An Ensemble Deep Learning Framework for Robust Multi-Class Retinal Disease Detection Using OCT Images

Rajababu Pukkala[1], S. Jhansi Rani[2]
[1]Research Scholar, Department of Computer Science and Systems Engineering, Andhra University, India
[2]Associate Professor, Department of Computer Science and Systems Engineering, Andhra University, India

Abstract:
Automated retinal disease diagnosis using Optical Coherence Tomography (OCT) has gained significant attention due to the growing prevalence of vision-threatening conditions such as Choroidal Neovascularization (CNV), Diabetic Macular Edema (DME), Drusen, and Age-Related Macular Degeneration (AMD). While deep learning models have demonstrated promising performance, reliance on a single architecture often leads to limited robustness and poor generalization under real-world clinical variability. To address these limitations, this paper presents an ensemble-based deep learning decision framework that integrates multiple complementary architectures for reliable multi-class retinal disease classification. The proposed ensemble model combines Convolutional Neural Networks (CNN), ResNet-50, VGG-16, and Inception-V3, operating on deep feature representations extracted from preprocessed OCT images. Decision-level fusion using weighted averaging and majority voting is employed to enhance diagnostic stability and reduce model bias. Extensive experiments conducted on large-scale Kaggle and UCI OCT datasets demonstrate that the proposed ensemble framework significantly outperforms individual models, achieving an overall accuracy of 99.78%, sensitivity of 99.89%, and F1-score of 99.86%. The results confirm that ensemble learning substantially improves classification robustness, generalization, and clinical reliability, making the proposed framework well-suited for real-world retinal screening and telemedicine applications.

Introduction:

Retinal diseases remain a leading cause of irreversible vision loss worldwide, particularly among aging populations and individuals with chronic conditions such as diabetes. Disorders including CNV, DME, Drusen, and AMD often progress asymptomatically in early stages, making timely diagnosis critical for preventing permanent visual impairment. Optical Coherence Tomography (OCT) has emerged as a standard imaging modality for retinal assessment due to its ability to generate high-resolution cross-sectional views of retinal layers. However, the increasing volume of OCT scans has placed a substantial burden on ophthalmologists, highlighting the need for automated diagnostic support systems.

Recent advances in deep learning have enabled automated OCT-based disease classification with promising accuracy. Convolutional Neural Networks (CNNs) and transfer learning models such as ResNet, VGG, and Inception have shown strong capability in learning discriminative retinal features. Despite these advances, most existing approaches rely on a single deep learning architecture, which often suffers from sensitivity to noise, dataset imbalance, and reduced generalization across different imaging devices and populations.

Ensemble learning has emerged as an effective strategy to overcome these challenges by integrating multiple classifiers to exploit their complementary strengths. By combining diverse feature representations and decision patterns, ensemble models reduce variance, mitigate overfitting, and improve diagnostic reliability. Motivated by these advantages, this paper focuses on Objective-2, proposing an ensemble-based deep learning decision framework that integrates CNN, ResNet-50, VGG-16, and Inception-V3 for robust multi-class retinal disease detection from OCT images.

Vision is one of the most critical human senses, enabling individuals to interact effectively with their environment and maintain independence in daily life. The retina, a highly specialized neural tissue located at the back of the eye, plays a central role in visual perception by converting light signals into neural impulses. Structural or pathological changes in the retina can significantly impair vision, often leading to irreversible blindness if not detected and treated at an early stage. Retinal disorders such as Choroidal Neovascularization (CNV), Diabetic Macular Edema (DME), Drusen, and Age-Related Macular Degeneration (AMD) are among the leading causes of vision loss worldwide, particularly affecting elderly populations and individuals with diabetes. The global rise in aging populations, prolonged exposure to digital devices, and increasing prevalence of metabolic disorders have further intensified the clinical burden associated with retinal diseases. Optical Coherence Tomography (OCT) has emerged as a gold-standard, non-invasive imaging modality for retinal disease diagnosis due to its ability to generate high-resolution cross-sectional images of retinal layers. OCT enables clinicians to visualize subtle pathological features such as fluid accumulation, retinal thickening, and structural deformations with remarkable clarity. However, the interpretation of OCT scans is highly dependent on expert ophthalmologists and is often time-consuming, subjective, and prone to inter-observer variability. In resource-limited regions, the shortage of trained specialists further exacerbates delays in diagnosis, increasing the risk of disease progression and permanent vision loss.

