSEN
DME
CNV
Normal

Fig. 5: Hybrid architecture combining Vision Transformer and CNN components

C. *Implementation Details*

1) Data Preprocessing:

- CLAHE enhancement and noise reduction
- Geometric corrections for alignment
- Edge preservation filtering

2) Training Configuration:

- Adam optimizer (lr=1e-4, batch size=32)
- Categorical cross-entropy loss
- Cosine annealing with warm restarts

D. *Key Benefits*

- **Architecture:** Multi-scale feature representation, efficient preprocessing
- **Performance:** Enhanced accuracy for imbalanced classes, better generalization
- **Clinical:** Real-time processing, explainable decisions, adaptable to various protocols

V. Result Analysis

A. *Demonstrating performance of the models*

First, we trained the VGG16 CNN model on our dataset for retinal disease classification, for which we obtained an accuracy of 88.88 after 10 epochs. Below are the graphs for accuracy and loss.

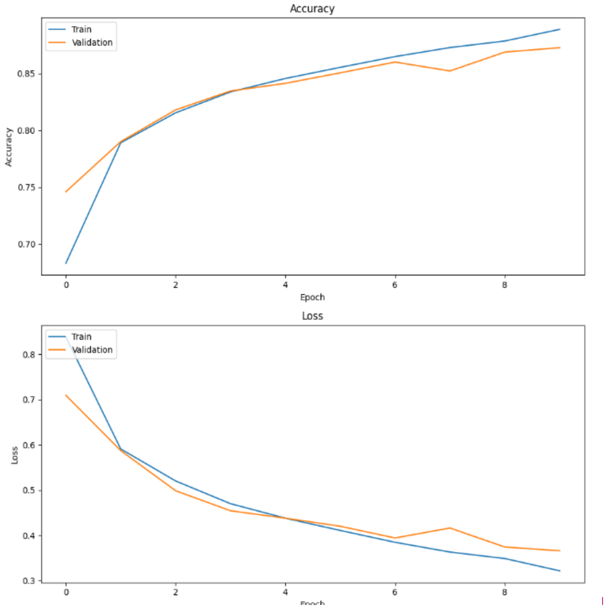


Fig. 6: Accuracy & Loss gained from CNN(VGG16)

Then we evaluated a base ViT(Vision Transformer) model with our dataset to do some classification, training for 10 epochs. We achieved an 91.39% accuracy, outperforming the VGG16 model. Below are the associated accuracy and loss graphs showing its performance.

Finally, we proposed to combine CNN (VGG16) with Swin Transformers as CNNs are capable of capturing spatial hierarchies while ViTs use attention mechanisms. 2E, this model improves accuracy in classifying retinal diseases as normal,

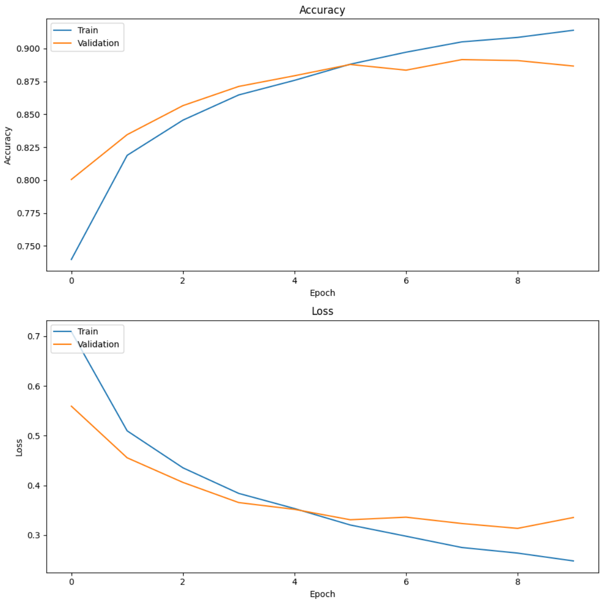


Fig. 7: Accuracy & Loss gained from ViT(Swin Transformer)

CNV, DRUSEN, and DME by extracting subtle features in retinal images.

