

An Efficient Hybrid Framework Combining Transformer Feature Encoding and Machine Learning for Diabetic Retinopathy Classification



Prepared By:

**Asraa Alwattar
Rahaf Barakat**



**Supervised By:
Dr. Wessam Ahmad**

College of Informatics,
Midocean University, UAE

Master of Science in Artificial Intelligence

Midocean University | College of Informatics

Thesis Defense Presentation

Presentation Outline

- 01** — Introduction and Background
- 02** — Contributions
- 03** — Literature Review
- 04** — Methodology
- 05** — Results and Analysis
- 06** — Conclusion and Future Work
- 07** — References

Introduction and Background



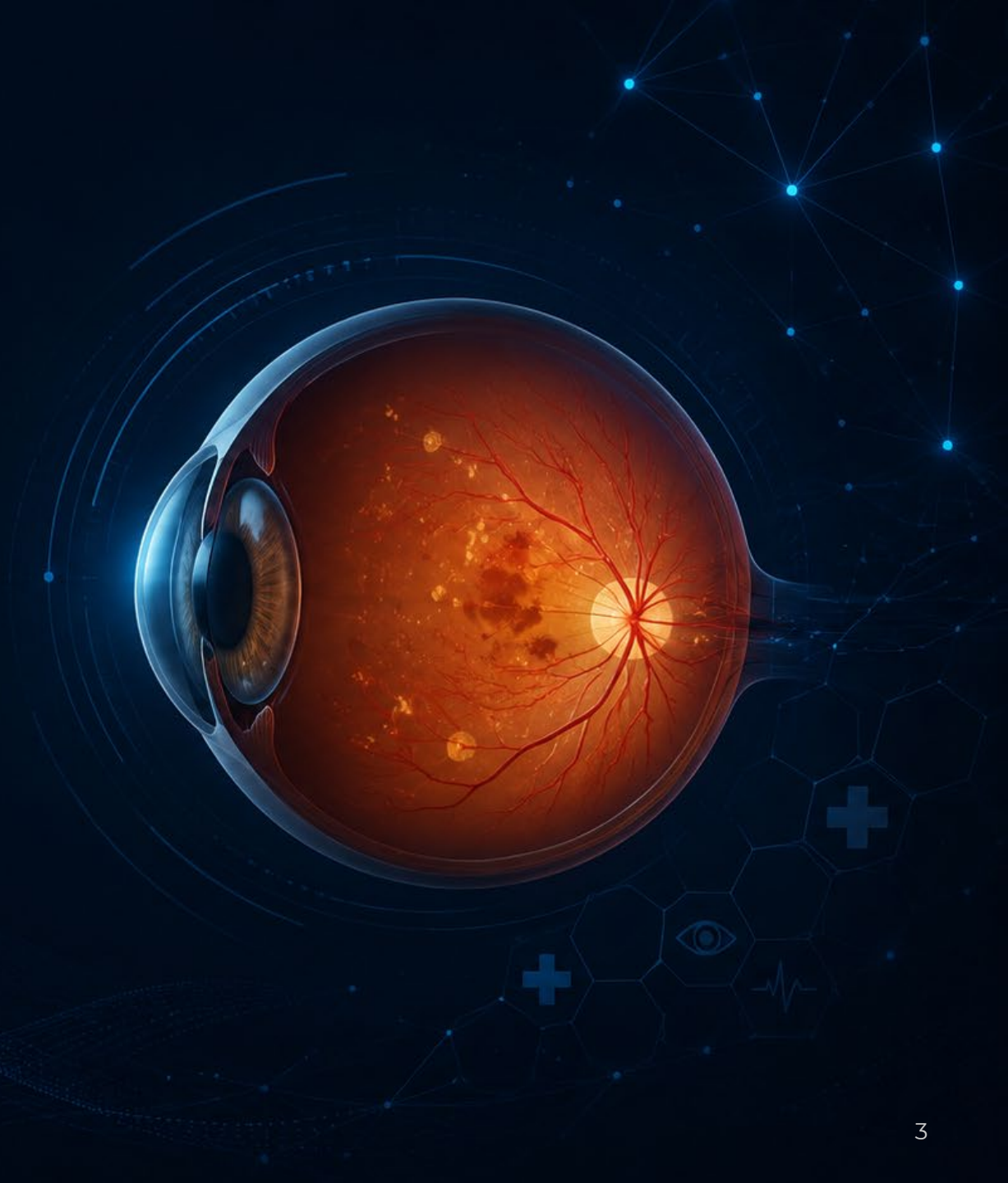
Diabetic retinopathy is a progressive microvascular complication of diabetes and a major cause of preventable visual impairment.



Manual grading of retinal fundus images is time-consuming, reader-dependent, and difficult to scale for population-level screening.



