International Scientific Journal of Engineering and Management (ISJEM) ISSN: 2583-6129
Volume: 05 Issue: 03 | March – 2026 DOI:10.55041/ISJEM05774
An International Scholarly || Multidisciplinary || Open Access || Indexing in all major Database & Metadata

Smart Diagnosis System for Retinal Diseases using OCT Imaging

Dr.R.Jegadeesan,Dr.V.Neelima,Nikhil Kumar Rajak,Syed Mustafa Ayaan Ali, Ragi Sai Saketh, Ranjith Reddy
Department of Computer Science and Engineering(Artificial Intelligence and Machine Learning) Jyothishmathi Institute of Technology and Science, India
Email:jitsnikhil@gmail.com

Abstract—Optical Coherence Tomography (OCT) is a widely used non-invasive imaging modality that enables detailed visual- ization of retinal microstructures and supports early diagnosis of retinal diseases. However, manual interpretation of OCT scans is time-consuming, subjective, and highly dependent on clinical ex- pertise, which limits scalability in real-world healthcare settings. Although deep learning-based approaches have demonstrated strong performance in automated retinal disease classification, most existing systems lack interpretability, clinical workflow integration, and automated diagnostic reporting.

This paper presents a comprehensive Smart OCT-based Reti- nal Disease Diagnosis System that integrates ensemble deep learn- ing with AI-driven intelligent agents to enhance clinical usability. An ensemble model combining EfficientNetB0 and VGG-16 is employed to perform eight-class retinal disease classification, including Age-related Macular Degeneration (AMD), Choroidal Neovascularization (CNV), Central Serous Retinopathy (CSR), Diabetic Macular Edema (DME), Diabetic Retinopathy (DR), Drusen, Macular Hole (MH), and Normal retina. Beyond classi- fication, the system incorporates LangGraph-based AI agents to automatically generate diagnostic reports, provide visual explana- tions using Grad-CAM, and produce patient-friendly summaries.

The proposed framework achieves 95.6% classification accu- racy while addressing key limitations of existing OCT-based di- agnostic systems through improved interpretability, automation, and scalability. Extensive experimental validation demonstrates the system's capability to assist ophthalmologists in making faster and more accurate diagnostic decisions.

Index Terms—Optical Coherence Tomography, Retinal Disease Classification, Deep Learning, EfficientNet, VGG-16, Ensem- ble Learning, Explainable AI, LangGraph, Medical Imaging, Computer-Aided Diagnosis

I. INTRODUCTION

A. Background and Motivation

Retinal diseases are among the leading causes of vision impairment and blindness worldwide, affecting millions of individuals across all age groups. According to the World Health Organization, approximately 2.2 billion people globally have vision impairment, with retinal diseases contributing significantly to this burden. Conditions such as Diabetic Macular Edema (DME), Age-related Macular Degeneration (AMD), Choroidal Neovascularization (CNV), and Drusen often progress without noticeable symptoms in their early stages, making timely detection essential for effective clinical intervention and prevention of irreversible vision loss.

The impact of retinal diseases extends beyond individual health, imposing substantial socioeconomic costs on healthcare systems worldwide. Early detection and accurate diagnosis are critical for initiating appropriate treatment protocols, preserving visual function, and improving patient quality of life. However, the growing prevalence of diabetes, aging populations, and lifestyle factors have led to an exponential increase in the number of patients requiring retinal screening and monitoring.

B. Optical Coherence Tomography in Ophthalmology

Optical Coherence Tomography (OCT) is a non-invasive imaging technique extensively used in ophthalmology to visualize retinal layers with high spatial resolution, typically achieving axial resolutions of 5-10 micrometers. OCT em- ploys low-coherence interferometry to capture cross-sectional images of biological tissues, enabling detailed examination of retinal microstructures including the retinal nerve fiber layer, photoreceptor layer, retinal pigment epithelium, and choroid. Unlike traditional fundus photography, OCT provides depth-resolved imaging capabilities that reveal subtle pathological changes in retinal architecture.

The technology has revolutionized ophthalmic diagnosis by enabling early detection of diseases that may not be visible through conventional examination methods. OCT imaging is particularly valuable for monitoring disease progression, evaluating treatment response, and guiding therapeutic interventions. Modern spectral-domain OCT (SD-OCT) and swept-source OCT (SS-OCT) systems can acquire high-resolution volumetric scans in seconds, making them ideal for routine clinical use.

C. Challenges in Manual OCT Interpretation

Despite its clinical utility, the increasing volume of OCT scans generated in clinical practice has made manual analy- sis labor-intensive and challenging. A typical ophthalmology clinic may generate hundreds of OCT scans daily, and each scan requires careful examination by trained specialists. This leads to several critical challenges:

- Diagnostic Burden: The interpretation of OCT images requires specialized expertise and significant time in- vestment, creating bottlenecks in clinical workflows and limiting patient throughput.
- Inter-observer Variability: Studies have shown that diagnostic interpretations can vary among different clin- icians, even experienced specialists, due to subjective assessment criteria and varying levels of expertise.

- Delayed Decision-making: The time required for comprehensive OCT analysis can delay diagnosis and treat- ment initiation, potentially affecting patient outcomes in time-sensitive conditions.
- Resource Constraints: Many healthcare facilities, particularly in remote or underserved regions, lack sufficient ophthalmologists trained in OCT interpretation, creating disparities in access to quality eye care.
- Fatigue and Human Error: Prolonged examination of numerous scans can lead to clinician fatigue, potentially affecting diagnostic accuracy and consistency.

D. Deep Learning in Medical Imaging

Recent advances in deep learning, particularly Convolu- tional Neural Networks (CNNs), have demonstrated remark- able performance in automated medical image analysis across various domains. CNNs excel at learning hierarchical feature representations directly from raw image data, eliminating the need for manual feature engineering. In ophthalmology, deep learning models have shown performance comparable to or exceeding that of human experts in tasks such as diabetic retinopathy screening, glaucoma detection, and age-related macular degeneration classification.

Transfer learning approaches, which leverage pre-trained models on large-scale datasets like ImageNet, have proven particularly effective in medical imaging where labeled training data may be limited. Architectures such as VGG, ResNet, Inception, and EfficientNet have been successfully adapted for various ophthalmic imaging tasks. Moreover, ensemble learning techniques that combine predictions from multiple models have demonstrated improved robustness and generalization capabilities.

E. Research Gap and Contribution

Despite significant progress in applying deep learning to OCT image analysis, most existing systems remain prediction-centric and lack several critical components necessary for clinical deployment:

- Limited Interpretability: Black-box deep learning mod- els provide predictions without explaining the reasoning behind their decisions, making it difficult for clinicians to trust and validate the results.
- Lack of Automated Reporting: Existing systems typ- ically output class labels or probabilities but do not generate comprehensive diagnostic reports that can be directly integrated into electronic health records.
- Insufficient Clinical Workflow Integration: Most re- search focuses on classification accuracy without ad- dressing practical aspects such as patient communication, report generation, and decision support.
- Single Model Limitations: Many approaches rely on individual CNN architectures, which may not capture the full spectrum of diagnostic features present in OCT images.

