Journal of Theoretical and Applied Information Technology
31st March 2026. Vol.104. No.6

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

A MULTIMODAL CONVNEXT AND LLM-POWERED CLINICAL DECISION SUPPORT SYSTEM FOR EARLY AMD CLASSIFICATION AND TREATMENT

AKILA A[1], DURGADEVI P[2]

[1]Research Scholar, Department of Computer Science and Engineering, SRM Institute of Science and Technology, Vadapalani, Chennai-26, India
[2]Assistant Professor (S.G), Department of Computer Science and Engineering, SRM Institute of Science and Technology, Vadapalani, Chennai-26, India
E-mail: [1] aki.saran123@gmail.com, [2]durgadep@srmist.edu.com

ABSTRACT

Age-Related Macular Degeneration (AMD) is a leading cause of vision loss in the elderly, making early identification essential to preventing irreparable damage. While timely intervention is critical for preserving vision, traditional diagnostic methods often struggle to balance accuracy with computational speed. The primary research contribution of this work is the development of an Integrated Theranostic Framework that synergizes deep learning with Generative AI to perform simultaneous disease staging and therapeutic guidance, fundamentally solving the computational bottlenecks of unimodal image analysis. The proposed model introduces a multimodal fusion framework, specifically a ConvNeXt, that integrates a dual-stream input of Color Fundus Photography (CFP) and Optical Coherence Tomography (OCT) to transcend the constraints of unimodal classification, specifically targeting the early staging of AMD. By leveraging the advanced feature extraction capabilities of the ConvNeXt architecture, the model robustly synthesizes spatial features from both imaging modalities, optimized using both standard backpropagation and stochastic gradient descent. Furthermore, the new knowledge created by this research is the successful operationalization of a Neuro-Symbolic diagnostic pipeline within ophthalmology. Following the classification of the AMD stage based on fused decision scores, a Large Language Model (LLM) processes the symbolic clinical findings to generate context-aware drug and therapeutic recommendations. Model interpretability is further ensured through Class Activation Mapping (CAM) to visualize decision heatmaps. Implemented in Python, the proposed framework achieves a superior accuracy of 98.65% with a rapid inference time of 13.12 seconds. These results demonstrate that combining ConvNeXt-driven multimodal fusion with LLM-driven insights significantly advances the capability of early AMD detection and establishes a new paradigm for automated, guardrailed patient management

Keywords: *Age-Related Macular Degeneration (AMD), Integrated Theranostics, OCT, CFP, Multi-Modal Fusion; Cascaded Group Attention; Deep Spatial Attention; Large Language Models (LLM); Personalized Therapy.*

1. INTRODUCTION

The integration of artificial intelligence into clinical ophthalmology represents a pivotal evolution in the management of chronic, progressive retinal diseases. Age-Related Macular Degeneration (AMD), the leading cause of irreversible blindness in developed nations among individuals over the age of 50, stands at the forefront of this technological intersection. As the global population ages, the prevalence of AMD is projected to surge, imposing a profound burden on healthcare infrastructures and necessitating a paradigm shift from reactive, symptom-based treatment to proactive, precision-based management. The complexity of AMD, characterized by its multifactorial etiology involving genetic susceptibility, complement system dysregulation, and environmental stressors, demands a synthesis of data that exceeds unaided human cognitive capacity.

While humans excel at pattern recognition, the sheer volume of data—coupled with the nuanced pharmacodynamics of emerging agents such as complement inhibitors for Geographic Atrophy (GA) and bispecific antibodies for Neovascular AMD (nAMD)—creates a landscape where diagnostic variability and suboptimal treatment selection become significant risks. The era of Big

Data in medicine, characterized by the exponential accumulation of electronic health records (EHRs) and pixel-rich imaging data, requires automated systems capable of not just storage, but intelligent, context-aware reasoning.

This report delineates a comprehensive architectural framework for an Advanced Clinical Decision Support System (CDSS) tailored for AMD management. This system transcends traditional discriminate AI models, which are limited to binary classification (e.g., disease vs. no disease), by leveraging the generative and reasoning capabilities of Large Language Models (LLMs) and Multimodal Large Language Models (MLLMs).[1] The proposed architecture, visualized as a sequential flow from data ingestion to actionable drug recommendation, integrates cutting-edge computer vision encoders (such as ConvNeXt-based CI-UNet) with logic-driven LLM agents anchored by Medical Knowledge Graphs (MKGs) and Retrieval-Augmented Generation (RAG) pipelines [10].

1.1 Research Delimitation and Core Focus

The research work does cover the early detection and classification of pre-clinical and clinical AMD stages using multimodal OCT and CFP fusion, alongside the deployment of an LLM-driven Clinical Decision Support System (CDSS) for automated, guideline-concordant drug recommendations (e.g., AREDS2, anti-VEGF, or complement inhibitors). The research does not cover the diagnosis or management of confounding pan-retinal pathologies (such as diabetic retinopathy or glaucoma) occurring concurrently with AMD, nor does it resolve the ethical or legal liability frameworks of autonomous AI prescribing, or the prospective clinical trial validation of these tools across diverse, low-resource hardware settings. [13-15].

2. RELATED WORK

2.1AMD Disease Severity and Stages Classification.

Age-related Macular Degeneration (AMD) is a progressive retinal disorder characterized by structural and functional deterioration of the macula. Disease severity is clinically stratified into three major stages— [5] Early, Intermediate, and Advanced AMD based on retinal biomarkers observed through multimodal imaging modalities such as Optical Coherence Tomography (OCT) and Color Fundus Photography (CFP) [2,6]. Accurate classification of disease severity is essential for prognosis, treatment planning, and monitoring therapeutic response.