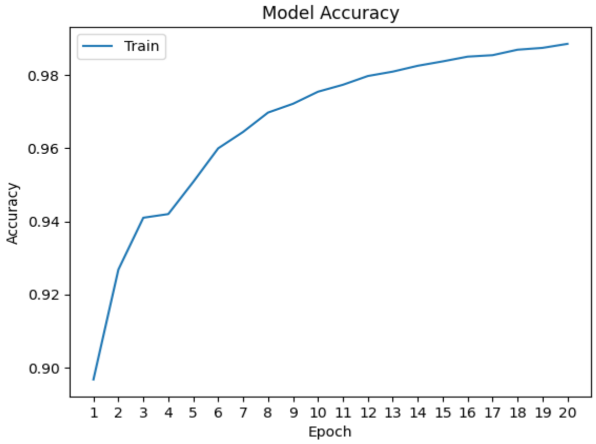


Fig. 8: Accuracy gained from Proposed Model

Our model fuses the best of both worlds, by combining the feature extraction capabilities of CNNs with the self-attention capabilities of ViT. As a result, as shown in studies like [22], ViT has been able to be as reliable and precise as CNNs. This hybridization provides local and global context while allowing it to learn the disease pattern of the retina. The model is designed to leverage the strengths of both between Convolutional embeds and Vision transformer hopes to take advantage of their specialties and provide better solutions for delivering accurate classification of Normal, CNV, DRUSEN and DME on retinal photos.

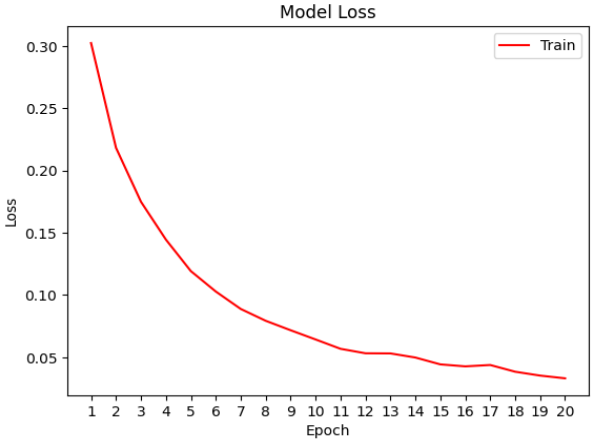


Fig. 9: Loss gained from Proposed Model

B. *Classification report confusion matrix*

As we see the classification report our model performs exceptionally well on CNV, DME, DRUSEN, and Normal classes with an accuracy of 96% to 100%. It accomplishes this with a total detection-rate of 95%–100% and an F1-Score of 97%–99%, indicating good consistency. The overall accuracy of the model is 98% which makes it highly reliable to detect patients with retinal diseases.

TABLE II: Classification Report

Class	**Precision (%)**	**Recall (%)**	**F1-Score (%)**	**Support**
CNV	96	100	98	242
DME	100	98	99	242
DRUSEN	98	95	97	242
NORMAL	98	99	98	242
Accuracy			98.80	968
Macro Avg	98	98	98	968
Weighted Avg	98	98	98	968

C. *Comparison of results*

According to this research[1], ResNET101 and EfficientNet-B0 had an accuracy of 0.94 percent and 0.88 percent respectively in improving retinal blood vessel segmentation and fundus categorization using DRIVE and FIVE datasets. In addition, they have demonstrated that emphasizing important regions Score-CAM is the most successful. Another research paper regarding Transformers performance on multi-modal medical image classification[2], combined CNN and Transformer architecture and achieving an improved score of 10 to 1.9 percent in average accuracy. On comparision our model produced remarkable results.

In this paper, a custom model is presented for image classification, and the model managed to score 98.8% accuracy. It surpasses the benchmarks of common models, including Swin-ViT, ResNet, VGG16, VGG19, and ViT, indicating its power

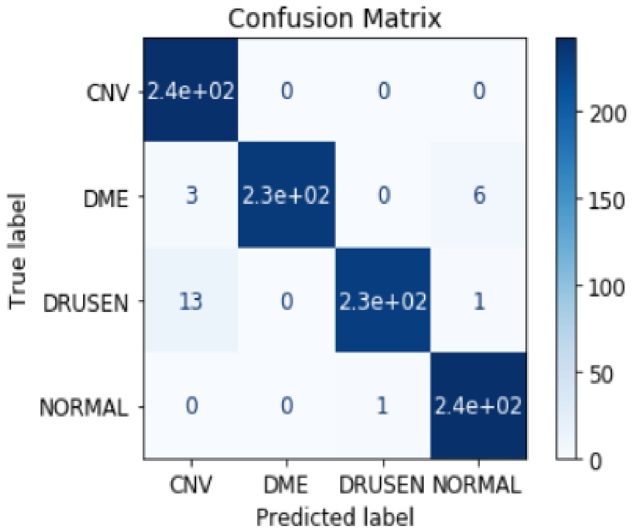


Fig. 10: Confusion matrix of our proposed model

and competitiveness. And its higher accuracy highlights the potential of this model as an advanced solution, and its value for the domain for image classification, standing out as an important contribution for the area.