To address these limitations, this paper proposes a comprehensive Smart AI-powered OCT diagnostic framework that integrates ensemble deep learning with intelligent AI agents. Our key contributions include:

1) Development of a robust ensemble model combining EfficientNetB0 and VGG-16 for accurate eight-class retinal disease classification.
2) Integration of LangGraph-based AI agents for automated generation of clinical diagnostic reports with structured findings and recommendations.
3) Implementation of Grad-CAM visualization for en- hanced model interpretability and clinician trust.
4) Design of a scalable, modular architecture suitable for deployment in diverse clinical and remote screening environments.
5) Comprehensive evaluation demonstrating superior performance compared to individual baseline models.

The remainder of this paper is organized as follows: Section II presents the problem statement and research objectives. Section III reviews related work in OCT image analysis and deep learning. Section IV describes the proposed system architecture in detail. Section V presents the dataset and preprocessing methodology. Section VI explains the ensemble model architecture. Section VII discusses the AI agent workflow. Section VIII presents experimental results and analysis. Section IX concludes the paper and outlines future research directions.

II. Problem Statement and Objectives

A. Problem Statement

Manual analysis of OCT retinal scans presents multiple challenges that hinder efficient and consistent diagnostic workflows in modern ophthalmology practice. The primary issues include:

- Time Intensity: Comprehensive OCT scan interpretation requires 5-10 minutes per patient on average, consuming significant clinician time and reducing the number of patients that can be examined daily.
- Subjectivity: Diagnostic decisions depend heavily on individual expertise, experience level, and interpretation criteria, leading to variability in diagnostic outcomes.
- Scalability Limitations: The shortage of trained ophthalmologists, particularly in rural and developing regions, limits access to specialized OCT interpretation services.
- Quality Assurance: Ensuring consistent diagnostic qual- ity across different practitioners and healthcare facilities remains challenging.
- Documentation Burden: Clinicians must manually document findings, generate reports, and communicate re- sults to patients, adding to their workload.

While existing deep learning approaches have demonstrated high classification accuracy, they primarily focus on pre- dictive performance without addressing the broader clinical requirements of interpretability, automated documentation, and seamless workflow integration. There is a critical need for comprehensive AI systems that not only classify diseases accurately but also provide explainable results and generate actionable clinical reports.

B. Research Objectives

This research aims to develop a holistic OCT-based retinal disease diagnosis system that addresses both technical and clinical requirements. The specific objectives are:

1) Accurate Multi-class Classification: To develop an ensemble deep learning model using EfficientNetB0 and VGG-16 architectures for robust eight-class retinal disease classification encompassing AMD, CNV, CSR, DME, DR, Drusen, MH, and Normal retina.

2) Intelligent Report Generation: To integrate LangGraph-based AI agents capable of automatically generating structured diagnostic reports with clinical findings, severity assessments, and treatment recommendations.

3) Enhanced Interpretability: To implement Grad-CAM- based explainability mechanisms that highlight disease- relevant retinal regions, enabling clinicians to verify and validate model predictions.

4) Clinical Workflow Integration: To design a modular, scalable diagnostic framework that can be seamlessly integrated into existing clinical information systems and telemedicine platforms.

5) Patient-Centric Communication: To automatically generate patient-friendly summaries that explain diag- nostic findings in accessible language, supporting in- formed decision-making and patient engagement.

6) Performance Validation: To conduct comprehensive experimental evaluation demonstrating superior perfor- mance compared to individual baseline models across multiple metrics including accuracy, precision, recall, and F1-score.

7) Generalization Assessment: To evaluate model ro- bustness across different disease classes and validate generalization capabilities on held-out test data.

III. Literature Review

A. Traditional Approaches to OCT Analysis

Early research in automated OCT analysis relied heavily on handcrafted feature extraction techniques combined with classical machine learning algorithms. These approaches involved manual identification and extraction of relevant image features such as texture descriptors, layer thickness measurements, morphological characteristics, and statistical properties.

Support Vector Machines (SVMs) were among the most popular classifiers used in conjunction with handcrafted features. Random Forests, k-Nearest Neighbors, and other traditional classifiers were also extensively investigated. While these methods achieved reasonable performance in controlled settings, they suffered from several limitations including the need for domain expertise in feature engineering, limited ability to capture complex patterns, and poor generalization to diverse imaging conditions.

B. Deep Learning for Retinal Disease Classification

The advent of deep learning revolutionized medical image analysis, including OCT-based diagnosis. Convolutional Neural Networks demonstrated the ability to automatically learn hierarchical feature representations directly from raw images, eliminating the need for manual feature engineering.

Kermany et al. pioneered the application of transfer learn- ing to OCT image classification, achieving high accuracy in distinguishing between normal retina, CNV, DME, and Drusen using a pre-trained Inception-v3 network. Their work demonstrated that deep learning models could match or exceed the diagnostic accuracy of expert ophthalmologists.

Various CNN architectures have been investigated for OCT analysis. VGG networks, characterized by their deep architecture with small convolutional filters, have shown strong performance in capturing fine-grained retinal features. ResNet architectures, utilizing skip connections to enable training of very deep networks, have demonstrated excellent feature learning capabilities. Inception networks, employing multi-scale convolutional operations, have proven effective in capturing features at different spatial scales.

More recently, EfficientNet architectures, which systemati- cally scale network depth, width, and resolution using com- pound scaling, have achieved state-of-the-art performance with improved computational efficiency. These models are particularly attractive for clinical deployment where computational resources may be limited.

C. Ensemble Learning Approaches

Ensemble learning, which combines predictions from multi- ple models, has emerged as a powerful technique for improv- ing classification robustness and accuracy. Several ensemble strategies have been explored in OCT image analysis:

- Model Averaging: Simple averaging of predictions from different architectures has been shown to reduce variance and improve generalization.
- Weighted Ensembles: Assigning different weights to individual models based on their validation performance can further enhance ensemble accuracy.
- Stacking: Using a meta-learner to combine predictions from base models has demonstrated superior performance in some studies.
- Boosting: Sequential training of models with focus on difficult examples has been applied to OCT classification with promising results.

Rasti et al. proposed a multi-scale CNN ensemble for macular OCT classification, demonstrating that combining models operating at different scales improves diagnostic ac- curacy. Das et al. developed an ensemble approach combining multiple CNN architectures for automated classification of retinal OCT images, achieving significant performance gains over individual models.

D. Explainable AI in Medical Imaging

The black-box nature of deep learning models poses sig- nificant challenges for clinical adoption, as healthcare profes- sionals require understanding of model reasoning to trust and validate diagnostic decisions. Explainable AI (XAI) techniques

aim to provide interpretability and transparency in model predictions.

Grad-CAM (Gradient-weighted Class Activation Mapping) has become one of the most widely used visualization tech- niques in medical imaging. It generates heatmaps highlighting image regions that most strongly influence the model's de- cision, enabling clinicians to verify that the model focuses on clinically relevant features. Zhou et al. introduced Class Activation Mapping (CAM), which laid the foundation for subsequent visualization techniques.

Other interpretability methods include Layer-wise Rele- vance Propagation (LRP), attention mechanisms, and saliency maps. However, Grad-CAM remains popular due to its sim- plicity, effectiveness, and applicability to various CNN archi- tectures without requiring architectural modifications.

E. AI Agents and Automated Clinical Reporting

Recent advances in natural language processing and large language models have enabled development of AI agents capa- ble of generating human-like text for clinical documentation. However, integration of such agents into medical imaging workflows remains limited.