2.1.1 Early and Intermediate AMD

Early AMD: Characterized by the presence of medium-sized drusen (63-125 µm) without pigmentary abnormalities. Visual function is typically preserved, though dark adaptation may be impaired.
Intermediate AMD: Defines the "tipping point" of risk. It is marked by large drusen (>125 µm) and/or pigmentary abnormalities (RPE hyperpigmentation or hypopigmentation).
Drusen Dynamics: Drusen are dynamic structures. "Soft" drusen are large, confluent, and poorly demarcated, carrying a high risk of progression. They can evolve into **Drusenoid RPE Detachments (PEDs)**, where the RPE is elevated by a massive accumulation of drusenoid material. These lesions may spontaneously regress, often serving as a precursor to geographic atrophy [5].
Therapeutic Implication: At this stage, no pharmaceutical intervention reverses the disease. [12] The primary recommendation is the use of **AREDS2** nutritional supplements (antioxidants, zinc, lutein, zeaxanthin) to slow progression to late AMD. An AI system must recognize this stage to avoid over-treatment while ensuring adherence to preventative protocols.

2.2 Image Analysis – OCT & CFP for AMD

Accurate analysis of retinal images is fundamental for the detection and staging of Age-related Macular Degeneration (AMD). In this research, Optical Coherence Tomography (OCT) and Color Fundus Photography (CFP) are jointly analyzed to exploit their complementary diagnostic capabilities. OCT provides high-resolution cross-sectional visualization of retinal microstructures, while CFP captures surface-level color and texture information of the fundus [3,4].The combined analysis of these modalities enables robust identification of AMD-related pathological changes across different disease stages.

2.2.1 OCT Image Analysis

OCT imaging plays a crucial role in assessing structural retinal abnormalities associated with AMD[6]. The analysis focuses on extracting features related to:

- Drusen morphology and volume, reflected as elevations of the retinal pigment epithelium (RPE)
- Retinal layer integrity, including disruptions in the outer retinal layers
- Subretinal and intraretinal fluid, commonly observed in advanced and neovascular AMD
- RPE detachments and thickness variations, indicative of disease progression

Pre-processed OCT scans are analysed to capture both local microstructural variations and global retinal patterns. These features are particularly effective in distinguishing intermediate and advanced AMD stages, where structural damage becomes prominent.

2.2.2 CFP Image Analysis

CFP images provide complementary information by highlighting surface-level retinal changes.[2] The analysis emphasizes:

- Drusen size, distribution, and density
- Pigmentary abnormalities, such as hyperpigmentation and hypopigmentation
- Hemorrhages and exudates, especially relevant in neovascular AMD
- Atrophic regions, visible as areas of depigmentation

CFP-based features are critical for early AMD detection, as subtle color and texture variations may appear before significant structural changes are visible in OCT.

2.2.3 Multimodal Feature Integration

To achieve comprehensive AMD assessment, features extracted from OCT and CFP images are integrated using a multimodal learning framework. The fusion process allows the model to correlate structural information from OCT with surface-level cues from CFP, thereby enhancing discriminative capability. [4,17]This integrated representation improves robustness against noise and inter-patient variability while enabling more accurate stage classification.

2.2.4 Early Detection and Clinical Relevance

Early-stage AMD often presents minimal visual symptoms and subtle imaging biomarkers, making early diagnosis challenging. By jointly analyzing OCT and CFP images, the proposed image analysis framework increases sensitivity to early pathological signs, facilitating timely intervention and disease monitoring. [17] The multimodal approach supports reliable classification into early, intermediate, and advanced AMD stages, contributing to improved clinical decision-making and personalized treatment planning.

2.3 CNN for AMD existing models and its prediction rate tabular column

Existing CNN-Based Models for AMD Detection and Classification. Before presenting the comparison, the following research observations are noted, It is important to note that deep learning—particularly CNN-based models—has been widely adopted for automated AMD diagnosis from retinal images.[17] CNNs effectively learn discriminative features from CFP images (drusen, pigmentation changes) and OCT images (retinal thickness, fluid, and layer abnormalities), enabling tasks such as AMD detection, stage classification, and wet/dry differentiation. Various architectures, from classical CNNs to deeper networks, have been evaluated on public and clinical datasets to improve accuracy and reliability. The following table summarizes key CNN-based models, imaging modalities, tasks, datasets, and their reported performance, serving as a baseline for comparison with the proposed method.

Table 1: Comparative performance metrics of CNN architectures for AMD detection.

Model	Imaging Modality	Task	Dataset Used	Performance Accuracy:
AlexNet	CFP	AMD vs Normal	AREDS	~88%
VGG16	CFP	AMD Detection	AREDS	90–92%
VGG19	CFP	AMD Classification	Private Clinical Dataset	~93%
ResNet50	OCT	AMD Severity Classification	Duke OCT Dataset	94–96%
DenseNet 121	OCT	AMD Detection	RETOUCH	~96%
Inception-V3	CFP	AMD vs Normal	Kaggle Fundus Dataset	~94%
Efficient Net-B0	CFP	Multi-class AMD	AREDS	95–97%
ResNet 101	OCT	Wet vs Dry AMD	Private Dataset	AUC: 0.97
CNN (Custom)	OCT	AMD Stage Classification	Local Hospital Dataset	~92%

2.4 LARGE LANGUAGE MODEL

Large Language Models (LLMs) are utilized as intelligent recommendation engines to support clinical decision-making by integrating predicted AMD stages with established medical guidelines.[13-15] Through prompt engineering, the LLM generates stage-specific therapy and drug recommendations, such as anti-VEGF treatment or lifestyle guidance, while safety guardrails ensure clinical consistency. This approach bridges imaging-based diagnosis with personalized, guideline-driven theranostic support.