TABLE III: Comparing different models' Accuracy

Model	Accuracy (%)
ResNet [23]	97.3
VGG19 [24]	97.8
VIT [25]	82.0
VGG16	88.88
Swin-ViT	91.39
Ours	98.8

D. *Grad-CAM Technique for Visualization*

Our model predicts each image to be in either Normal, CNV, DRUSEN and DME classes. We use Grad-CAM [5] to visualize the parts of the input regions that lead the model to its prediction using activation maps on FD (Fundus Digital) images, similar to what is done in [26] for interpreting ECG signal activity . This improves the transparency of models.

VI. Grad-CAM Analysis

In Grad-CAM images, the most contributive regions are highlighted in red, while less significant areas are in blue. The model zooms into certain FD (Fundus Digital images) regions of each class, as shown in the figures. The red regions are indicative of important areas for CNV, DRUSEN and DME. These results show the model's capacity to differentiate between Normal, CNV, DRUSEN, and DME based on distinctive fractal properties.

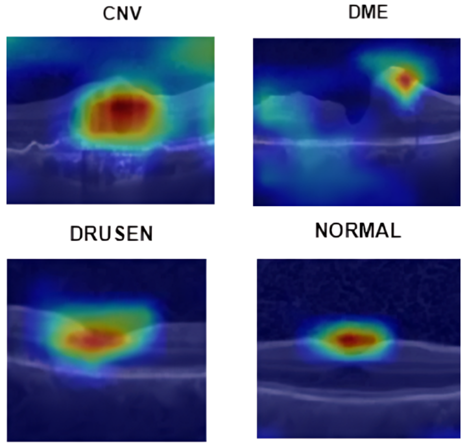


Fig. 11: Grad-CAM technique for visualization

VII. Conclusion

The proposed model, integrating VGG16 and Swin Transformer with Grad-CAM visualization, demonstrates a robust and interpretable approach to retinal disease diagnosis. By leveraging VGG16's feature extraction capabilities and the Swin Transformer's self-attention mechanism, this hybrid model captures both local and global contextual information in retinal images. Grad-CAM enhances interpretability, providing critical visual insights into the model's decisions, which is crucial for clinical applications.

Achieving a remarkable accuracy of 98.8% in classifying Normal, CNV, DRUSEN, and DME conditions, this approach surpasses existing benchmarks such as VGG16 at 88.88% and Swin Transformer at 94.61%. The visual heatmaps generated by Grad-CAM effectively highlight key regions contributing to classification, enhancing trust and usability in medical diagnostics.

While the results are promising, future efforts should address limitations such as dataset size, bias, and the need for extensive clinical validation. Expanding to larger, diverse datasets and comparing with state-of-the-art approaches will further refine the model's sensitivity, specificity, and practical utility. Ultimately, this framework holds great potential for early detection, precise diagnosis, and better patient outcomes in retinal disease management.

References

[1] H. Nazimul, K. Rohit, and H. Anjli, "Trend of retinal diseases in developing countries," *Expert Review of Ophthalmology*, vol. 3, no. 1, pp. 43–50, 2008.

[2] J. W. Miller, L. L. D'Anieri, D. Husain, J. B. Miller, and D. G. Vavvas, "Age-related macular degeneration (amd): a view to the future," 2021.

[3] J. De Fauw, J. R. Ledsam, B. Romera-Paredes, S. Nikolov, N. Tomasev, S. Blackwell, H. Askham, X. Glorot, B. O'Donoghue, D. Visentin, *et al.*, "Clinically applicable deep learning for diagnosis and referral in retinal disease," *Nature medicine*, vol. 24, no. 9, pp. 1342–1350, 2018.

[4] M. Badar, M. Haris, and A. Fatima, "Application of deep learning for retinal image analysis: A review," *Computer Science Review*, vol. 35, p. 100203, 2020.

[5] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, "Grad-cam: Visual explanations from deep networks via gradient-based localization," in *Proceedings of the IEEE international conference on computer vision*, pp. 618–626, 2017.

[6] D. Zhou, B. Kang, X. Jin, L. Yang, X. Lian, Z. Jiang, Q. Hou, and J. Feng, "Deepvit: Towards deeper vision transformer," *arXiv preprint arXiv:2103.11886*, 2021.

[7] A. Parvaiz, M. A. Khalid, R. Zafar, H. Ameer, M. Ali, and M. M. Fraz, "Vision transformers in medical computer vision—a contemplative retrospection," *Engineering Applications of Artificial Intelligence*, vol. 122, p. 106126, 2023.

