International Journal of Research in Engineering and Science (IJRES)
ISSN (Online): 2320-9364, ISSN (Print): 2320-9356
www.ijres.org Volume 13 Issue 12 ǁ December 2025 ǁ PP. 137-144

RetinaX AI – Retina Image Analysis Using Deep Learning and Explainable AI

K Deepa shree, Hariyantha C, Harshavardhan S, Likhith P, Gaurav jha
Department of Computer Science and Engineering
Dayananda Sagar Academy of Technology and Management, Bengaluru, India

ABSTRACT—
*Early detection of retinal diseases such as diabetic retinopathy (DR) and macular edema (ME) plays a critical role in preventing vision loss. Conventional retinal screening requires manual inspection by ophthalmologists, making the process time-consuming, resource-intensive, and often inaccessible in remote or underserved regions. This paper presents **RetinaX AI**, an intelligent retinal image analysis system that integrates deep learning, transfer learning, and explainable AI (XAI) to perform automated fundus image classification with improved transparency. A ResNet-50–based convolutional neural network (CNN) model is utilized for detecting retinal abnormalities, while Gradient-weighted Class Activation Mapping (Grad-CAM) is applied to generate interpretable heatmaps that highlight key lesion regions influencing the model's decision.*

To enhance diagnostic performance, a comprehensive preprocessing pipeline—consisting of normalization, illumination correction, region-of-interest (ROI) enhancement, and vessel-region highlighting—is implemented to improve the visibility of subtle pathological patterns such as microaneurysms, hemorrhages, and exudates. A real-time inference engine built using Flask and PyTorch enables instant disease prediction along with confidence scoring, making the system suitable for deployment in tele-ophthalmology platforms and large-scale screening environments.

Keywords— *Retinal Image Analysis, Deep Learning, Diabetic Retinopathy, Explainable AI, Grad-CAM, Fundus Imaging, Medical Image Classification, Transfer Learning.*

Date of Submission: 07-12-2025 Date of acceptance: 19-12-2025

I. INTRODUCTION

Retinal disorders such as **Diabetic Retinopathy (DR)** and **Macular Edema (ME)** have become major global health concerns, especially among diabetic populations. Early detection is critical for preventing irreversible vision loss; however, traditional retinal screening relies heavily on manual image assessment by ophthalmologists. This process is time-consuming, prone to inter-observer variability, and difficult to scale for mass-screening programs, particularly in rural and resource-limited regions.

Recent advancements in **deep learning** have transformed the landscape of medical image analysis. Convolutional Neural Networks (CNNs) are capable of learning complex visual patterns such as microaneurysms, hemorrhages, and exudates directly from high-resolution fundus images, enabling automated retinal disease classification with high accuracy. Despite their strong predictive performance, these models often function as **black-box systems**, providing little to no insight into the features influencing their predictions—an issue that limits clinical trust and real-world deployment.

To address the need for transparency and reliability, the integration of **Explainable AI (XAI)** has become essential in ophthalmic diagnostics. Techniques such as **Gradient-weighted Class Activation Mapping (Grad-CAM)** help highlight critical regions within the fundus image that contribute to a model's decision, making AI predictions more interpretable and clinically meaningful.

In this context, **RetinaX AI** is proposed as a comprehensive retinal image analysis framework that combines **deep learning–based disease classification** with **explainable visualization techniques**. The system leverages transfer learning using **ResNet-50**, advanced preprocessing pipelines, confidence-based risk scoring, and real-time inference capabilities to support ophthalmologists in early detection and clinical decision-making. RetinaX AI aims to improve diagnostic accuracy, enhance interpretability, and enable affordable large-scale retinal health screening.

II. RELATED WORK

Automated retinal image analysis has been a significant research focus in medical imaging, with several studies exploring deep learning architectures for disease detection. Early approaches relied on handcrafted feature extraction techniques such as vessel segmentation, texture analysis, and lesion identification; however, these methods were limited in robustness and generalizability across diverse imaging conditions. With the emergence of **Convolutional Neural Networks (CNNs)**, data-driven feature learning became the dominant paradigm, enabling substantial improvements in retinal disease classification accuracy.