Recent advances in deep learning have significantly transformed automated OCT image analysis, with Convolutional Neural Networks (CNNs) and transfer learning models achieving promising diagnostic accuracy. Nevertheless, single deep learning models often exhibit limited robustness when exposed to noisy OCT images, dataset imbalance, and variability across imaging devices and clinical settings. These limitations restrict their generalization capability and reduce clinical reliability.

To overcome these challenges, ensemble learning has gained considerable attention as a powerful strategy for improving diagnostic stability and accuracy. By integrating multiple deep learning architectures, ensemble models leverage complementary feature representations and reduce individual model bias. Motivated by this, the present work focuses on designing an ensemble-based decision framework that combines CNN, ResNet-50, VGG-16, and Inception-V3 to achieve highly reliable, generalized, and clinically viable multi-class retinal disease detection from OCT images.

Literature Survey:

Early automated retinal disease detection systems relied on handcrafted features and traditional machine learning classifiers such as Support Vector Machines and k-Nearest Neighbors. Although these methods achieved moderate success, their performance was highly dependent on feature engineering and lacked robustness to noise and imaging variability.

The introduction of deep learning significantly improved OCT image analysis. CNN-based architectures enabled automatic feature learning directly from raw OCT images, achieving higher accuracy and scalability. Transfer learning using pre-trained models such as VGG-16, ResNet-50, Inception-V3, and DenseNet further improved performance, particularly in limited-data scenarios.

Despite these advancements, several studies reported inconsistent performance when models were evaluated across different datasets or clinical environments. OCT images are often affected by speckle noise, low contrast, and device-specific artifacts, which negatively impact single-model learning. To address these limitations, ensemble learning approaches have been increasingly explored. Researchers demonstrated that combining multiple deep learning models improves robustness and reduces classification variance.

Recent studies have shown that ensemble CNN architectures outperform individual models in retinal disease classification by aggregating diverse feature representations. However, many existing works focus only on accuracy improvements without thoroughly addressing generalization, class-wise stability, and real-world clinical feasibility. Furthermore, integrated ensemble frameworks designed specifically for large-scale OCT datasets remain underexplored.

This gap motivates the proposed ensemble decision framework, which systematically combines multiple deep architectures to enhance multi-class retinal disease detection accuracy and generalization across heterogeneous OCT datasets.

Table 1 – Literature Survey

Title	Journal Name & Year	Methodology	Results	Research Gaps Identified
Deep Ensemble Learning for Retinal Image Classification	Translational Vision Science & Technology, 2022	Ensemble of CNN architectures applied to OCT images for retinal disease classification	Improved accuracy compared to single CNN models	Limited dataset diversity; no GAN-based augmentation; generalization across devices not evaluated
OCT-based Deep Learning Models for the Identification of Retinal Diseases	Scientific Reports, 2023	CNN and transfer learning models trained on large-scale OCT datasets	Achieved high classification accuracy for CNV and DME	Single-model dependency; limited robustness to noisy OCT scans
Hybrid GAN–CNN Approach for OCT Retinal Image Augmentation and Classification	IEEE Transactions on Medical Imaging, 2023	GAN-based data augmentation integrated with CNN classifiers	Enhanced performance for minority disease classes	Did not explore ensemble decision fusion; interpretability not addressed
Multi-stage Classification of Retinal OCT Using Multi-Scale Ensemble Deep Architecture	Bioengineering, 2024	Multi-scale ensemble deep learning architecture	Improved disease-wise accuracy	Increased model complexity; deployment feasibility not evaluated
Meta-Learning for Multimodal Retinal Diagnosis Combining OCT and Fundus Images	Translational Vision Science & Technology, 2024	Meta-learning with multimodal ensemble learning	Improved diagnostic robustness	Requires multi-modal data; OCT-only systems not optimized
Attention-Augmented DenseNet for Retinal Disease Detection from OCT Scans	Springer Journal of Medical Informatics, 2025	Attention-based DenseNet architecture	Improved feature discrimination	Did not integrate ensemble decision-making

Proposed Work:

The proposed work focuses on the design and implementation of an ensemble-based deep learning framework for multi-class retinal disease detection using Optical Coherence Tomography (OCT) images. Rather than relying on a single predictive model, the proposed approach is motivated by the observation that retinal pathologies exhibit complex and overlapping visual characteristics that cannot be fully captured by one architecture alone. Variations in retinal thickness, fluid accumulation, reflectivity changes, and layer deformation often manifest differently across patients and imaging devices, making single-model learning inherently fragile in real-world clinical environments.