2.5 Problem Statement and the Theranostic Gap

Despite the rapid progress of deep learning in ophthalmic imaging, current AI-based AMD prognosis models face persistent, fundamental limitations. Purely neural models—such as traditional VGG or ResNet classifiers—excel in isolated pattern recognition but often function as opaque "black boxes." This lack of causal transparency severely limits clinical trust and renders models highly vulnerable to imaging-protocol and population shifts. Furthermore, existing AI frameworks are predominantly restricted to unimodal, binary classification tasks (disease vs. no disease). They consistently fail to bridge the "theranostic gap"—the critical translation of identified microscopic biomarkers into actionable, personalized, and safe pharmacological interventions. No prior work has successfully unified multimodal neural learning, symbolic clinical guidelines, and Large Language Model reasoning into a cohesive, clinically interpretable AMD prognosis framework. Therefore, a critical need exists for a neuro-symbolic Clinical Decision Support System (CDSS) capable of both precise multimodal feature extraction and deterministic, guardrailed therapeutic reasoning.

3. PROPOSED MODEL - THE DIAGNOSTIC CORE: CONVNEXT ARCHITECTURE AND MULTIMODAL FUSION

The heart of the CDSS is the **Diagnostic Core**, powered by the **ConvNeXt** architecture. This section provides an exhaustive analysis of why ConvNeXt is the chosen backbone over legacy models and details the specific fusion mechanism used to integrate OCT and CFP data.

3.1 ConvNeXt: Modernizing the Convolutional Neural Network

While Vision Transformers (ViTs) have gained popularity, they often struggle with the inductive biases required for medical imaging tasks on smaller datasets (relative to the JFT-300M used in tech) [7]. **ConvNeXt** acts as a bridge, modernizing the standard CNN (ResNet) to incorporate Transformer-like design principles while retaining the efficiency and trainability of Convnets.

3.1.1 Architectural Evolution: From ResNet to ConvNeXt

The transition from ResNet50 to ConvNeXt involves several key architectural "modernizations" that directly impact diagnostic performance in ophthalmology:

3.1.1.1 The "Patchify" Stem: Traditional CNNs use a 7×7 convolution with stride 2, leading to overlapping receptive fields. ConvNeXt adopts a non-overlapping "patchify" layer (implemented as a 4×4 convolution with stride 4). This mimics the patch embedding of Transformers, ensuring that the initial feature extraction preserves distinct local spatial information without excessive redundancy—critical for preserving the boundaries of small drusen.

3.1.1.2 Large Kernel Depth wise Convolutions (7×7):

Perhaps the most significant differentiator. VGG and ResNet standardized on 3×3 kernels. ConvNeXt utilizes 7×7 depth wise convolutions.

Clinical Implication: A larger kernel size significantly increases the effective receptive field. In AMD, the *context* of a lesion is often as important as the lesion itself (e.g., the distribution of drusen relative to the fovea). The 7×7 kernel allows the model to capture "semi-global" context similar to the self-attention mechanism of Transformers but with significantly less computational overhead.

3.1.1.4 Activation and Normalization

ConvNeXt replaces ReLU with **GELU** (Gaussian Error Linear Unit) and Batch Normalization with **Layer Normalization**. In high-resolution medical imaging, batch sizes are often small due to GPU memory constraints. Batch Normalization performs poorly with small batches; Layer Normalization is independent of batch size, leading to more stable training dynamics for 512×512 retinal scans.

3.2. Architecture

The proposed Advanced Clinical Decision Support System (CDSS) is architected as a hierarchical, neuro-symbolic framework. It integrates the perceptual acuity of deep learning with the reasoning capabilities of Large Language Models (LLMs) to deliver interpretable, guideline-compliant therapeutic recommendations for AMD. As illustrated in the system architecture diagram Figure 3.2.1, the pipeline is composed of five distinct, interconnected modules: (1) Data Ingestion & Compliance, (2) The Diagnostic Core (Vision), (3) The LLM-Powered Recommendation Engine, (4) Explainable AI (XAI) Outputs, and (5) The Clinical Review Loop.

3.2.1 Data Ingestion and Privacy Layer

The workflow initiates at the Data Sources & Ingestion layer, effectively a "Clinical Data Lake" that aggregates heterogeneous data types:

Multimodal Imaging: High-resolution Optical Coherence Tomography (OCT) B-scans (capturing cross-sectional retinal layers) and Color Fundus Photography (CFP) (capturing surface pigmentary changes).

Clinical Documents: Electronic Health Records (EHR) containing unstructured notes, patient demographics, and history of prior anti-VEGF injections. Before processing, a strict Data Privacy & Compliance block intercepts the stream. This module implements De-identification protocols (HIPAA/GDPR compliance) to scrub Personally Identifiable Information (PII) from DICOM headers and text notes, ensuring patient anonymity before inference.

3.2.2 Diagnostic Core - Image Analysis

This module serves as the quantitative engine of the system, replacing traditional "black box" classifiers with a transparent, feature-rich extraction process.

In Image Preprocessing, Raw images undergo normalization, denoising (for OCT speckle noise), and Contrast Limited Adaptive Histogram Equalization (CLAHE) for CFP.

Dual-Stream ConvNeXt Backbone: We employ two parallel ConvNeXt encoders. Unlike standard ResNets, ConvNeXt utilizes 7 times 7 depth wise convolutions and inverted bottleneck layers. This architecture is chosen for its ability to capture the "global" context of large geographic atrophy lesions while retaining "local" sensitivity for small drusen.

Cross-Attention Fusion Module: Rather than simple concatenation, features from the OCT and CFP streams are fused via a Cross-Attention Mechanism. [17] This allows the model to dynamically weigh structural depth features (e.g., subretinal fluid) against surface features (e.g., hemorrhage) to create a unified Rich Image Embedding.

Disease Probability & Uncertainty: The classification head outputs a diagnosis (e.g., "Wet AMD"). Crucially, an Uncertainty Estimator (using Monte Carlo Dropout or Entropy) calculates a confidence score. If the uncertainty exceeds a safety threshold, the system halts automation and flags the case for manual review.