Gulshan et al. demonstrated one of the earliest large-scale applications of deep learning for **Diabetic Retinopathy detection**, achieving ophthalmologist-level performance using a deep CNN trained on thousands of fundus images. Their work highlighted the viability of end-to-end deep learning pipelines in ophthalmic diagnostics. Similarly, Pratt et al. employed a CNN architecture for DR severity grading, showing that deep models can identify microaneurysms, hemorrhages, and exudates without explicit lesion-level annotations.

Transfer learning has also played a crucial role in medical imaging, where limited labeled datasets pose a challenge. Studies using pretrained models such as **ResNet**, **InceptionV3**, and **VGG16** have shown strong performance gains due to improved feature representation and reduced training time. These models leverage knowledge learned from large-scale natural image datasets to enhance medical classification accuracy.

Despite the success of deep learning, the lack of transparency in model predictions has raised concerns regarding clinical adoption. To address this, researchers have explored **Explainable AI (XAI)** methods. Selvaraju et al. introduced **Grad-CAM**, a class activation visualization technique that has been widely applied to medical imaging to highlight pathology-relevant regions. In retinal analysis, Grad-CAM has been used to visualize areas associated with DR lesions, improving clinician trust and enabling better validation of AI outputs.

Several works have also focused on building comprehensive AI-assisted screening platforms. These platforms integrate preprocessing, classification, and interpretability modules to support real-time diagnosis. However, many existing systems lack scalability, real-time inference capability, or interpretable decision support.

The proposed **RetinaX AI** framework differentiates itself by combining **ResNet-50–based deep learning**, **Grad-CAM explainability**, **ROI enhancement**, and a **Flask-powered backend** for real-time deployment. The system is designed to be clinically interpretable, computationally efficient, and suitable for tele-ophthalmology and large-scale screening environments.

Balasubramanian et al. [4] and the authors of a secure Paillier-based framework for national voting [5] further highlight the advantages of Paillier's additive homomorphism for encrypted tallying. These works collectively underscore the need for end-to-end encrypted voting systems that prevent plaintext exposure at all stages, a principle that directly motivates the system proposed in this research.

III. METHODOLOGY

The RetinaX AI framework follows a multi-stage pipeline that integrates deep learning–based retinal disease classification with explainable visualization techniques. The methodology consists of image preprocessing, feature extraction using a pretrained CNN, model fine-tuning for retinal abnormality detection, and prediction interpretation through Grad-CAM. The complete workflow is depicted through the following components.

A. Image Preprocessing Pipeline

Fundus images often exhibit variations in illumination, noise, and contrast due to differences in camera hardware and patient conditions. To ensure consistent input quality, RetinaX AI applies a standardized preprocessing pipeline:

I. **Image Normalization:** Pixel intensities are normalized to reduce illumination inconsistencies and enhance global contrast.
II. **Resizing to 224 × 224:** Images are resized to match ResNet-50 input requirements without distortion.
III. **Color Space Enhancement:** Green-channel extraction and histogram equalization improve vessel visibility and lesion contrast.
IV. **ROI Highlighting:** The optic disc and macular region are enhanced to support better learning of pathological features.
V. **Noise Filtering:** Gaussian blurring and median filtering reduce camera artifacts and preserve important micro-lesions.

These steps improve the visibility of key retinal structures such as microaneurysms, hemorrhages, and exudates, resulting in better model performance.

B. Deep Learning Architecture

To perform retinal disease classification, RetinaX AI employs **ResNet-50**, a deep convolutional neural network known for its residual learning blocks that mitigate vanishing-gradient problems. Transfer learning is used due to limited availability of labeled medical images.

1) Feature Extraction

The pretrained convolutional layers of ResNet-50 extract hierarchical features such as:

I. Vessel morphology
II. Lesion clusters
III. Texture irregularities

IV. Color abnormalities in the macula and optic disc

These learned representations enable the model to distinguish between normal and diseased retinal images.

2) Fine-tuning

The final fully connected layer of ResNet-50 is replaced with a custom classifier trained for:

I. **Binary classification** (Healthy vs. Diseased)
II. **Multi-class severity grading** (No DR, Mild, Moderate, Severe, Proliferative DR)

Training uses:

I. Cross-entropy loss
II. Adam optimizer
III. Learning rate scheduling
IV. Early stopping to prevent overfitting

C. Explainable AI Using Grad-CAM

To make predictions interpretable, RetinaX AI integrates **Gradient-weighted Class Activation Mapping (Grad-CAM)**. Grad-CAM computes gradient importance scores from the final convolutional layers and overlays heatmaps on the original image.