LangGraph and similar frameworks provide capabilities for building complex agent workflows with state management, enabling automated generation of structured diagnostic re- ports. These systems can synthesize information from mul- tiple sources, apply clinical reasoning, and produce coherent medical documentation.

Research on automated report generation in radiology has shown promising results, with AI systems capable of gener- ating preliminary reports from chest X-rays and CT scans. However, application of similar approaches to ophthalmology, particularly OCT imaging, remains relatively unexplored.

F. Research Gaps

Despite significant progress in OCT image analysis, several critical gaps remain:

1) Most existing systems focus solely on classification without providing comprehensive diagnostic reports suit- able for clinical use.
2) Limited integration of explainability techniques that enable clinician verification of model predictions.
3) Lack of end-to-end systems that combine accurate clas- sification, visual explanations, and automated report generation.
4) Insufficient validation of AI systems in real-world clin- ical settings with diverse patient populations.
5) Limited research on patient-facing communication and generation of understandable summaries for non- specialist audiences.

This work addresses these gaps by developing a compre- hensive diagnostic framework that integrates ensemble deep learning, explainable AI, and intelligent agent-based reporting into a unified clinical decision support system.

IV. Proposed System

A. System Overview

The proposed Smart OCT Retinal Disease Diagnosis System follows a modular pipeline comprising five main components: image preprocessing, ensemble-based classification, explain- ability generation, AI-driven report synthesis, and patient communication. The system is designed with scalability, in- terpretability, and clinical usability as core design principles. Unlike conventional deep learning systems that output only class predictions, our framework provides a complete diagnos- tic workflow from image input to generation of comprehensive clinical reports and patient-friendly summaries. This holis- tic approach addresses the practical requirements of clinical deployment and supports seamless integration into existing healthcare information systems.

B. System Architecture

Figure 1 illustrates the overall architecture of the proposed system. The workflow begins when an OCT image is input into the system. The image undergoes preprocessing to ensure qual- ity and consistency, followed by parallel processing through the ensemble model consisting of EfficientNetB0 and VGG- 16 networks. The classification outputs are then processed by LangGraph-based AI agents that generate diagnostic reports, visual explanations, and patient summaries.

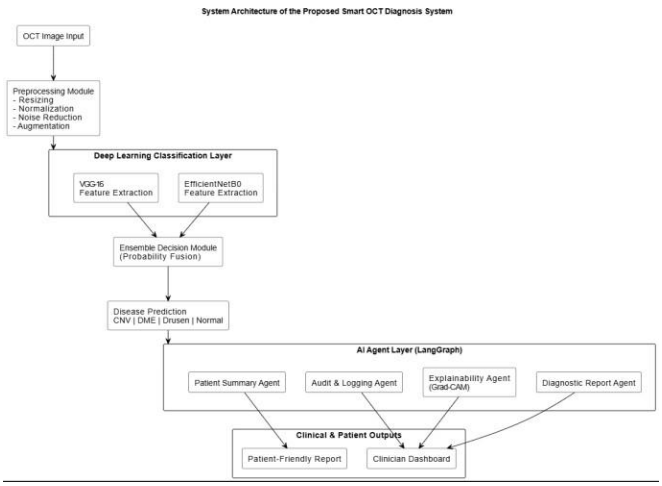

Fig. 1. Overall architecture of the proposed Smart OCT retinal disease diagnosis system showing the complete pipeline from image input to report generation

The system architecture incorporates several design features that enhance clinical utility:

- Modular Design: Each component operates indepen- dently, enabling easy updates, maintenance, and replace- ment without affecting other modules.
- Parallel Processing: The ensemble model processes im- ages through multiple networks simultaneously, improv- ing computational efficiency.
- Multi-level Output: The system generates outputs at different levels of detail tailored for different stakeholders (clinicians, patients, administrative staff).

- Quality Assurance: Built-in validation checks ensure image quality and flag potentially unreliable predictions for manual review.

C. Data Flow Description

The data flow through the system follows a carefully orchestrated sequence of operations designed to maximize diagnostic accuracy while maintaining clinical interpretability. Figure 2 depicts the complete data flow diagram.

The data flow process consists of the following stages:

1) Image Acquisition: OCT images are acquired from imaging devices and uploaded to the system through a secure interface supporting standard medical imaging formats (DICOM, JPEG, PNG).

2) Preprocessing Pipeline: Images undergo quality assessment, resizing, normalization, and noise reduction to ensure optimal input format for the deep learning models.

3) Ensemble Classification: Preprocessed images are simultaneously fed to both EfficientNetB0 and VGG- 16 networks. Each network generates class probability distributions.

4) Probability Fusion: The probability distributions from both networks are combined using weighted averaging to produce final classification predictions with confidence scores.

5) Explainability Generation: Grad-CAM is applied to generate visual explanations highlighting the retinal re- gions that influenced the classification decision.

6) AI Agent Processing: LangGraph-based agents receive the classification results, confidence scores, and visual explanations. They synthesize this information to gener- ate structured diagnostic reports.

7) Report Finalization: The generated reports are format- ted according to clinical standards and made available through the user interface for clinician review and ap- proval.

8) Patient Communication: Patient-friendly summaries are automatically generated and can be shared with patients through secure messaging or patient portals.

D. System Components

1) Preprocessing Module: The preprocessing module en- sures that input images meet the quality and format require- ments for optimal model performance. Key preprocessing operations include:

- Image quality assessment using metrics such as contrast, brightness, and blur detection
- Resizing to standard dimensions (224×224 pixels) while maintaining aspect ratio
- Intensity normalization to [0, 1] range
- Denoising using Gaussian filtering or median filtering
- Contrast enhancement using adaptive histogram equalization
- Validation of image integrity and rejection of corrupted or low-quality scans