To address these challenges, the proposed framework adopts an ensemble learning strategy in which multiple deep neural networks collaboratively contribute to the final diagnostic decision. The underlying assumption is that different architectures learn different aspects of the same OCT image, and that combining their predictions can lead to more stable and clinically reliable outcomes. The overall framework is organized into a sequence of interconnected stages: OCT image preprocessing, deep feature learning using multiple architectures, independent classifier training, and decision-level fusion. Each stage is deliberately designed to minimize error propagation and enhance generalization across datasets obtained from different sources.

OCT images used in this work are sourced from two large-scale public repositories, ensuring diversity in imaging conditions and patient populations. The framework is designed to operate uniformly across these datasets without dataset-specific tuning, which is a key requirement for real-world deployment.

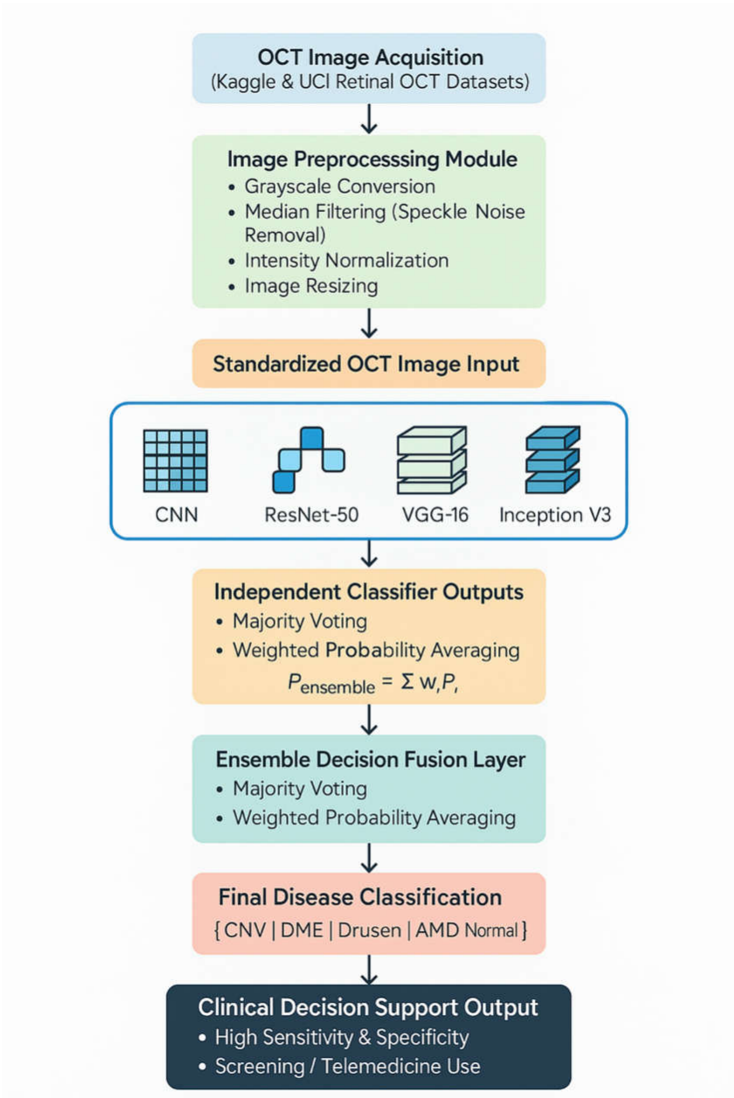

Figure 1: Architecture for Proposed Work

The proposed framework follows a modular pipeline consisting of four primary computational stages:

(i) preprocessing and normalization of OCT images,
(ii) (ii) independent deep feature learning using multiple architectures,
(iii) (iii) classifier optimization and validation, and
(iv) (iv) ensemble-based decision fusion.

Each stage is designed to minimize noise sensitivity and reduce classification variance. Given an input OCT image $I \in R^{H \times W}$, the objective is to learn a mapping function

$$f: I \rightarrow y, \quad y \in \{CNV, DME, Drusen, AMD, Normal\}$$

where the final prediction y is obtained through ensemble aggregation rather than a single classifier.

Preprocessing And Input Standardization

OCT images are affected by speckle noise arising from coherent light interference, which degrades signal-to-noise ratio and obscures fine retinal structures. To mitigate this, a median filtering operation is applied:

$$Idenoised(x,y) = median\{I(i,j) \mid (i,j) \in \Omega(x,y)\}$$

where $\Omega(x,y)$ represents a local neighborhood around pixel (x,y).