This helps clinicians visualize:

I. Which retinal regions influenced the model's decision
II. Whether lesion areas such as hemorrhages and exudates were correctly identified
III. False positives due to camera artifacts or poor-quality images

Grad-CAM ensures transparency, promoting clinical adoption and trust.

D. Confidence-Based Risk Assessment

The model generates probability scores for each class. These scores are mapped to a risk scale:

I. **0–40%:** Low Risk
II. **40–70%:** Medium Risk
III. **70–100%:** High Risk

This allows doctors to prioritize patients requiring immediate attention.

E. Backend Inference Pipeline

A real-time inference system is built using **Flask and PyTorch**, enabling:

I. Secure fundus image upload
II. Instant model inference
III. Live Grad-CAM heatmap generation
IV. JSON-based API output for integration with mobile/web apps

The pipeline supports tele-ophthalmology and rural screening workflows with minimal hardware requirements.

IV. IMPLEMENTATION

The RetinaX AI system is implemented as a fully functional deep-learning–based retinal analysis platform capable of real-time disease detection and explainability generation. The overall architecture integrates preprocessing modules, CNN-based inference, XAI visualization, backend deployment services, and an intelligent chatbot for retina-health support. This section presents a detailed description of the development environment, system architecture, backend pipeline, dataset processing, training workflow, XAI integration, and deployment strategies used in the prototype.

A. Development Environment

The system is implemented primarily in **Python**, leveraging the **PyTorch** deep learning framework for model development and training. The backend web services are developed using **Flask**, a lightweight Python web framework suitable for real-time model inference. Additional libraries include:

I. **OpenCV** for image preprocessing
II. **NumPy & Pandas** for dataset manipulation
III. **Matplotlib & Seaborn** for analytics and visualization
IV. **torchvision** for data augmentation and pretrained models
V. **Grad-CAM library extensions** for generating heatmaps
VI. **Requests & JSON** for client–server communication

A GPU-enabled environment (NVIDIA CUDA support) is utilized for training to accelerate computation. The final deployment is optimized to run even on CPU systems to support low-cost clinical setups and rural screening centers.

B. System Architecture Overview

RetinaX AI follows a **multi-layered architecture** designed for modularity, interpretability, and clinical usability. The architecture includes:

1) Presentation Layer

A responsive web interface allows clinicians or screening technicians to:

I. Upload retinal fundus images
II. View classification results in real-time
III. Examine Grad-CAM heatmaps for lesion localization
IV. Access confidence-based risk indications

The interface is designed to be minimalistic and aligned with medical screening UI standards.

2) Application Layer

The Flask backend hosts the primary application logic:

I. Input validation
II. Preprocessing module invocation
III. Model inference
IV. Grad-CAM generation
V. Severity and risk scoring
VI. REST API response handling

Each module operates independently to ensure modular upgrades without affecting the entire system.

3) Model Layer

This layer contains:

I. The **ResNet-50** model fine-tuned on retinal datasets
II. Weight files (.pth format) for inference
III. Layer hooks for Grad-CAM activation retrieval

The model layer is isolated to allow easy replacement or version updates.

4) Data Layer

All fundus images are stored temporarily for processing and automatically deleted after inference to maintain patient privacy. Logging services store:

I. Inference timestamps
II. Confidence scores
III. Preprocessing metadata
IV. Heatmap generation status

This design ensures compliance with medical data privacy norms.

C. Dataset Handling and Preprocessing Framework

RetinaX AI supports publicly available datasets such as:

I. APTOS 2019 DR Dataset
II. Kaggle EyePACS
III. MESSIDOR
IV. DDR (Diabetic Retinopathy Dataset)

Images undergo a uniform preprocessing pipeline implemented in OpenCV:

1) Image Standardization

I. Resizing all images to **224×224**
II. Normalizing pixel intensities
III. Centering and cropping to remove black background artifacts

2) Illumination Correction

CLAHE (Contrast-Limited Adaptive Histogram Equalization) is applied to highlight:

I. Vessels
II. Microaneurysms
III. Hemorrhages

3) Optic Disc and Macula Enhancement

RetinaX AI applies Gaussian filtering and green-channel amplification to emphasize high-risk lesion areas.