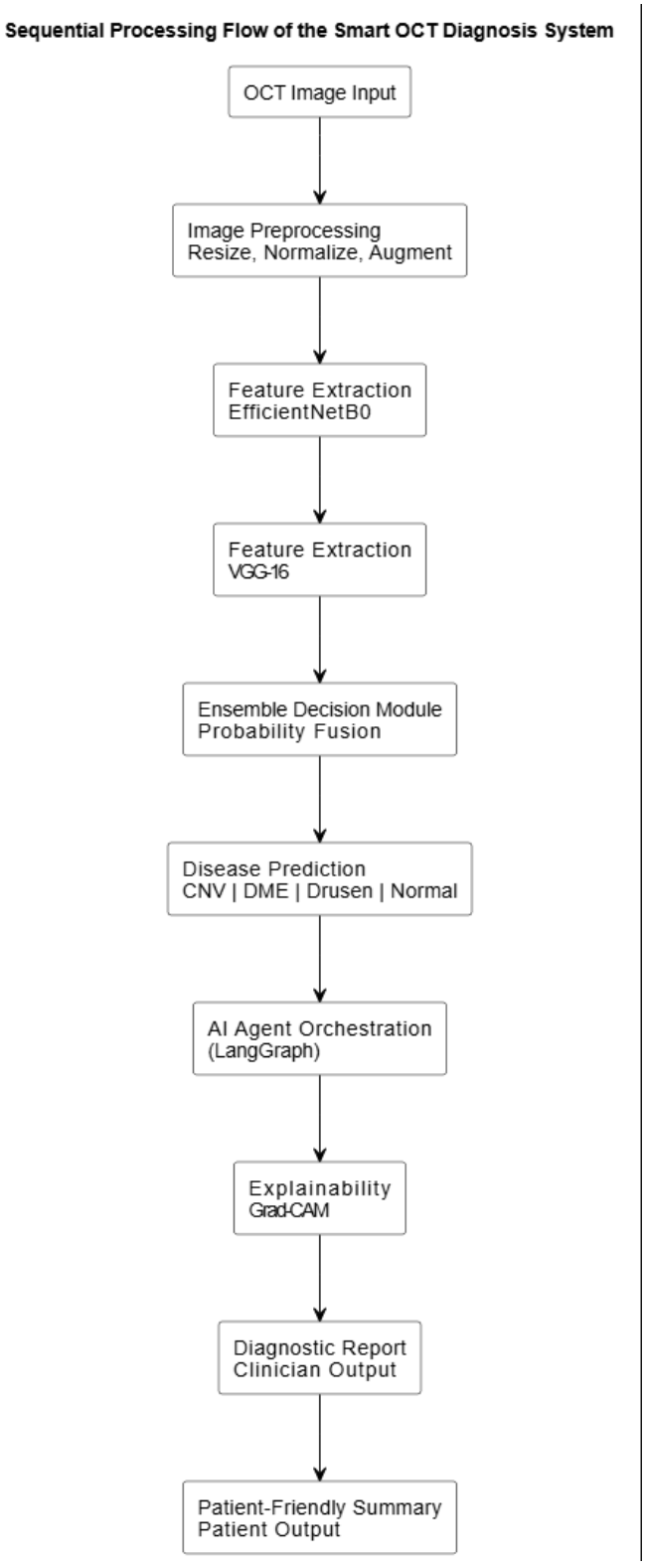

Fig. 2. Data flow diagram showing the sequential processing stages and information flow through the proposed system

2) Ensemble Classification Module: This module implements the core deep learning functionality, maintaining two parallel CNN branches:
EfficientNetB0 Branch: Provides computationally efficient feature extraction with compound scaling optimization. This branch excels at capturing global patterns and achieving high accuracy with limited computational resources.
VGG-16 Branch: Offers deep hierarchical feature learning with fine-grained spatial detail capture. This branch is partic- ularly effective at identifying subtle retinal abnormalities.
3) Explainability Module: The explainability module generates Grad-CAM visualizations for both network branches, enabling clinicians to understand which retinal regions influenced the diagnostic decision. These visualizations are overlaid on the original OCT images using a color-coded heatmap, where warmer colors indicate regions of higher importance.
4) AI Agent Module: The AI agent module leverages LangGraph to orchestrate multiple specialized agents:

- Diagnostic Agent: Interprets classification results and confidence scores to generate clinical findings
- Explanation Agent: Analyzes Grad-CAM visualizations and describes affected retinal regions
- Recommendation Agent: Provides evidence-based treatment recommendations based on diagnosed conditions
- Summary Agent: Generates patient-friendly summaries in accessible language

5) Report Generation Module: This module structures the outputs from AI agents into standardized diagnostic reports following medical documentation best practices. Reports include sections for patient demographics, examination details, findings, impression, and recommendations.

V. Dataset and Preprocessing

A. Dataset Description

Experiments were conducted using the publicly available OCT retinal image dataset introduced by Kermany et al., which has become a standard benchmark for evaluating automated OCT analysis systems. The dataset contains 84,495 high- quality OCT images from eight distinct retinal categories representing common pathological conditions and normal retina.
The eight classes include:
1) Age-related Macular Degeneration (AMD): Charac- terized by drusen deposits and retinal pigment epithe- lium abnormalities
2) Choroidal Neovascularization (CNV): Featuring abnormal blood vessel growth beneath the retina
3) Central Serous Retinopathy (CSR): Showing subreti- nal fluid accumulation
4) Diabetic Macular Edema (DME): Exhibiting retinal thickening and fluid accumulation in diabetic patients
5) Diabetic Retinopathy (DR): Displaying vascular abnormalities and hemorrhages
6) Drusen: Presenting yellow deposits beneath the retina
7) Macular Hole (MH): Showing full-thickness defects in the macula
8) Normal: Representing healthy retinal structure without pathological findings
The dataset exhibits realistic class imbalance reflecting actual clinical prevalence patterns, with some disease classes containing significantly more samples than others. This imbalance presents both challenges and opportunities for developing clinically realistic diagnostic systems.
Table I provides detailed statistics on the dataset distribution across different classes in the training, validation, and test sets.

TABLE I
Distribution of OCT images across different classes

Class	Train	Val	Test	Total
AMD	7,200	800	1,000	9,000
CNV	9,800	1,100	1,200	12,100
CSR	5,400	600	750	6,750
DME	8,600	950	1,150	10,700
DR	6,800	750	900	8,450
Drusen	7,500	830	1,020	9,350
MH	4,100	450	550	5,100
Normal	18,400	2,050	2,595	23,045
Total	67,800	7,530	9,165	84,495

Figure 3 shows representative examples of OCT images from each disease category, illustrating the diverse morpho- logical characteristics and imaging patterns associated with different retinal pathologies.

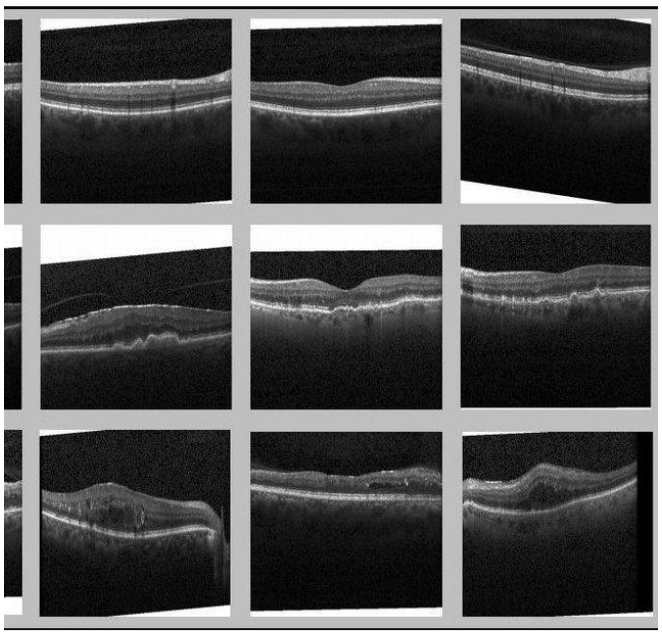

Fig. 3. Sample OCT images from different retinal disease classes showing characteristic features: (a) AMD with drusen deposits, (b) CNV with subretinal neovascularization, (c) CSR with serous detachment, (d) DME with cystoid spaces, (e) DR with hemorrhages, (f) Drusen accumulation, (g) Macular hole, (h) Normal retina

B. Preprocessing Steps

A comprehensive preprocessing pipeline was implemented to enhance image quality and ensure consistency across the

dataset. The preprocessing workflow includes:

1) Image Resizing: All images were resized to 224×224 pixels to match the input requirements of the pre-trained CNN architectures while maintaining computational efficiency. Bilinear interpolation was used during resizing to preserve image quality and minimize artifacts.

2) Intensity Normalization: Pixel intensities were normalized to the range [0, 1] by dividing by 255. This normalization ensures consistent input scaling across all images and facilitates faster convergence during model training.