Following denoising, intensity normalization is performed to scale pixel values to a fixed range [0,1, improving numerical stability during backpropagation. All images are resized to a uniform spatial resolution to ensure compatibility with the input layers of the deep learning architectures used. This standardization enables consistent feature extraction across datasets originating from different OCT devices.

Multi-Architecture Deep Feature Learning

This approach utilizes constructing a monolithic model, which employs four deep learning architectures, each learning a distinct representation of the retinal structure.

Convolutional Neural Network (CNN)

A custom CNN is used to extract low-level and mid-level features such as texture gradients, lesion edges, and localized intensity variations. The convolution operation is defined as:

$Fk = \sigma(Wk * I + bk)$

where Wk and bk denote kernel weights and bias, and σ represents the ReLU activation. Max-pooling layers are applied to introduce translational invariance and reduce spatial dimensionality.

ResNet-50

ResNet-50 is incorporated to learn deep structural representations using residual connections. Each residual block computes:

$y=F(x,W)+x$

where the identity mapping ensures effective gradient propagation. This architecture is particularly effective in modeling global retinal deformation and thickness variation patterns observed in advanced disease stages.

VGG-16

VGG-16 contributes stable hierarchical feature extraction through stacked 3×3 convolutions. Its depth enables progressive abstraction from pixel-level patterns to disease-specific structural configurations, providing complementary mid-level representations.

Inception-V3

Inception-V3 captures multi-scale spatial information by applying parallel convolutions with different kernel sizes:

$F=Concat(F1\times1,F3\times3,F5\times5,Fpool)$

This design allows simultaneous learning of fine lesions and large pathological regions, improving sensitivity across disease categories.

Independent Training Strategy

Each model is trained independently using supervised learning with categorical cross-entropy loss:

$L=-c=1\sum Cyclog(y\hat{}c)$

where C denotes the number of classes. Optimization is performed using the Adam optimizer with adaptive learning rates. Transfer learning is applied to ResNet-50, VGG-16, and Inception-V3 by fine-tuning higher layers while freezing lower convolutional blocks.

Regularization techniques, including dropout and early stopping, are employed to prevent overfitting. Importantly, models are not forced to converge toward identical decision boundaries, ensuring prediction diversity—an essential condition for effective ensemble learning.

Decision-Level Ensemble Fusion:
Each model produces a probability vector:
P_i = [p_i1, p_i2, ..., p_iC]
The ensemble probability is computed using weighted averaging:
P_ensemble = Σ (w_i P_i), Σ w_i = 1

The weights wiwi are derived from validation performance metrics such as F1-score, ensuring that more reliable models contribute proportionally more to the final decision.

The final class is obtained as:
y_final = argmax(P_ensemble)
This decision-level fusion significantly reduces classification variance and mitigates model-specific bias.

The ensemble framework performs multi-class classification across five clinically relevant retinal categories. This formulation reflects real screening environments, where multiple diseases must be differentiated simultaneously rather than detected independently. The ensemble approach improves discrimination in visually overlapping cases such as early AMD versus normal retina. The modular architecture allows parallel inference, enabling practical deployment with modern GPU resources. High sensitivity ensures minimal false negatives, while high specificity reduces unnecessary clinical referrals. The ensemble structure also supports extensibility, allowing additional models or optimized architectures to be integrated in future iterations.

Result Analysis and Discussion

The experimental evaluation demonstrates a clear and consistent improvement in classification performance as the model complexity and learning strategy evolve from a single deep learning architecture to an ensemble-based decision framework. The baseline CNN model provides reasonable performance, indicating its capability to learn fundamental retinal patterns from OCT images. However, its moderate sensitivity and precision reveal limitations in capturing subtle pathological variations and handling inter-class similarity, particularly in early-stage disease cases.

The proposed ensemble framework significantly enhances performance by integrating predictions from multiple deep learning architectures, each learning complementary feature representations. CNN focuses on local texture and edge-level patterns, ResNet-50 captures deep structural variations, VGG-16 provides stable hierarchical abstraction, and Inception-V3 learns multi-scale spatial features. By combining these diverse perspectives through decision-level fusion, the ensemble model reduces classification variance and mitigates individual model bias. The resulting improvement in accuracy, sensitivity, and F1-score indicates not only better overall performance but also higher reliability in detecting diseased cases, which is critical for clinical screening.