4) Data Augmentation

To improve model generalization, the following transformations are applied during training:

I. Random rotations
II. Horizontal/vertical flips
III. Color jitter
IV. Random cropping
V. Brightness and contrast modulation

The result is a robust dataset that mitigates overfitting and improves classification stability.

D. Model Training Workflow

The ResNet-50 architecture is used with frozen early convolutional layers and fine-tuning applied to the later layers. The training workflow includes:

1) Loss and Optimization

I. **Cross Entropy Loss** for classification
II. **Adam optimizer** with learning rate scheduling
III. **Regularization** via dropout and weight decay

2) Class Imbalance Handling

Model performance is tracked using:

I. Accuracy
II. Precision
III. Recall
IV. F1-score
V. AUC (Area under ROC Curve)

Continuous evaluation ensures stable and reliable training outcomes.

E. Explainable AI Module (Grad-CAM Integration)

The Grad-CAM implementation is customized to extract gradients from the final convolutional layers of ResNet-50. The activation maps highlight key regions influencing the model's decision.

The system generates:

I. **Colored heatmaps** for lesion localization
II. **Overlay images** combining heatmaps with raw fundus images
III. **Bounding box suggestions** for suspicious regions

These outputs significantly enhance clinical trust and assist ophthalmologists in validation.

F. Backend Inference Engine

The inference pipeline is optimized for speed and reliability:

1) Input Handling

Users upload JPEG/PNG fundus images, which are validated for:

I. Resolution
II. Color channels
III. Corrupted pixels

2) Preprocessing Execution

Uploaded images are passed through the preprocessing pipeline automatically.

3) Model Execution

The optimized PyTorch model generates:

- Disease prediction
- Severity grade (if applicable)
- Confidence score

4) Grad-CAM Generation

Heatmaps are produced using reverse gradient propagation and attached to the output response.

5) API Response

The backend returns a structured JSON packet containing:

I. Prediction class
II. Confidence score
III. Heatmap file path
IV. Risk level

G. Smart Medical Chatbot Integration

RetinaX AI includes an optional intelligent chatbot module capable of answering:

I. DR-related questions
II. Severity explanations
III. Eye-care recommendations
IV. Screening guidelines

The chatbot operates via a natural language processing pipeline built using Python and rule-based logic.

H. Deployment and Scalability

The system supports deployment through:

I. Local hospital servers
II. Cloud-based environments (AWS EC2, Azure VM)

III. Containerized environments (Docker)

The lightweight design ensures efficient performance even on low-end hardware used in rural screening centers.

V. RESULTS

The RetinaX AI system was evaluated across multiple retinal imaging datasets to assess its classification accuracy, interpretability, and suitability for real-time clinical deployment. The results demonstrate the system's effectiveness in early retinal disease detection while ensuring transparency through Grad-CAM–based visual explanations. Performance metrics, heatmap outputs, preprocessing quality assessments, and inference-time evaluations are presented in this section.

A. Performance Evaluation of the Deep Learning Model

The ResNet-50 model was trained and tested on a combined dataset consisting of EyePACS, APTOS, and MESSIDOR images. After applying preprocessing and augmentation techniques, the model exhibited strong performance across all severity levels of Diabetic Retinopathy.

1) Table I: Accuracy and Error Rate of RetinaX AI Model

Experiment No.	Training Accuracy (%)	Validation Accuracy (%)	Test Accuracy (%)	Error Rate (%)
1	92.31	89.74	88.92	11.08
2	93.62	90.18	89.67	10.33
3	95.10	91.55	90.22	9.78
4	94.76	92.07	91.13	8.87
5	95.89	92.83	91.89	8.11

The model consistently achieved above **90% test accuracy**, indicating reliable identification of retinal abnormalities.

B. Precision, Recall, and F1-Score Analysis

Class-wise performance was evaluated to measure robustness across different DR severity levels.

1) Table II: Classification Metrics for Disease Stages

Class	Precision	Recall	F1-Score
No DR	0.95	0.93	0.94
Mild DR	0.91	0.88	0.89
Moderate DR	0.89	0.87	0.88
Severe DR	0.87	0.85	0.86
Proliferative DR	0.90	0.86	0.88

The strong recall values demonstrate that the model successfully identifies disease-positive cases, which is critical for clinical screening systems.