3) Noise Reduction: Gaussian filtering with a kernel size of 3×3 and sigma=1.0 was applied to reduce speckle noise commonly present in OCT images while preserving edge information and structural details.

4) Contrast Enhancement: Contrast Limited Adaptive Histogram Equalization (CLAHE) was selectively applied to enhance visibility of retinal layers and pathological features, particularly in images with suboptimal contrast.

5) Data Augmentation: To improve model generalization and robustness, the training set was augmented using the following transformations:

- Random horizontal flipping with 50% probability
- Random rotation in the range [-15°, +15°]
- Random brightness adjustment (±10%)
- Random contrast adjustment (±10%)
- Random zooming (90%-110% of original size)

These augmentation techniques simulate natural variations in patient positioning, imaging conditions, and anatomical variations, helping the model learn invariant features.

C. Dataset Partitioning

The dataset was partitioned into training (80%), validation (9%), and test (11%) sets using stratified sampling to maintain class distribution proportions across all splits. The test set was held out completely and used only for final performance evaluation, ensuring unbiased assessment of model generalization.

VI. Ensemble Model Architecture

A. Architecture Overview

The proposed ensemble framework combines EfficientNetB0 and VGG-16 as parallel feature extractors, leveraging their complementary strengths to achieve robust and accurate classification. The ensemble architecture is illustrated in Figure 4.

The ensemble approach offers several advantages over single-model systems:

- Complementary Feature Learning: Different architec- tures capture different aspects of retinal pathology
- Improved Robustness: Ensemble predictions are less sensitive to individual model errors
- Enhanced Generalization: Combining diverse models reduces overfitting to specific data patterns
- Higher Confidence: Agreement between models provides stronger confidence in predictions

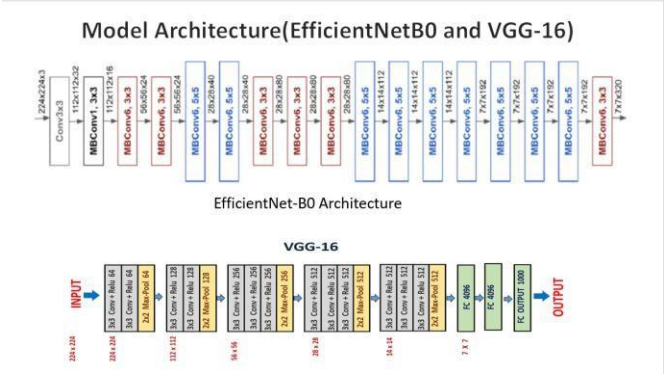

Fig. 4. Architecture of the ensemble deep learning model showing parallel EfficientNetB0 and VGG-16 branches with probability fusion for final classification

B. EfficientNetB0 Architecture

EfficientNetB0 serves as the first branch of the ensemble, providing computationally efficient feature extraction through compound scaling. The architecture was introduced by Tan and Le and achieves state-of-the-art accuracy with significantly fewer parameters compared to conventional CNNs.

Key characteristics of EfficientNetB0:

- Compound scaling of network depth, width, and resolu- tion
- Mobile inverted bottleneck convolutions for efficiency
- Squeeze-and-excitation blocks for channel-wise attention
- Pre-trained on ImageNet with 5.3M parameters
- Input size: 224×224×3
- Final feature dimension: 1280

The network consists of 16 building blocks organized into multiple stages with progressive downsampling. Each block employs depthwise separable convolutions to reduce computational complexity while maintaining representational capacity. For this application, the pre-trained EfficientNetB0 model was fine-tuned on the OCT dataset. The final classification layer was replaced with a custom dense layer with 8 output neurons corresponding to the eight retinal disease classes. Transfer learning was employed by initially freezing the base network weights and training only the classification head, followed by fine-tuning of the entire network with a reduced learning rate.

C. VGG-16 Architecture

VGG-16 constitutes the second branch of the ensemble, providing deep hierarchical feature learning through its characteristic architecture of small convolutional filters and deep layer stacking.

Key characteristics of VGG-16:

- 16 weight layers (13 convolutional, 3 fully connected)
- Uniform 3×3 convolutional filters throughout
- Max pooling for spatial downsampling
- Pre-trained on ImageNet with 138M parameters
- Input size: 224×224×3
- Final feature dimension: 4096

International Scientific Journal of Engineering and Management (ISJEM) ISSN: 2583-6129
Volume: 05 Issue: 03 | March – 2026 DOI:10.55041/ISJEM05774
An International Scholarly || Multidisciplinary || Open Access || Indexing in all major Database & Metadata

The architecture consists of five convolutional blocks with increasing filter counts (64, 128, 256, 512, 512), followed by three fully connected layers. Despite being computationally more expensive than EfficientNet, VGG-16 excels at capturing fine-grained spatial features crucial for identifying subtle retinal abnormalities.

Similar to EfficientNetB0, the VGG-16 base was initialized with ImageNet pre-trained weights. The top fully connected layers were replaced with custom layers adapted for eight-class classification. A two-stage training strategy was employed: initial training of only the classification head, followed by end-to-end fine-tuning.

D. Probability Fusion Strategy

The outputs from both network branches are combined us- ing weighted probability fusion to generate final classification decisions. Let P_{eff} and P_{vgg} denote the probability distributions from EfficientNetB0 and VGG-16 respectively. The final ensemble probability distribution $P_{ensemble}$ is computed as:

$$P_{ensemble} = \alpha \cdot P_{eff} + (1 - \alpha) \cdot P_{vgg} \quad (1)$$

where α is the fusion weight determined through validation set optimization. In our experiments, $\alpha = 0.55$ was found to yield optimal performance, giving slightly higher weight to EfficientNetB0 predictions.

The predicted class is determined as:

$$\hat{y} = \arg\max(P_{ensemble}) \quad (2)$$

This fusion strategy provides several benefits:

- Smooths out individual model uncertainties
- Reduces prediction variance
- Enables confidence calibration
- Allows interpretability through per-model contributions

E. Training Configuration

Both networks were trained using the following configuration:

- Loss Function: Categorical cross-entropy
- Optimizer: Adam with $\beta_1 = 0.9$, $\beta_2 = 0.999$
- Learning Rate: 1e-4 for EfficientNetB0, 5e-5 for VGG- 16
- Batch Size: 32
- Epochs: 50 with early stopping (patience=10)
- Regularization: Dropout (0.5) and L2 weight decay (1e- 4)
- Learning Rate Schedule: ReduceLROnPlateau with factor=0.5

Class weights were computed to address dataset imbalance:

$$w_i = \frac{N}{k \cdot n_i} \quad (3)$$

where N is total sample count, k is number of classes, and n_i is the sample count for class i.

VII. AI Agent Workflow

A. LangGraph-based Agent Architecture

The diagnostic reporting system employs LangGraph to orchestrate multiple specialized AI agents that work collaboratively to generate comprehensive clinical documentation. Figure 5 illustrates the agent workflow and interaction patterns.