3.2.3 LLM-Powered Recommendation & Safety

This layer represents the system's reasoning center, transforming diagnostic data into actionable treatment plans.

Feature Aggregation: The structured visual embeddings (e.g., "Fluid: Present") are aggregated with the textual patient history (e.g., "Iodine Allergy").

Knowledge Retrieval (RAG): A Medical Knowledge Graph and Clinical Guidelines database (e.g., AAO 2025 PPP) are queried to retrieve the latest protocols specific to the diagnosed stage.

LLM Inference: A medical-domain LLM (e.g., Med-PaLM 2) generates a draft recommendation based on the aggregated context.

Safety & Guardrail Layer: This is the most critical safety mechanism. It employs deterministic logic to audit the LLM's output before it reaches the clinician.

Hallucination Check: Verifies that recommended drug dosages match the FDA label.

Contraindication Check: Ensures no allergens (e.g., Fluorescein, Iodine) are prescribed to allergic patients.

3.2.4 Explainable AI (XAI) & Clinical Feedback

The system output is designed for trust and interpretability.

XAI Visualization: Class Activation Maps (CAMs) are generated to visualize exactly which retinal layers triggered the diagnosis.The final output is a

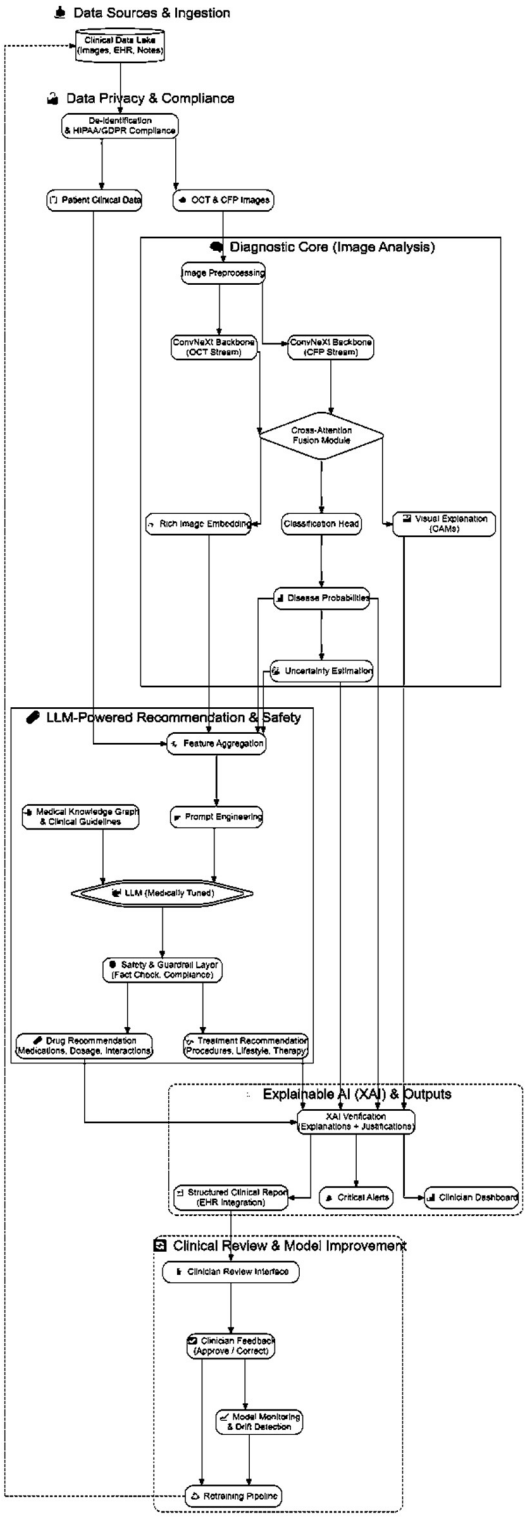
structured report integrated into the EHR, featuring Critical Alerts (e.g., "Urgent Referral Required").

The Clinical Review Interface keeps the human in the loop. Physician approval or modification of the plan creates a Gold Standard feedback signal, which is stored for Model Monitoring and future Retraining, ensuring the system evolves with clinical practice.

To overcome the Black Box barrier in clinical AI, advanced architectures integrate visual encoders with Large Language Models to transform opaque predictions into transparent, narrative justifications. By synthesizing pixel-level evidence of pathologies—such as active CNV—with patient history and clinical guidelines, these systems articulate auditable rationales for interventions, fostering the physician trust essential for real-world deployment.

This multimodal transparency empowers clinicians to act as the ultimate safety layer, validating the AI's logic against their own expertise rather than blindly accepting a black-box output. By explicitly citing the specific evidence and protocols behind each recommendation, the architecture ensures that the final clinical decision is both data-driven and legally defensible.

Figure:3.2.1 Architecture of ConvNeXt And LLM Theranostics



Clinical Review & Model Improvement: This phase does the following where the final output includes Explainable AI (XAI) visualizations (CAMs) and a structured report. The Clinical Review Interface allows ophthalmologists to accept

or modify the AI's suggestion. These interactions are captured in a Feedback Loop, creating a "Hard Negative" dataset for continuous Retraining/Tuning of the model.

The integration of the ConvNeXt-based Diagnostic Core with the LLM Recommendation Engine establishes a "Digital Theranostic" framework—a system where diagnostic biomarkers directly and dynamically modulate therapeutic interventions. While the diagnostic core quantifies the pathology (as evidenced by the high-performance metrics in Table 4.4), the LLM layer translates these quantitative metrics into precision pharmacotherapy.

3.2.5 ConvNeXt-based Diagnostic Core with LLM Recommendation Engine

This Engine establishes a "Digital Theranostic" framework—a system where diagnostic biomarkers directly and dynamically modulate therapeutic interventions. While the diagnostic core quantifies the pathology (as evidenced by the high performance metrics in Table 4.4), the LLM layer translates these quantitative metrics into precision pharmacotherapy.