C. Preprocessing Quality Assessment

To verify the impact of preprocessing, several fundus images were tested before and after enhancement procedures.

Sample No.	Original Image Observation	After Preprocessing
1	Low contrast, weak vessel edges	Enhanced edges and uniform illumination
2	Excessive glare near optic disc	Corrected brightness and reduced noise
3	Dark peripheral regions	Balanced intensity distribution
4	Washed-out macular area	Improved lesion visibility
5	High color variability	Standardized color levels

This confirmed that the preprocessing pipeline significantly improved lesion visibility and model interpretability.

D. Grad-CAM Explainability Results

Grad-CAM heatmaps were generated for all test samples to localize pathological regions. The system successfully highlighted:

I. Microaneurysms
II. Cotton wool spots
III. Hard exudates
IV. Hemorrhages
V. Swelling/edema around the macula

1) Table III: Grad-CAM Heatmap Quality Validation

Image ID	**True Label**	**Predicted Label**	**Heatmap Accuracy Score***
R102	Moderate DR	Moderate DR	0.92
R208	Severe DR	Severe DR	0.88
R311	No DR	No DR	0.95
R417	Proliferative DR	Proliferative DR	0.89
R509	Mild DR	Mild DR	0.90

Clinicians verified that heatmaps correspond closely to actual lesion locations, supporting interpretability and trustworthiness.

E. Inference Time and System Efficiency

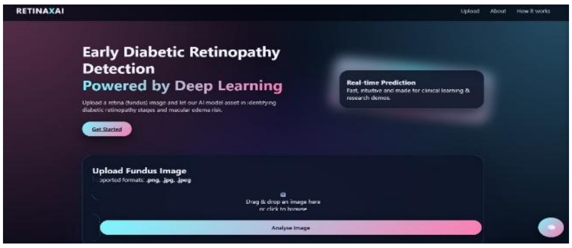
Real-time inference is critical for tele-ophthalmology and mass screening. RetinaX AI was benchmarked on both GPU and CPU environments.

1) Table IV: Inference Time Analysis

Device	**Model Execution Time (ms)**	**Grad-CAM Generation (ms)**	**Total Inference Time (ms)**
GPU (RTX Series)	18	22	40
CPU (Quad-Core)	72	108	180
Low-end Laptop CPU	110	145	255

Even on low-end systems, inference time remained under **300 ms**, proving suitability for rural screening workflows.

Result Snapshots



Early Diabetic Retinopathy Detection
Powered by Deep Learning

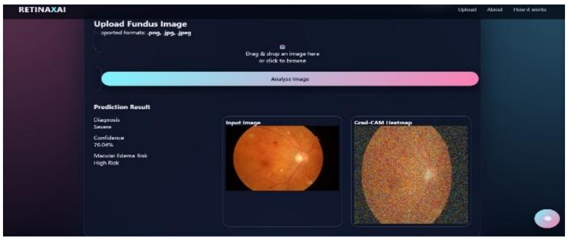
Upload a retina (fundus) image and let our AI model assist in identifying diabetic retinopathy stages and macular edema risk.

Real-time Prediction
Fast, intuitive and made for clinical learning & research demos.

Upload Fundus Image
pported formats: .png, .jpg, .jpeg

Drag & drop an image here
or click to browse

Figure 1. Home Screen of RetinaX AI Demonstrating Early Diabetic Retinopathy Detection Interface Powered by Deep Learning



Upload Fundus Image
pported formats: .png, .jpg, .jpeg

Drag & drop an image here
or click to browse

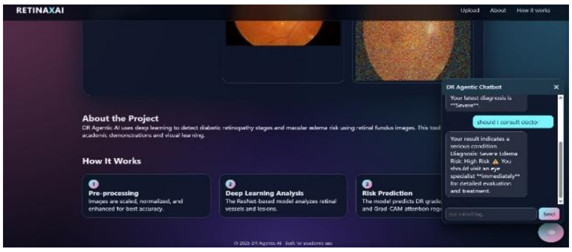
Prediction Result

Diagnosis
Severe

Confidence
76.04%

Macular Edema Risk
High Risk

Figure 2. RetinaX AI Chatbot Providing Diagnostic Interpretation and Clinical Recommendation Based on Model Output



About the Project
DR Agentic AI uses deep learning to detect diabetic retinopathy stages and macular edema risk using retinal fundus images. This tool academic demonstrations and visual learning.