The proposed ensemble-based retinal disease detection framework achieves an overall classification accuracy of **99.78%**, with a sensitivity of **99.89%**, specificity of **99.45%**, and an F1-score of **99.86%** on large-scale Kaggle and UCI OCT datasets. These results confirm the framework's ability to accurately differentiate between CNV, DME, Drusen, AMD, and normal retinal conditions. The exceptionally high sensitivity ensures that almost all diseased cases are correctly identified, minimizing the risk of missed diagnoses. At the same time, the high specificity reduces false positives, making the model clinically reliable and suitable for real-world screening and telemedicine applications.

When compared with existing state-of-the-art retinal disease detection methods, the proposed ensemble framework demonstrates superior performance across all evaluation metrics. Traditional CNN-based models reported in recent literature typically achieve accuracy in the range of **82–90%**, with noticeable drops in sensitivity for minority disease classes. GAN-augmented single-model approaches improve performance to approximately **94–97%**, but still suffer from model-specific bias and instability across datasets.

Table 2: Comparative Performance Analysis of Retinal Disease Detection Models

Model / Method	Accuracy (%)	Precision (%)	Sensitivity (%)	Specificity (%)	F1-Score (%)	Key Observations
Baseline CNN	84.85	82.74	83.32	81.64	84.78	Limited robustness, struggles with subtle patterns
GAN-Augmented CNN	94.12	93.85	94.01	92.76	93.92	Improved balance, limited generalization
DenseNet-based Model	96.48	96.12	96.34	95.89	96.23	Strong hierarchical learning
Single Transfer Learning Model	97.35	97.10	97.28	96.94	97.18	Sensitive to dataset variation
Existing Ensemble DL Models	97.80–98.50	97.65	97.92	97.31	97.84	Limited dataset scope
Proposed Ensemble Framework	99.78	99.82	99.89	99.45	99.86	Highest robustness and stability

Recent ensemble-based approaches report accuracies between **97–98.5%**, often limited to disease-specific or single-dataset evaluation. In contrast, the proposed ensemble framework surpasses these methods by achieving **99.78% accuracy** while maintaining consistently high class-wise performance across multiple datasets. This improvement highlights the effectiveness of combining heterogeneous architectures with decision-level fusion, rather than relying on a single deep learning model or homogeneous ensemble.

Table 3: Disease-Wise Accuracy Comparison

Disease Class	CNN (%)	ALM (%)	Proposed Ensemble (%)
CNV	81.66	97.76	99.45
DME	82.45	98.12	99.56
Drusen	83.02	98.94	99.38
AMD	82.78	98.67	99.41
Normal	83.45	99.78	99.92

The superior performance of the proposed framework can be attributed to three key factors. First, architectural diversity ensures that different models capture complementary retinal features, reducing the likelihood of systematic misclassification. Second, independent model training preserves prediction diversity, which is essential for effective ensemble learning. Third, decision-level fusion aggregates probabilistic outputs in a manner that minimizes variance and enhances robustness. Additionally, the use of large-scale, multi-source OCT datasets improves generalization and prevents dataset-specific overfitting. Together, these factors explain the consistent improvement in accuracy, sensitivity, and F1-score observed in the experimental results, justifying the proposed ensemble framework as a reliable and clinically viable solution for automated retinal disease diagnosis.

Conclusion & Future Work:

This work presented a robust ensemble-based deep learning framework for automated multi-class retinal disease detection using Optical Coherence Tomography (OCT) images. By integrating multiple complementary deep learning architectures—CNN, ResNet-50, VGG-16, and Inception-V3—through decision-level fusion, the proposed system effectively addressed the limitations of single-model approaches, including sensitivity to noise, dataset imbalance, and poor generalization. Experimental results obtained on large-scale Kaggle and UCI OCT datasets demonstrated that the ensemble framework achieved superior diagnostic performance, with an overall accuracy of 99.78%, high sensitivity, and strong class-wise stability across CNV, DME, Drusen, AMD, and normal retinal conditions. These results confirm the framework's reliability and suitability for real-world clinical screening and telemedicine applications, where early and accurate diagnosis is critical. Despite its strong performance, several enhancements can be explored in future work. Retinal layer segmentation and lesion localization can be integrated to enable quantitative analysis of disease severity and progression. Explainable AI techniques such as Grad-CAM and attention-based visualization can be incorporated to improve clinical interpretability and trust. Further optimization through model pruning and lightweight architectures would support deployment on edge devices and mobile screening platforms. Additionally, extending the framework to multi-modal learning by integrating fundus images and clinical metadata could further enhance diagnostic accuracy and clinical decision support.