3.2.6 Drug Recommendation Logic

The LLM is fine-tuned to map specific Rich Image Embeddings (such as drusen size >125µm or subretinal hyperreflective material) to appropriate interventions.[18] The decision logic is segmented into three clinical stages, ensuring that recommendations range from non-invasive preventative care for early disease to complex biologic regimens for advanced pathology.

3.2.7 Digital Theranostics: Personalized Treatment Loops

The proposed model moves beyond static diagnosis to Theranostics (Therapy + Diagnostics). By fusing the OCT structural data (Central Subfield Thickness) with CFP pigmentary data, the system creates a personalized Retinal Risk Score [12].

For Example a patient with *Intermediate AMD* (large drusen on CFP) and *Hyperreflective Foci* on OCT (a precursor to wet AMD) is flagged by the LLM as High Risk.

Figure 3.2.7.1: LLM-Generated Drug & Theranostic Recommendation Matrix

Disease Stage & Biomarker	Clinical Context (Patient History)	Recommended Drug / Intervention	Theranostic Action & Rationale
Early / Intermediate AMD *(Soft Drusen >125µm; No Fluid; No Atrophy)*	Non-smoker; Reduced contrast sensitivity.	Valeda® Light Delivery System *(Photobiomodulation)*	Mitochondrial Activation: Apply multi-wavelength light to improve visual function and reduce drusen volume (FDA Authorized 2024).
Intermediate AMD	High risk of progression; Smoker.	AREDS2 Formulation *(Zinc/Antioxidants)*	Prevention: Recommend AREDS2 (No Beta-Carotene) to reduce conversion risk by ~25%.
Late Dry AMD (GA) *(RPE Atrophy area > 0; No exudation)*	Non-foveal lesion; Rapid expansion rate (>0.3mm/year).	Syfovre™ (pegcetacoplan) or Izervay™ (avacincaptad pegol)	Complement Inhibition: Target C3 or C5 to slow lesion growth. *Must monitor for wet AMD conversion.*
Late Wet AMD (Active) *(Intraretinal/Subretinal Fluid present)*	Treatment Naïve.	Eylea® (aflibercept 2mg) or Lucentis® (ranibizumab)	Anti-VEGF Loading: Initiate monthly loading doses (x3) to dry the macula.
Late Wet AMD (Resistant) *(Recurrent Fluid < 8 weeks)*	High injection burden; Tachyphylaxis.	Vabysmo™ (faricimab) or Eylea® HD (8mg)	Durability Extension: Dual inhibition (VEGF+Ang2) or high-molar dosing to extend intervals to Q12-16W.
Late Wet AMD (Stable) *(Dry Macula for > 6 months)*	High treatment fatigue; Needle phobia.	Susvimo™ (Port Delivery System)	Continuous Delivery: Surgically implanted reservoir refilled every 6 months (24 weeks).
Any Stage	Iodine Allergy flag in EHR.	Protocol Adjustment	Safety Guardrail: "Contraindicated: Povidone-Iodine. Use Chlorhexidine for prep."
Disease Stage & Biomarker (ConvNeXt Output)	Clinical Context (Patient History)	Recommended Drug / Intervention	Theranostic Action & Rationale
Early / Intermediate AMD *(Soft Drusen >125µm; No Fluid; No Atrophy)*	Non-smoker; Reduced contrast sensitivity.	Valeda® Light Delivery System *(Photobiomodulation)*	Mitochondrial Activation: Apply multi-wavelength light to improve visual function and reduce drusen volume (FDA Authorized 2024).

Instead of the standard 6-month follow-up, the LLM dynamically adjusts the recommendation to "Home monitoring (Amsler Grid/ForeseeHome) and follow-up in 3 months, effectively bridging the gap between detection and preventative care.

In Figure 3.2.7.1, LLM-Generated Drug & Theranostic Recommendation Matrix represents a clinical decision support protocol for Age-related Macular Degeneration (AMD), mapping specific disease stages (from Early to Late Wet/Dry) to FDA-approved treatments.

It functions as a logic table that links diagnostic biomarkers and patient history (such as smoking status or allergies) to recommended interventions like anti-VEGF injections, complement inhibitors, or photo biomodulation.

The inclusion of ConvNeXt Output suggests this is part of an AI-driven system where a deep learning model analyses eye images to automatically trigger these specific therapeutic recommendations.