[8] K. Han, Y. Wang, H. Chen, X. Chen, J. Guo, Z. Liu, Y. Tang, A. Xiao, C. Xu, Y. Xu, Z. Yang, Y. Zhang, and D. Tao, "A survey on vision transformer," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 45, no. 1, pp. 87–110, 2023.

[9] S. S. M. Sheet, T.-S. Tan, M. As'ari, W. H. W. Hitam, and J. S. Sia, "Retinal disease identification using upgraded clahe filter and transfer convolution neural network," *ICT Express*, vol. 8, no. 1, pp. 142–150, 2022.

[10] S. A. Kamran, K. F. Hossain, A. Tavakkoli, S. L. Zuckerbrod, and S. A. Baker, "Vtgan: Semi-supervised retinal image synthesis and disease prediction using vision transformers," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 3235–3245, 2021.

[11] A. K. Bitto and I. Mahmud, "Multi categorical of common eye disease detect using convolutional neural network: a transfer learning approach," *Bulletin of Electrical Engineering and Informatics*, vol. 11, no. 4, pp. 2378–2387, 2022.

[12] M. Subramanian, M. S. Kumar, V. Sathishkumar, J. Prabhu, A. Karthick, S. S. Ganesh, and M. A. Meem, "Diagnosis of retinal diseases based on bayesian optimization deep learning network using optical coherence tomography images," *Computational Intelligence and Neuroscience*, vol. 2022, 2022.

[13] V. Annavarjula, "Computer-vision based retinal image analysis for diagnosis and treatment," 2017.

[14] R. Fan, K. Alipour, C. Bowd, M. Christopher, N. Brye, J. A. Proudfoot, M. H. Goldbaum, A. Belghith, C. A. Girkin, M. A. Fazio, *et al.*, "Detecting glaucoma from fundus photographs using deep learning without convolutions: Transformer for improved generalization," *Ophthalmology Science*, vol. 3, no. 1, p. 100233, 2023.

[15] Y. Ma and W. Lv, "Identification of pneumonia in chest x-ray image based on transformer," *International Journal of Antennas and Propagation*, vol. 2022, 2022.

[16] K. Przystalski, M. Jungiewicz, P. Wawryka, and K. Sabatowski, "Vision transformer in stenosis detection of coronary arteries," *Available at SSRN 4175204*.

[17] Y. Chen, X. Gu, Z. Liu, and J. Liang, "A fast inference vision transformer for automatic pavement image classification and its visual interpretation method," *Remote Sensing*, vol. 14, no. 8, p. 1877, 2022.

[18] D. Kermany, "Large dataset of labeled optical coherence tomography (oct) and chest x-ray images," Jun 2018.

[19] A. D. Kulkarni and B. D. Kuppermann, "Wet age-related macular degeneration," *Advanced drug delivery reviews*, vol. 57, no. 14, pp. 1994–2009, 2005.

[20] C. A. Curcio, "Soft drusen in age-related macular degeneration: biology and targeting via the oil spill strategies," *Investigative ophthalmology & visual science*, vol. 59, no. 4, pp. AMD160–AMD181, 2018.

[21] X. Zhang, H. Zeng, S. Bao, N. Wang, and M. C. Gillies, "Diabetic macular edema: new concepts in patho-physiology and treatment," *Cell & bioscience*, vol. 4, no. 1, pp. 1–14, 2014.

[22] S. Cuenat and R. Couturier, "Convolutional neural network (cnn) vs vision transformer (vit) for digital holography," in *2022 2nd International Conference on Computer, Control and Robotics (ICCCR)*, pp. 235–240, IEEE, 2022.

[23] D. Wang and L. Wang, "On oct image classification via deep learning," *IEEE Photonics Journal*, vol. 11, no. 5, pp. 1–14, 2019.

[24] J. Kim and L. Tran, "Retinal disease classification from oct images using deep learning algorithms," in *2021 IEEE Conference on Computational Intelligence in Bioinformatics and Computational Biology (CIBCB)*, pp. 1–6, IEEE, 2021.

[25] L. Cai, C. Wen, J. Jiang, C. Liang, H. Zheng, Y. Su, and C. Chen, "Classification of diabetic maculopathy based on optical coherence tomography images using a vision transformer model," *BMJ Open Ophthalmology*, vol. 8, p. e001423, 12 2023.

[26] V. Jahmunah, E. Y. K. Ng, R.-S. Tan, S. L. Oh, and U. R. Acharya, "Explainable detection of myocardial infarction using deep learning models with grad-cam technique on ecg signals," *Computers in Biology and Medicine*, vol. 146, p. 105550, 2022.