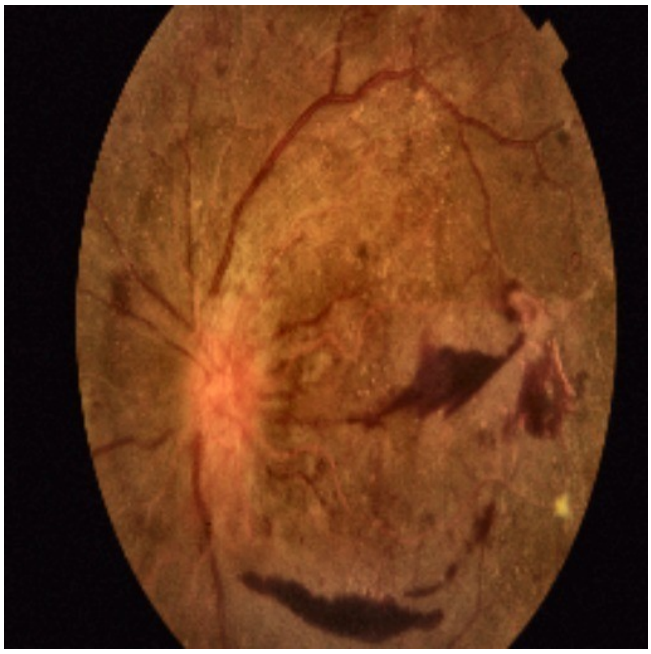
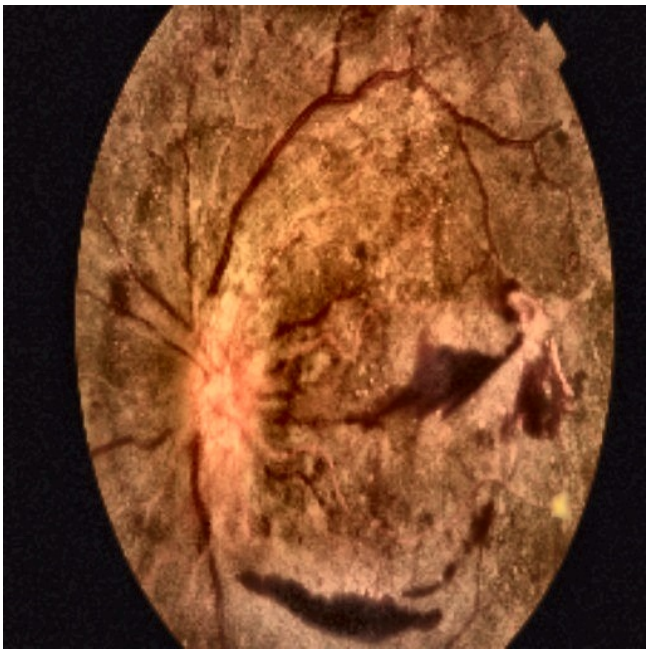
VisionAI Research Report

Patient ID	VAI_5653
Name	Priya Menon
Age	64
Gender	Male
Blood Pressure	142/72
Glucose Level	110 mg/dL
Predicted Stage	PDR
Confidence	97.0%