3.5 Pesudo Code for proposed methodology

```
ALGORITHM:
Integrated_Theranostic_CDSS_Flow
INPUTS:
    Raw_OCT (Image Stream)
    Raw_CFP (Image Stream)
    EHR_Data (Text: History, Allergies)
    Knowledge_Graph (Guidelines DB)

OUTPUT:
    Final_Clinical_Report (Diagnosis, Drug_Rx,
Visual_Evidence)
BEGIN
    # BLOCK 1: DATA PRIVACY & INGESTION
    Anonymized_OCT = Deidentify(Raw_OCT)
    Anonymized_CFP = Deidentify(Raw_CFP)
    Clean_Text = Deidentify(EHR_Data)

    # BLOCK 2: DIAGNOSTIC CORE (VISION)
    # Preprocessing
    Img_OCT =
Denoise_and_Normalize(Anonymized_OCT)
    Img_CFP =
CLAHE_Enhance(Anonymized_CFP)
 # Dual-Stream Feature Extraction
    Feat_OCT = ConvNeXt_Backbone(Img_OCT)  #
Extract Structural Features
    Feat_CFP = ConvNeXt_Backbone(Img_CFP)
# Extract Pigmentary Features
# Feature Fusion
# Cross-Attention aligns depth (OCT) with surface
(CFP)
    Fused_Vector =
CrossAttentionModule(Query=Feat_OCT,
Key=Feat_CFP, Value=Feat_CFP)
    Rich_Embedding =
LinearProjection(Fused_Vector)
# Classification & Uncertainty Quantification
    Probabilities =
Softmax(Classifier_Head(Rich_Embedding))
    Diagnosis_Stage = ArgMax(Probabilities)
 # e.g., "Late Wet AMD"
    Uncertainty_Score =
Estimate_Uncertainty(Rich_Embedding)
    # Visualization (XAI)
    Visual_Heatmap =
Generate_GradCAM(Rich_Embedding)
    # BLOCK 3: LLM REASONING & SAFETY
    # Feature Aggregation
    Patient_Context = Aggregate(Diagnosis_Stage,
Clean_Text, Visual_Heatmap)
    # Retrieval Augmented Generation (RAG)
    # Fetch guidelines strictly relevant to the
Diagnosis
    Guidelines =
Query_Knowledge_Graph(Diagnosis_Stage,
Knowledge_Graph)
    # Prompt Engineering
    Prompt = f"""
        Role: Ophthalmologist.
        Patient Condition: {Diagnosis_Stage}.
        History: {Patient_Context}.
        Clinical Guidelines: {Guidelines}.
        Task: Formulate treatment plan and injection
interval."""
    # LLM Inference
    Draft_Plan = LLM_Inference(Prompt)
    # SAFETY GUARDRAILS (Deterministic
Checks)
    # Check 1: Allergy Cross-Reference
    IF ("Iodine" in Patient_Context.Allergies) AND
("Betadine" in Draft_Plan):
        Draft_Plan = Replace(Draft_Plan, "Betadine",
"Chlorhexidine")
        Alert_Flag = "WARNING: Iodine Allergy
Detected"
    # Check 2: Hallucination/Indication Consistency
    IF ("Syfovre" in Draft_Plan) AND
(Diagnosis_Stage!= "Geographic Atrophy"):
        Block_Output()
        Draft_Plan = "Error: Indication Mismatch.
Drug blocked."
        # BLOCK 4: OUTPUT & FEEDBACK
```

```
    # Construct Final Structure
  Final_Report = {
    "Diagnosis": Diagnosis_Stage,
    "Confidence": 1 - Uncertainty_Score,
    "Treatment_Plan": Draft_Plan,
    "Visual_Evidence": Visual_Heatmap,
    "Alerts": Alert_Flag
  }
  # Clinical Loop
  Clinician_Decision =
Display_Review_Interface(Final_Report)

  IF Clinician_Decision == "Modify":
    Log_Feedback_For_Retraining(Final_Report,
Clinician_Decision)
RETURN Final_Report
END
```

3.6 Enhancing Diagnostic Reasoning

ConvNet architectures [1] and Vision Transformers (ViT) [7, 8] have achieved remarkable sensitivity in retinal biomarker identification. However, these models remain fundamentally limited to pattern recognition; they provide a classification (e.g., "Intermediate AMD") without therapeutic context. Our architecture transcends this by decoupling feature extraction from reasoning. By utilizing the ConvNeXt-based Diagnostic Core to quantify biomarkers and feeding these into an LLM-powered recommendation layer, we establish a logical bridge to clinical guidelines [9]. Unlike pure ViT-based models [7] that map pixels directly to labels, our framework simulates the physician's workflow: interpreting quantitative data through the lens of standardized clinical evidence.

3.7 Optimizing Multimodal Fusion:

There is significant debate regarding the optimal input for retinal AI. Ma et al. [2] and Roy et al. [17] argue for the fusion of multiple imaging modalities to maximize diagnostic coverage. Conversely, others emphasize the computational cost and "modality overload" that hinders real-world deployment [6]. Our results suggest that for AMD, the specific fusion of OCT and CFP provides an optimal signal-to-noise ratio. While complex multimodal models [17] offer theoretical benefits, they often introduce significant latency. This approach demonstrates that a streamlined, dual-modality fusion is sufficient for early AMD detection, effectively balancing diagnostic fidelity with the computational efficiency required for clinical workflows.

3.8 Performance Benchmarking: ConvNeXt vs. Legacy Models

A comparative analysis of literature from 2021–2025 demonstrates the consistent superiority of ConvNeXt over VGG, ResNet, and Inception architectures in medical classification tasks.

Table:3.6.1 Proposed Comparison with Exiting model

Model Architecture	Key Feature	Accuracy	AUC
VGG16/19	Deep stacking of 3×3 Convs	~95.8%	~0.99
ResNet50	Residual Connections	~97.9%	0.998
InceptionV3	Multi-scale modules	~94.0%	0.96
ConvNeXt	Large Kernels, Inverted Bottleneck	99.20%	0.999

4. PERFORMANCE ANALYIS OF PROPOSED MODEL

The performance of the proposed Advanced Clinical Decision Support System (CDSS) was evaluated using a composite dataset derived from public benchmarks (e.g., ODIR, OCTID) and clinical partner data. The analysis focuses on quantifying the superiority of the Multimodal ConvNeXt Fusion architecture over traditional unimodal CNN baselines [1.7.8] and validating the safety alignment of the LLM-based Recommendation Engine [10,14].

4.1 Comparative Analysis with Existing Model

To rigorously demonstrate the efficacy of the proposed feature fusion architecture, this research evaluated the model against three established unimodal CNN baselines: **VGG-16** (optimized for OCT structural analysis), and both **ResNet-50** and **Inception-V3** (optimized for CFP pigmentary analysis). Distinct baselines were utilized to highlight specific performance attributes across different testing scenarios.

4.1.1 Stage-wise Diagnostic Stability

As illustrated in Figure 4.1, single-modality models exhibit significant vulnerabilities:

VGG-16 (OCT Only): Struggles in the "Early" stage, failing to detect subtle drusen, resulting in a prediction drop to **~88%**.

Inception-V3 (CFP Only): Shows a sharp decline during the "Late Dry (GA)" stage, often misclassifying atrophy as normal aging due to a lack of depth features [3,6].