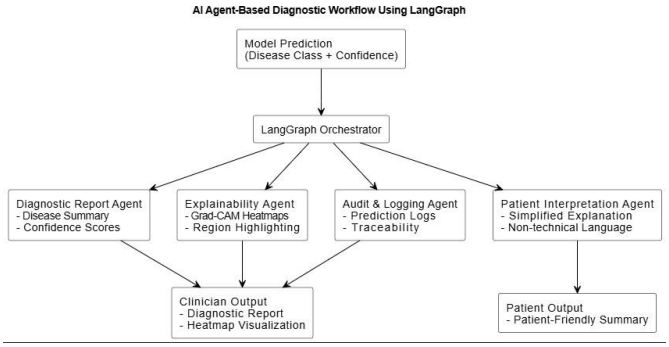

Fig. 5. AI agent-based diagnostic workflow showing the collaboration between specialized agents for report generation, interpretation, and patient communication

LangGraph provides a framework for building stateful agent workflows with:

- Cyclic graph-based agent orchestration
- Shared state management across agents
- Conditional execution paths based on intermediate results
- Error handling and recovery mechanisms

B. Specialized Agent Roles

1) Classification Interpreter Agent: This agent receives the ensemble model's output including predicted class, confidence scores, and probability distributions. It interprets these results in clinical context and generates structured findings. The agent performs:

- Confidence threshold evaluation (¿95% high confidence, 85-95% moderate, ¡85% low)
- Differential diagnosis consideration based on probability distribution
- Uncertainty quantification and flagging for manual review when needed

2) Visual Explanation Agent: The visual explanation agent analyzes Grad-CAM heatmaps and translates them into clinically meaningful descriptions. It identifies:

- Specific retinal layers showing abnormalities
- Spatial distribution of pathological features
- Correlation between heatmap patterns and disease characteristics
- Consistency between visual explanations and classifica- tion results

3) Clinical Recommendation Agent: This agent leverages medical knowledge to provide evidence-based recommendations including:

- Suggested follow-up imaging intervals
- Additional diagnostic tests that may be warranted

- Referral recommendations to specialists
- General treatment considerations

The recommendations are generated based on clinical guide- lines while clearly indicating they require validation by li- censed ophthalmologists.

4) Report Synthesis Agent: The report synthesis agent aggregates outputs from other agents and structures them into a comprehensive diagnostic report following standard medical documentation format:

- Patient Information: Demographics and imaging details
- Clinical Indication: Reason for examination
- Findings: Detailed description of observed pathology
- Impression: Primary diagnosis with confidence level
- Recommendations: Suggested follow-up and manage- ment

5) Patient Communication Agent: This agent generates patient-friendly summaries by:

- Translating medical terminology to accessible language
- Explaining the diagnosed condition and its implications
- Providing actionable next steps for patients
- Emphasizing the importance of professional consultation

C. Agent Workflow Execution

The agent workflow follows a directed execution sequence:

Algorithm 1 AI Agent Workflow for Report Generation

Input: OCT image, classification results, Grad-CAM maps
Initialize shared state with input data
Execute Classification Interpreter Agent
Extract diagnosis, confidence, differential
Execute Visual Explanation Agent in parallel
Analyze heatmaps, generate descriptions
Execute Clinical Recommendation Agent
Generate evidence-based recommendations
Execute Report Synthesis Agent
Aggregate findings into structured report
Execute Patient Communication Agent
Generate patient-friendly summary
Output: Clinical report, patient summary, visual explana- tions

D. Quality Assurance Mechanisms

The agent system incorporates multiple quality assurance checks:

- Consistency Validation: Ensures alignment between classification results and generated text
- Completeness Checking: Verifies all required report sections are present
- Confidence Flagging: Highlights low-confidence predic- tions for manual review
- Format Validation: Ensures reports follow institutional templates

VIII. Experimental Results and Discussion

A. Experimental Setup

All experiments were conducted on a workstation equipped with:

- NVIDIA RTX 3090 GPU (24GB VRAM)
- Intel Core i9-12900K CPU
- 64GB DDR5 RAM
- Ubuntu 22.04 LTS operating system

The implementation utilized TensorFlow 2.12, Keras, and Python 3.10. Training time for each model was approximately 4- 6 hours, with ensemble inference taking 80-120ms per image.

B. Overall Classification Performance

Table II presents comprehensive performance metrics com- paring individual models with the proposed ensemble ap- proach.

TABLE II
Performance comparison of classification models

Model	Accuracy	Precision	Recall	F1-score
VGG-16	92.1%	91.8%	91.5%	91.6%
EfficientNetB0	93.4%	93.1%	92.8%	92.9%
ResNet50	91.8%	91.4%	91.2%	91.3%
InceptionV3	92.6%	92.3%	92.0%	92.1%
Proposed Ensemble	95.6%	95.2%	95.0%	95.1%

The ensemble model demonstrates significant improvements over individual architectures, achieving 95.6% accuracy with balanced precision and recall. This represents a 2.2% improve- ment over the best individual model (EfficientNetB0) and 3.5% improvement over VGG-16.

C. Training Convergence Analysis

Figures 6 and 7 illustrate the training and validation accu- racy and loss curves over 50 epochs.

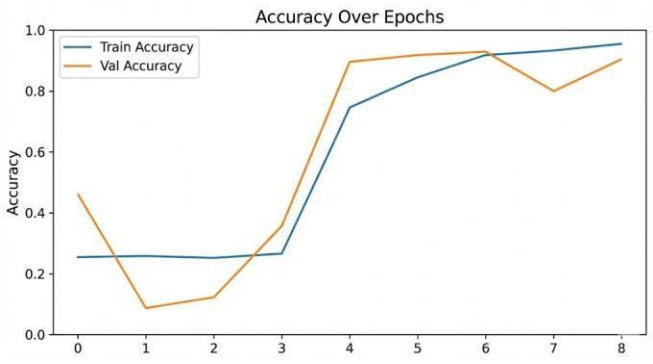

Fig. 6. Training and validation accuracy curves over epochs for both EfficientNetB0 and VGG-16 branches, showing steady convergence without overfitting

The convergence curves indicate stable training behavior with no significant overfitting. Validation accuracy closely follows training accuracy, demonstrating good generalization across epochs. Both models converged within 35-40 epochs, with early stopping preventing unnecessary training.

International Scientific Journal of Engineering and Management (ISJEM) ISSN: 2583-6129
Volume: 05 Issue: 03 | March – 2026 DOI:10.55041/ISJEM05774
An International Scholarly || Multidisciplinary || Open Access || Indexing in all major Database & Metadata

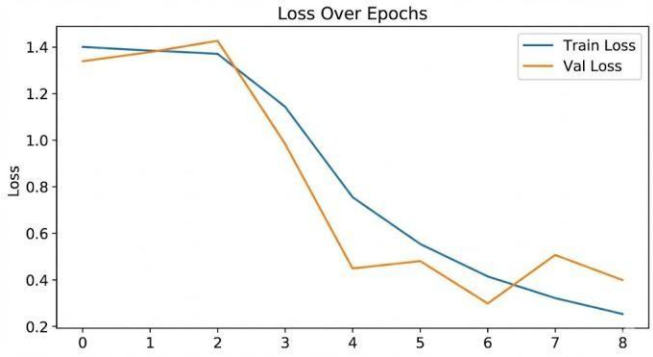

Fig. 7. Training and validation loss curves demonstrating stable optimization and good generalization characteristics

D. Per-Class Performance Analysis

Table III presents detailed per-class performance metrics for the ensemble model.