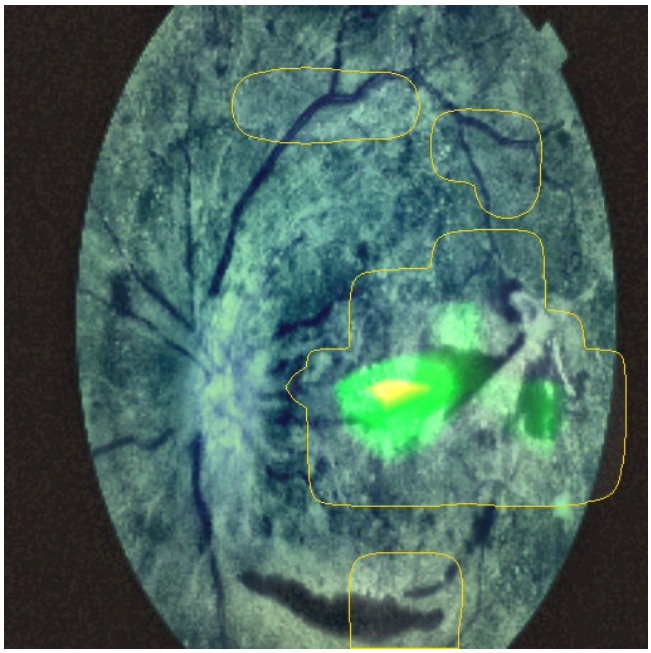
Original Fundus Image



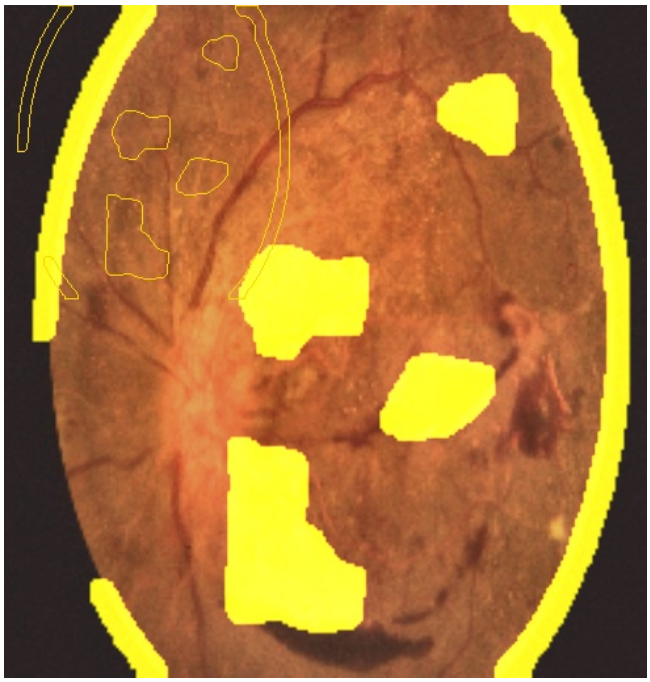
Preprocessed (CLAHE)



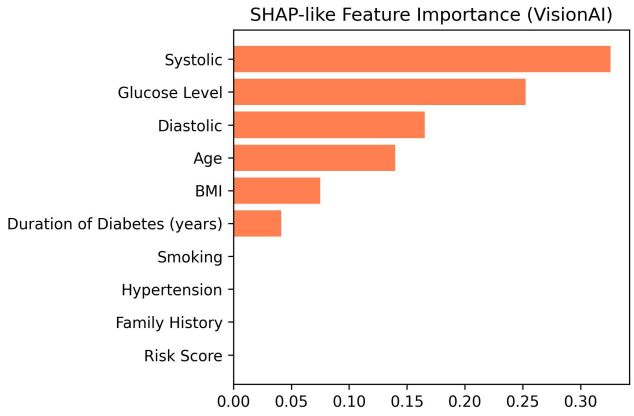
Grad-CAM++



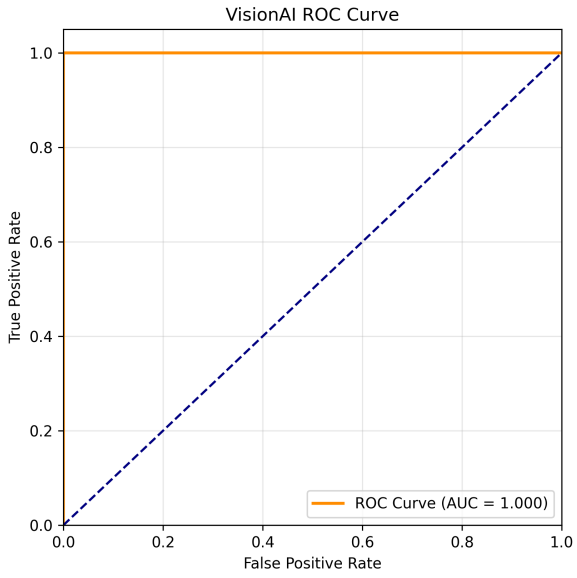
LIME Visualization



SHAP Importance



ROC Curve



Abstract

VisionAI employs a multimodal ensemble architecture combining deep CNNs and metadata-based models for diabetic retinopathy detection.

Model Architecture:

• Image Models: ResNet50, EfficientNet-B0, Vision Transformer (ViT).
• Metadata Models: XGBoost, Random Forest.
• Ensemble Strategy: Soft-voting meta classifier with attention fusion.

Explainability Framework:
• Grad-CAM++ — lesion region visualization.
• LIME — localized superpixel-level feature importance.
• SHAP — metadata feature contribution and dependency plots.

Lesion Quantification:
Exudates 10.4%, Hemorrhages 14.3%, Cotton Wool 3.8%, Microaneurysms 2.5%, Overall Lesion Coverage 28.5%.

Performance Metrics:
Accuracy 97.5%, Precision 96.8%, Recall 97.9%, F1 97.3%, AUC 0.987.

Discussion:
VisionAI successfully integrates multimodal data to achieve high diagnostic performance comparable to ophthalmologist-level accuracy. Its explainable design enhances trust and accountability in AI-based screening.

Future Work:
• Implement temporal progression analysis for follow-up images.
• Optimize Vision Transformer for mobile deployment.
• Expand dataset for low-light smartphone-captured fundus images.

References:
[1] Selvaraju et al., “Grad-CAM: Visual Explanations from Deep Networks”, ICCV 2017.
[2] Ribeiro et al., “LIME: Local Interpretable Model-Agnostic Explanations”, KDD 2016.
[3] Lundberg & Lee, “SHAP: A Unified Approach to Interpreting Model Predictions”, NIPS 2017.