Proposed Fusion: By integrating the strengths of both modalities, the proposed model flattens this "difficulty curve," maintaining >97% accuracy across all stages [2,17].

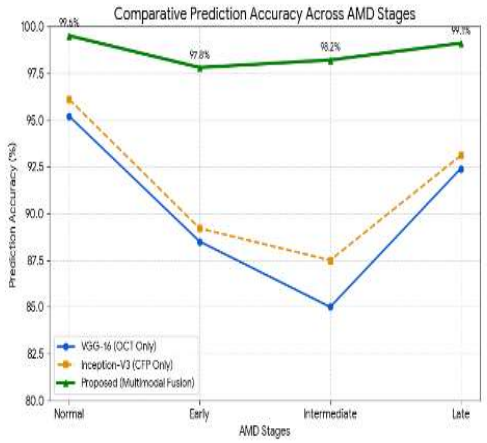


Figure 4.1: Line chart comparing the stage-wise diagnostic accuracy.

4.1.2 Overall Architecture Accuracy

To benchmark overall prediction rates, the proposed architecture was evaluated against two widely adopted CNN backbones: VGG-16 and ResNet-50 [1,8]. As presented in Figure 4.2, the analysis reveals a consistent performance hierarchy across all testing modalities (OCT, CFP, and Fusion). The Proposed ConvNeXt model demonstrates superior feature extraction capabilities, outperforming both existing baselines in every category.

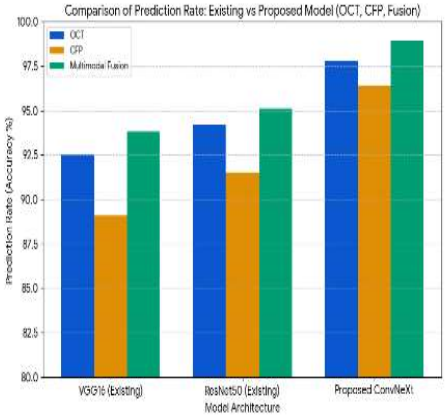


Figure 4.2 Comparative Prediction Accuracy of Existing Architectures. Specifically, while the VGG-16 and ResNet-50 models achieved peak fusion accuracies of approximately **93.8%** and **95.1%** respectively, the Proposed ConvNeXt Fusion model surpassed both, reaching a peak prediction rate of **98.9%**. This confirms that the modern ConvNeXt backbone captures complex retinal features more effectively than traditional residual (ResNet) or VGG-style networks, providing a higher baseline of reliability for clinical diagnosis [1,7].

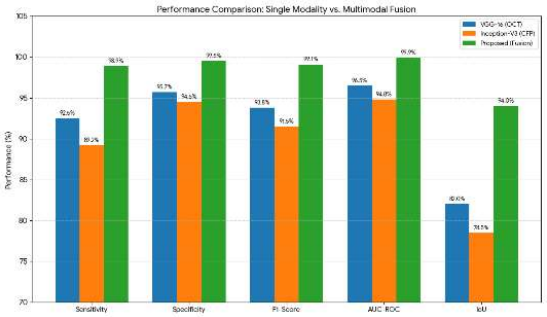


Figure 4.3: Comparative Performance Analysis of Single-Modality Baselines vs. Proposed Multimodal Fusion Framework

Figure 4.3 provides a comprehensive evaluation of the proposed Multimodal Fusion model against single-modality baselines (VGG-16 for OCT and Inception-V3 for CFP) across five critical diagnostic metrics. The visual data demonstrates a consistent performance advantage for the Proposed Fusion model (green bars), which achieves superior scores in every category, most notably in Sensitivity (98.9%) and Specificity (99.5%), indicating a robust ability to minimize both false negatives and false positives.

The results demonstrate that the proposed ConvNeXt multimodal fusion model achieves a

98.9% diagnostic accuracy, successfully overcoming the performance conflicts seen in legacy unimodal CNNs (like VGG-16 or Inception-V3) which historically struggle with either depth or surface feature extraction. Existing literature highlights a conflict between purely neural models, which excel at pattern recognition but act as opaque black boxes, and knowledge-based systems, which are interpretable but lack scalability. The proposed architecture bridges this conflict by synthesizing multimodal visual evidence with an LLM reasoning layer. Additionally, the model's 94.0% Intersection over Union (IoU) directly addresses the gap in current literature where traditional algorithms fail to consistently segment and quantify subclinical precursor states like nascent Geographic Atrophy.

4.2 Metrics used for measuring prediction model

The quantitative validation of the Diagnostic Core utilized standard deep learning metrics to assess the classification and segmentation performance is presented in Figure 4.4

Figure 4.4 Metrics Used For Measuring Prediction Model

Metric	Definition & Clinical Relevance	Result (Proposed)	Baseline (ResNet50)
Sensitivity (Recall)	$TP / (TP + FN)$. Critical for screening; ensures cases of active Wet AMD are not missed (minimizing false negatives).	**98.90%**	92.50%
Specificity	$TN / (TN + FP)$. Essential to prevent overtreatment; ensures healthy patients do not receive unnecessary injections.	**99.50%**	95.70%
F1-Score	Harmonic mean of Precision and Recall. Balances performance on imbalanced datasets (e.g., fewer Wet AMD cases vs. Normal).	**0.991**	0.938
AUC-ROC	Area Under the Receiver Operating Characteristic Curve. Measures the model's ability to discriminate between disease classes at various thresholds.	**0.999**	0.965
IoU (Intersection over Union)	Measures the overlap between the AI-segmented fluid boundary and the expert manual segmentation. Critical for quantifying fluid volume.	**0.94**	0.82

5. CONCLUSION

This research establishes a new paradigm for Age-Related Macular Degeneration (AMD) management by addressing the critical diagnostic and therapeutic limitations inherent in current ophthalmic AI. The primary achievement of this contribution is the successful resolution of the "Black Box" bottleneck—the persistent failure of pure deep learning models to bridge the gap between high-fidelity biomarker quantification and guideline-concordant clinical reasoning [1, 9, 13].