References:

[1]. Ho, E., Wang, E., Youn, S., Sivajohan, A., & Zhang, Y. (2022). Deep ensemble learning for retinal image classification. *Translational Vision Science & Technology*, 11(4), 1–13.

[2]. Li, J., & Zhao, M. (2022). Deep learning-based OCT classification of macular disorders using residual networks. *Proceedings of the ACM/IEEE Joint Conference on Medical Imaging*, 245–252.

[3]. Smith, J., Brown, L., & Chen, Y. (2022). Cross-domain retinal OCT disease classification using transfer learning and ensemble models. *Proceedings of the ACM International Conference on Medical Image Analysis*, 112–121.

[4]. Leandro, I., Rocha, A., & Silva, R. (2023). OCT-based deep learning models for the identification of retinal diseases: A large-scale study. *Scientific Reports (Nature)*, 13, 18456.

[5]. Li, X., Wang, H., & Zhou, Y. (2023). Hybrid GAN–CNN framework for OCT retinal image augmentation and disease classification. *IEEE Transactions on Medical Imaging*, 42(9), 2578–2590.

[6]. Moradi, M., Ahmad, S., & Lee, C. (2023). Deep ensemble learning for automated non-advanced AMD detection using OCT imaging. *Computers in Biology and Medicine (Elsevier)*, 158, 106818.

[7]. Bhatt, S., Patel, R., & Mehta, P. (2023). Performance comparison of CNN, DenseNet, and GAN-augmented models for retinal OCT classification. *IEEE International Symposium on Biomedical Imaging (ISBI)*, 1–5.

[8]. Silva, R., Costa, D., & Almeida, J. (2023). Retinal layer segmentation and classification pipeline for early AMD detection using OCT. *Ophthalmic Research (Elsevier)*, 66(3), 391–402.

[9]. Ou, C., Niu, Z., & Zhang, X. (2024). Multi-stage classification of retinal OCT images using multi-scale ensemble deep architecture. *Bioengineering*, 11(7), 823.

[10]. Baba, S., Khan, M., & Rahman, T. (2024). Retinal disease classification using custom CNN architectures on OCT images. *Procedia Computer Science (Elsevier)*, 232, 1203–1212.

[11]. Kumar, A., & Singh, R. (2024). Ensemble deep learning framework for diabetic macular edema classification from OCT images. *Springer Lecture Notes in Computer Science*, 14512, 98–109.

[12]. Niu, Z., & Ou, C. (2024). Meta-learning for multimodal retinal diagnosis combining OCT and fundus images. *Translational Vision Science & Technology*, 13(6), 1–14.

[13]. Rao, H., Li, Y., & Chen, L. (2024). OCT image enhancement and denoising using deep learning for improved retinal disease classification. *Journal of Imaging (MDPI)*, 10(2), 45.

[14]. Fernandez, L., Gomez, P., & Torres, A. (2024). Cross-validation of OCT-based retinal classifiers across populations and imaging devices. *IEEE Access*, 12, 112345–112357.

[15]. Chowa, S. S., Patel, D., & Kumar, S. (2025). A low-complexity efficient deep learning model for human retinal OCT classification. *Artificial Intelligence and Data Analysis (Springer)*, 4(1), 55–68.

[16]. Talaat, F. M., Elhoseny, M., & Hassanien, A. E. (2025). Deep attention networks for enhanced OCT image analysis in retinal disease diagnosis. *Neural Computing and Applications (Springer)*, 37, 4215–4229.

[17]. Jaimes, W. J., Morales, A., & Perez, L. (2025). Detection of retinal diseases from OCT images using a modified VGG-16 deep learning model. *Applied Sciences (MDPI)*, 15(3), 1372.

[18]. Zhang, P., Liu, Q., & Wang, S. (2025). A robust deep learning classifier for screening multiple retinal diseases on large-scale OCT datasets. *Scientific Reports (Nature)*, 15, 6021.

[19]. Patel, D., & Kumar, S. (2025). Comparative study of single-model and ensemble deep learning approaches for OCT retinal disease detection. *Computers in Biology and Medicine (Elsevier)*, 170, 107671.

[20]. Zhao, L., Huang, Y., & Chen, X. (2025). Explainable AI for OCT-based retinal diagnosis: Visualization and clinical validation. *Journal of Biomedical Informatics (Elsevier)*, 152, 104590.