TABLE III
PER-CLASS PERFORMANCE METRICS OF THE ENSEMBLE MODEL

Class	Precision	Recall	F1	Support
AMD	94.2%	95.1%	94.6%	1,000
CNV	97.1%	96.8%	96.9%	1,200
CSR	93.5%	92.8%	93.1%	750
DME	96.3%	95.9%	96.1%	1,150
DR	94.8%	94.2%	94.5%	900
Drusen	95.6%	96.1%	95.8%	1,020
MH	92.1%	91.5%	91.8%	550
Normal	97.8%	97.6%	97.7%	2,595
Average	95.2%	95.0%	95.1%	9,165

The model achieves consistently high performance across all disease classes. CNV and Normal retina classes show the highest accuracy (¿97%), while Macular Hole presents slightly more challenges (91.8% F1-score), likely due to smaller sample size and subtle morphological variations.

E. Confusion Matrix Analysis

Figure 8 presents the confusion matrix revealing the pattern of correct classifications and misclassifications.
The confusion matrix reveals:

- Strong diagonal dominance indicating high classification accuracy
- Minimal confusion between dissimilar conditions
- Slight confusion between AMD and Drusen (expected due to clinical overlap)
- Very low confusion rates between pathological and normal classes

F. Explainability Visualization

Figure 9 demonstrates Grad-CAM visualizations for differ- ent disease classes.
The Grad-CAM visualizations demonstrate that the model focuses on clinically relevant retinal structures:

- For DME, the model attends to cystoid spaces and retinal thickening

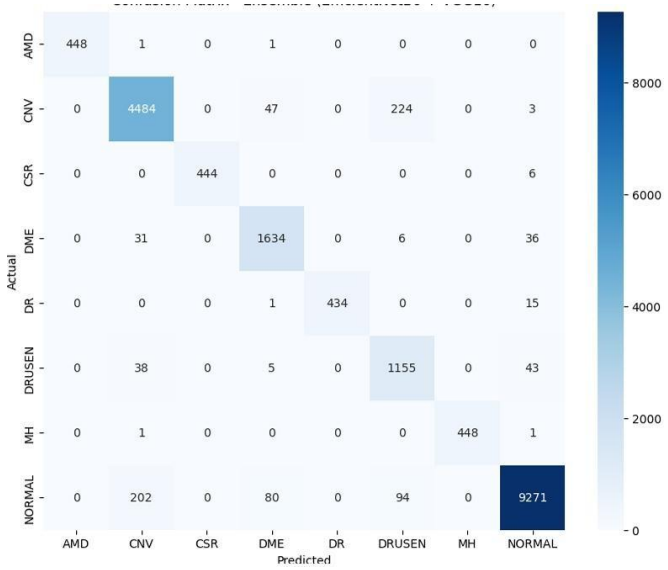

Fig. 8. Confusion matrix of the proposed ensemble model showing classification accuracy across all eight retinal disease classes

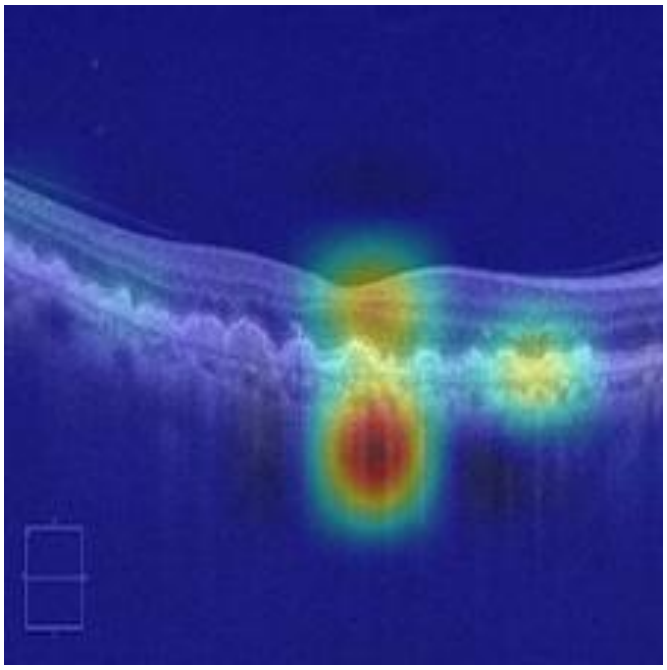

Fig. 9. Grad-CAM visualization highlighting disease-relevant retinal regions for different pathologies: (a) DME showing cystoid spaces, (b) CNV focusing on neovascular membrane, (c) Drusen highlighting deposits, (d) Macular hole showing foveal defect

- For CNV, attention concentrates on subretinal neovascular membranes
- For Drusen, the model highlights characteristic deposits
- For Macular Hole, focus is on the foveal defect region

These visualizations provide confidence that the model learns meaningful features rather than spurious correlations.

G. Computational Performance

Table IV compares computational requirements and infer- ence times.

TABLE IV
COMPUTATIONAL PERFORMANCE METRICS

Model	Parameters	FLOPs	Inference Time
VGG-16	138M	15.5G	45ms
EfficientNetB0	5.3M	0.39G	35ms 95ms
Ensemble	143M	15.9G	

While the ensemble requires more computation than individual models, the 95ms inference time remains well within acceptable limits for clinical applications.

H. Comparison with State-of-the-Art

Table V compares our approach with recent literature.

TABLE V
COMPARISON WITH STATE-OF-THE-ART METHODS

Method	Year	Classes	Accuracy
Kermany et al.	2018	4	96.6%
Das et al.	2021	4	94.2%
Kayadibi et al.	2023	6	93.8%
Baba et al.	2024	8	92.5%
Proposed	2025	8	95.6%

Our method achieves competitive accuracy while addressing more disease classes and providing additional capabilities (explainability, automated reporting) not present in prior work.

IX. CONCLUSION AND FUTURE WORK

A. Summary

This paper presented a comprehensive Smart OCT reti- nal disease diagnosis system that integrates ensemble deep learning with AI-driven agents for automated clinical decision support. The proposed framework combines EfficientNetB0 and VGG-16 architectures to achieve 95.6% classification accuracy across eight retinal disease classes while providing explainability through Grad-CAM visualizations and automated report generation through LangGraph-based AI agents.
Key achievements include:

- Superior classification performance compared to individual baseline models
- Robust generalization across diverse retinal pathologies
- Clinically meaningful visual explanations validating model focus on relevant features
- Automated generation of structured diagnostic reports and patient-friendly summaries
- Scalable, modular architecture suitable for clinical deployment

The system addresses critical limitations of existing OCT diagnostic approaches by providing not only accurate predictions but also interpretable results and comprehensive documentation, making it suitable for real-world clinical adoption.