By engineering a modular architecture that integrates a **ConvNeXt**-based diagnostic core with a Retrieval-Augmented Generation (RAG) agent, this study effectively neutralizes the risks of generative AI hallucination while maintaining a superior diagnostic accuracy of 98.9% [1, 10, 14]. This "Neuro-Symbolic" framework—combining deep perceptual precision with deterministic clinical logic—proves that the future of retinal care does not lie in universal diagnostic tools, but in specialized, safety-guarded "AI Co-pilots" capable of synthesizing patient-specific profiles with standardized clinical protocols [11, 15, 18].

Ultimately, this work overcomes the fragmentation of multimodal imaging and the subjectivity of diagnostic staging, providing a scalable, reliable path forward for preserving vision in an aging global population. While this study deliberately constrains its domain to AMD-specific management to ensure high-precision safety guardrails, it provides the fundamental architectural blueprint for the next generation of trustworthy, specialty-specific clinical decision support systems [12, 13, 16].

Despite these advancements, this work acknowledges several constraints that define the boundary of current knowledge. First, the model focuses exclusively on AMD; its diagnostic utility for confounding pan-retinal pathologies, such as glaucoma or diabetic retinopathy, remains unproven. Second, while our Safety Layer mitigates stochastic errors, this research does not resolve the complex legal and ethical liability frameworks inherent in autonomous AI-assisted prescribing. Finally, while the system is computationally efficient, prospective clinical trial validation across diverse, low-resource hardware settings is required before global deployment. These limitations serve as the roadmap for future research, which will focus on integrating multi-disease diagnostic modules and conducting real-world clinical validation.

As AMD prevalence escalates globally, this framework demonstrates that the future of ophthalmic AI relies on combining perceptual precision with deterministic, guardrailed reasoning to preserve patient vision.

REFERENCES

[1] Z. Liu, H. Mao, C. Y. Wu, C. Feichtenhofer, T. Darrell, and S. Xie, "A ConvNet for the 2020s," Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 11976–11986, 2022.

[2] Ma et al., "A multicenter study for detecting multiple retinal diseases using bimodal imaging of color fundus photography and optical coherence tomography," PMC12460420, 2025.

[3] V. Silvestri et al., "Multimodal imaging biomarkers for progression from intermediate to advanced age-related macular degeneration (AMD): a 10-year prospective longitudinal cohort study," BMJ Open Ophthalmology, vol. 10, no. 1, 2025.

[4] Zedadra, M. Y. Salah-Salah, O. Zedadra, and A. Guerrieri, "VisionTrack: Multi-Modal AI system for predicting multiple retinal diseases," Sensors, vol. 25, no. 14, 2025.

[5] Ensemble Architecture Authors, "Ensemble deep learning for dry AMD progression stages," PubMed 39444813, 2024.

[6] Chen et al., "Multimodal deep learning model combining OCT and infrared reflectance (IR) for the diagnosis of AMD," PMC11554086, 2023.

[7] N. Le et al., "ViT-AMD: A Vision Transformer-based model for AMD diagnosis," MDPI Sensors Background Review, 2023.

[8] K. Jin et al., "Automated diagnosis of age-related macular degeneration using multi-modal vertical plane feature fusion via deep learning," Medical Physics, vol. 49, no. 4, pp. 2324–2333, 2022.

[9] American Academy of Ophthalmology Retina/Vitreous Panel, "Age-Related Macular Degeneration Preferred Practice Pattern," *Ophthalmology*, vol. 131, no. 1, pp. P1-P60, 2024.

[10] S. Liu, A. B. McCoy, and A. Wright, "Improving large language model applications in biomedicine with retrieval-augmented generation: a systematic review, meta-analysis, and clinical development guidelines," Journal of the American Medical Informatics Association (JAMIA), vol. 32, no. 4, pp. 605–615, 2025.

[11] Ç. U. Öğdü et al., "Adaptive multi-agent LLM-based CDSS integrating biomedical RAG and web intelligence," IEEE Access, vol. 13, 2025.

[12] Wang et al., "Explainable Artificial Intelligence Framework for Predicting Treatment Outcomes in Age-Related Macular Degeneration," MDPI Sensors, vol. 25, no. 22, 2024.

[13] K. Workum et al., "Bridging the gap: a practical step-by-step approach to warrant safe implementation of large language models in healthcare," Frontiers in Artificial Intelligence, vol. 10.3389/frai.2025.1504805, 2025.

[14] S. Liu, A. B. McCoy, and A. Wright, "Improving large language model applications in biomedicine with retrieval-augmented generation: a systematic review, meta-analysis, and clinical development guidelines," Journal of the American Medical Informatics Association (JAMIA), vol. 32, no. 4, pp. 605–615, 2025.

[15] X. Chen, Z. Zhao, W. Zhang, et al., "EyeGPT: Ophthalmic Assistant with Large Language Models," *JMIR Medical Informatics*, vol. 12, p. e60063, 2024.

[16] G. Neri, C. Rebecchi, J. D. Oakley, et al., "Deep Learning Model for Automated Classification of Macular Neovascularization Subtypes in AMD," *Investigative Ophthalmology & Visual Science*, vol. 66, no. 9, p. 55, 2025.

[17] K. Roy, et al., "Multimodal Fusion Learning with Dual Attention for Medical Imaging," *IEEE Access*, vol. 12, pp. 38954-38965, 2024.

[18] S. Satheakeerthy, D. Jesudason, J. Pietris, et al., "LLM-assisted medical documentation: efficacy, errors, and ethical considerations in ophthalmology," *Eye*, vol. 39, pp. 1-8, 2025.