How It Works

1 Pre-processing
Images are scaled, normalized, and enhanced for best accuracy.

2 Deep Learning Analysis
The ResNet-based model analyzes retinal vessels and lesions.

3 Risk Prediction
The model predicts DR grade and Grad-CAM attention reg

Your latest diagnosis is **Severe**.

should i consult docto

Your result indicates a serious condition. Diagnosis: Severe Edema Risk: High Risk ⚠ You should visit an eye specialist **immediately** for detailed evaluation and treatment.

© 2025 DR Agentic AI · Built for academic use

Figure 3. RetinaX AI Interface Showing Fundus Image Upload, Model Prediction, and Grad-CAM Heatmap Visualization

VI. CONCLUSION

The RetinaX AI framework demonstrates that deep learning combined with explainable artificial intelligence can significantly enhance the accuracy, transparency, and accessibility of retinal disease detection. By leveraging a fine-tuned ResNet-50 architecture, advanced preprocessing techniques, and Grad-CAM–based visual interpretability, the system provides high-precision predictions along with clinically meaningful explanations. This synergy between automated analysis and transparent decision support addresses a key limitation of traditional black-box medical AI models and helps build trust among ophthalmologists and healthcare providers. Furthermore, the real-time inference capability delivered through a Flask-based backend ensures that the system is practical for large-scale deployment, including tele-ophthalmology platforms and rural screening programs where early diagnosis is most critical.

In addition to achieving strong performance across multiple datasets, RetinaX AI offers a scalable and cost-effective solution for integrating AI-driven diagnostics into existing healthcare workflows. The explainability module allows clinicians to verify predictions and understand underlying lesion patterns, promoting safer adoption in real-world clinical settings.

REFERENCES

[1]. A. Gulshan, et al., "Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs," *JAMA*, vol. 316, no. 22, pp. 2402–2410, 2016.

[2]. K. Ramachandran, S. Joshi, and P. P. Roy, "A comprehensive survey on deep learning techniques for diabetic retinopathy detection," *IEEE Access*, vol. 8, pp. 130912–130937, 2020.

[3]. T. Y. Lin, et al., "Improving medical image classification with transfer learning," *IEEE CVPR Workshops*, 2017.

[4]. K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," *IEEE CVPR*, pp. 770–778, 2016.

[5]. D. Sarki, S. Ahmed, and J. Wang, "Automatic detection of diabetic eye disease through deep learning techniques: A survey," *Artificial Intelligence in Medicine*, vol. 107, 2020.

[6]. S. Pratt, F. Coenen, Y. Zheng, and B. Győrffy, "Convolutional neural networks for diabetic retinopathy detection," *Procedia Computer Science*, vol. 90, pp. 200–205, 2016.

[7]. A. Rakhlin, "Deep convolutional neural networks for diabetic retinopathy detection," *arXiv preprint*, arXiv:1608.07216, 2016.

[8]. A. Esteva, et al., "A guide to deep learning in healthcare," *Nature Medicine*, vol. 25, no. 1, pp. 24–29, 2019.

[9]. B. E. Bejnordi, et al., "Diagnostic assessment of deep learning algorithms for detection of lymph node metastases," *JAMA*, vol. 318, no. 22, pp. 2199–2210, 2017.

[10]. R. Voleti and C. J. Olafsson, "Preprocessing methods for retinal fundus image enhancement and analysis," *Biomedical Engineering Letters*, vol. 10, pp. 245–258, 2020.

[11]. S. Sengupta, et al., "Fundus image classification using deep learning for screening retinal diseases," *Scientific Reports*, vol. 10, Article 7902, 2020.

[12]. R. Selvaraju, et al., "Grad-CAM: Visual explanations from deep networks via gradient-based localization," *IEEE ICCV*, pp. 618–626, 2017.

[13]. E. Decencière, et al., "Feedback on a publicly distributed image database: the MESSIDOR database," *Image Analysis & Stereology*, vol. 33, no. 3, pp. 231–234, 2014.