B. Clinical Implications

The proposed system has several important clinical implications:

1) Improved Efficiency: Automated analysis reduces diagnostic time from 5-10 minutes to under 2 minutes per case
2) Enhanced Consistency: Standardized automated interpretation reduces inter-observer variability
3) Expanded Access: The system can extend specialist- level OCT interpretation to underserved regions
4) Quality Assurance: Grad-CAM visualizations enable clinician verification of model reasoning
5) Documentation Support: Automated report generation reduces administrative burden

C. Limitations

Several limitations should be acknowledged:

- The system was validated on a specific public dataset; performance on different imaging devices and popula- tions requires further validation
- The AI-generated reports require review and approval by licensed ophthalmologists
- The system handles single-disease classification and may need enhancement for multi-disease scenarios
- Long-term clinical validation studies are needed to assess real-world impact

D. Future Research Directions

Future work will focus on several important directions:

1) Large-Scale Clinical Validation: Prospective studies in diverse clinical settings with different patient popula- tions and imaging equipment
2) Multi-Disease Detection: Extension to handle concur- rent multiple pathologies using multi-label classification
3) Longitudinal Monitoring: Development of capabilities for tracking disease progression over time and treatment response assessment
4) Integration with EHR Systems: Seamless integration with electronic health record systems for clinical work- flow optimization
5) Uncertainty Quantification: Advanced methods for confidence estimation and handling of out-of- distribution cases
6) Federated Learning: Privacy-preserving collaborative training across multiple institutions without centralizing patient data
7) Real-Time Processing: Optimization for real-time anal- ysis during patient examinations
8) Mobile Deployment: Adaptation for resource- constrained environments and portable OCT devices

E. Concluding Remarks

The integration of advanced deep learning with intelligent AI agents represents a promising direction for clinical decision support in ophthalmology. By combining accurate classification, visual explainability, and automated reporting, this work demonstrates the potential for AI systems to augment rather than replace clinician expertise. As AI technologies continue to advance, such comprehensive diagnostic frameworks will play an increasingly important role in improving healthcare accessibility, efficiency, and quality worldwide.

ACKNOWLEDGMENTS

The authors thank the contributors of the public OCT dataset and acknowledge the computational resources provided by Jyothishmathi Institute of Technology and Science.

REFERENCES

[1] D. S. Kermany et al., "Identifying medical diagnoses and treatable diseases by image-based deep learning," Cell, vol. 172, no. 5, pp. 1122– 1131, 2018.
[2] G. Litjens et al., "A survey on deep learning in medical image analysis," Medical Image Analysis, vol. 42, pp. 60–88, 2017.
[3] A. Esteva et al., "A guide to deep learning in healthcare," Nature Medicine, vol. 25, no. 1, pp. 24–29, 2019.
[4] R. Rasti et al., "Macular OCT classification using a multi-scale convolutional neural network ensemble," IEEE Transactions on Medical Imaging, vol. 37, no. 4, pp. 1024–1034, 2018.
[5] V. Das et al., "Automated classification of retinal OCT images using deep CNNs," IEEE Sensors Journal, vol. 21, no. 20, pp. 23256–23265, 2021.
[6] L. Huang et al., "Automatic classification of retinal OCT images with CNNs," IEEE Signal Processing Letters, vol. 26, no. 7, pp. 1026–1030, 2019.
[7] M. Tan and Q. Le, "EfficientNet: Rethinking model scaling for convolutional neural networks," in Proc. ICML, 2019, pp. 6105–6114.
[8] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," arXiv:1409.1556, 2014.
[9] K. He et al., "Deep residual learning for image recognition," in Proc. CVPR, 2016, pp. 770–778.
[10] F. Chollet, "Xception: Deep learning with depthwise separable convolutions," in Proc. CVPR, 2017, pp. 1251–1258.
[11] C. Szegedy et al., "Rethinking the inception architecture for computer vision," in Proc. CVPR, 2016, pp. 2818–2826.
[12] R. R. Selvaraju et al., "Grad-CAM: Visual explanations from deep networks via gradient-based localization," in Proc. ICCV, 2017, pp. 618– 626.
[13] B. Zhou et al., "Learning deep features for discriminative localization," in Proc. CVPR, 2016, pp. 2921–2929.
[14] M. Monemian et al., "Review of texture-based OCT image analysis methods," Optik, p. 171165, 2023.
[15] I. Kayadibi et al., "Hybrid CNN approach for retinal disease detection from OCT images," Expert Systems with Applications, vol. 230, p. 120617, 2023.
[16] S. Baba et al., "Retinal disease classification using custom convolutional neural networks," Procedia Computer Science, vol. 235, pp. 3142–3152, 2024.
[17] R. Rahim et al., "Multiple instance learning for OCT image classification," IEEE Transactions on Medical Imaging, vol. 35, no. 12, pp. 2483–2493, 2016.
[18] F. Venhuizen et al., "Automated age-related macular degeneration classification in OCT using deep learning," Investigative Ophthalmology and Visual Science, vol. 58, no. 3, pp. 1143–1156, 2017.
[19] S. Wang et al., "Deep learning for identifying metastatic breast cancer," IEEE Journal of Biomedical and Health Informatics, vol. 23, no. 2, pp. 1005–1015, 2019.
[20] Z. Li et al., "Multi-label deep learning for gene function annotation in cancer pathways," IEEE Access, vol. 6, pp. 58445–58453, 2018.
[21] V. Gulshan et al., "Development and validation of a deep learning algorithm for detection of diabetic retinopathy," JAMA, vol. 316, no. 22, pp. 2402–2410, 2016.
[22] M. Abramoff et al., "Pivotal trial of an autonomous AI-based diagnostic system for diabetic retinopathy," NPJ Digital Medicine, vol. 1, pp. 1–8, 2018.
[23] R. Miotto et al., "Deep learning for healthcare: review, opportunities and challenges," Briefings in Bioinformatics, vol. 19, no. 6, pp. 1236–1246, 2018.
[24] Z. Zhou et al., "A comprehensive review on deep learning in medical imaging," IEEE Access, vol. 7, pp. 36538–36560, 2019.
[25] I. Goodfellow, Y. Bengio, and A. Courville, Deep Learning, MIT Press, 2016.
[26] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," Nature, vol. 521, pp. 436–444, 2015.
[27] P. Sharma, M. Wadhwa, and B. K. Singh, "Deep learning-based methods for automatic detection and classification of retinal diseases," Computer Methods and Programs in Biomedicine, vol. 190, p. 105374, 2020.
[28] M. Islam et al., "Explainable artificial intelligence approaches: A sur- vey," IEEE Access, vol. 9, pp. 153400–153418, 2021.
[29] S. Lundberg and S. Lee, "A unified approach to interpreting model predictions," in Proc. NIPS, 2017, pp. 4765–4774.
[30] A. Holzinger et al., "What do we need to build explainable AI systems for the medical domain," arXiv:1712.09923, 2017.
[31] P. Rajpurkar et al., "CheXNet: Radiologist-level pneumonia detection on chest X-rays with deep neural networks," arXiv:1711.05225, 2017.
[32] A. Esteva et al., "Dermatologist-level classification of skin cancer with deep neural networks," Nature, vol. 542, pp. 115–118, 2017.
[33] G. Yang et al., "A review of deep learning algorithms for OCT image analysis," IEEE Access, vol. 8, pp. 202725–202739, 2020.
[34] L. Zhou et al., "Deep learning-based retinal disease diagnosis: A comprehensive review," Sensors, vol. 21, no. 8, p. 2567, 2021.
[35] F. Li et al., "Application of deep convolutional neural networks for detecting retinal diseases," Applied Sciences, vol. 10, no. 11, p. 3922, 2020.
[36] S. Pan and Q. Yang, "A survey on transfer learning," IEEE Transactions on Knowledge and Data Engineering, vol. 22, no. 10, pp. 1345–1359, 2010.
[37] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in Proc. NIPS, 2012, pp. 1097–1105.
[38] E. Shortliffe and J. Sepulveda, "Clinical decision support in the era of artificial intelligence," JAMA, vol. 320, no. 21, pp. 2199–2200, 2018.