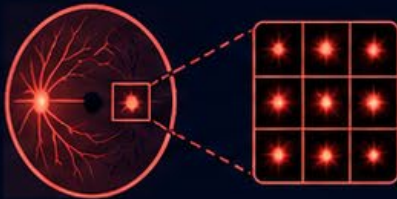
AI-based retinal-image analysis enables automated learning of disease-related patterns and supports more efficient severity grading.



Problem, Objective, and Significance



Automated five-class DR grading is challenged by image-quality variation, class imbalance, and subtle differences between neighboring disease stages.



Severe and proliferative classes are underrepresented, while local lesion patterns And broader retinal context must both be represented.



Enhance a reproducible hybrid framework that fuses ConvNeXt-Tiny and Swin-V2-Tiny embeddings and classifies them using Random Forest.

Key Research Contributions

1

Hybrid Visual Representation

Fuses ConvNeXt -Tiny and Swin -V2-Tiny features to combine fine-grained local lesion patterns with global contextual relationships.

2

Regularized Final Classifier

Implements Random Forest as the multiclass classifier and selects its configuration without using the testing subset for model selection.

3

Class-Sensitive Evaluation

Evaluates the hybrid classifier using per-class and aggregate precision, recall, F1-score, specificity, confusion matrix, and ROC-AUC.

Comparative Analysis of Related Work

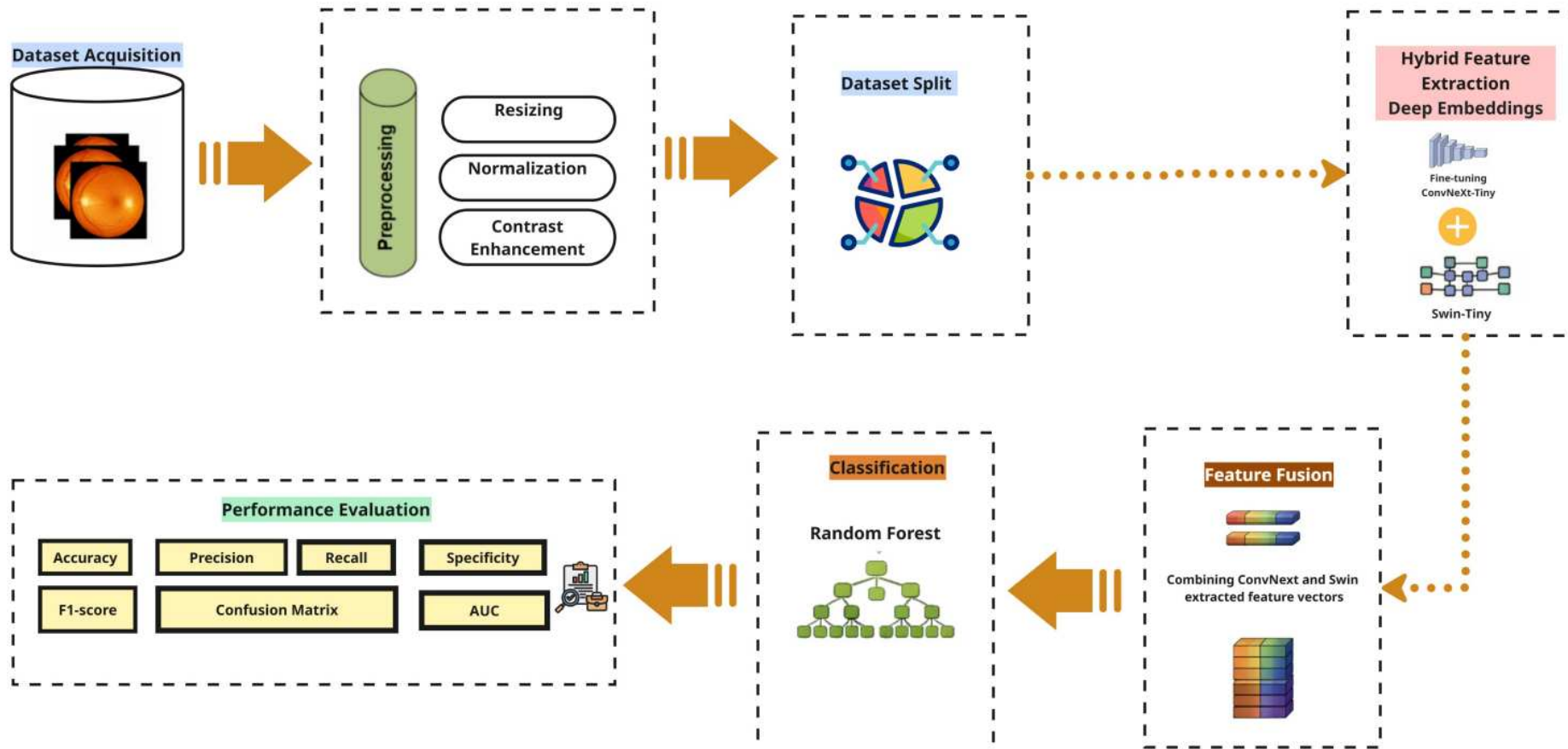


Authors	Year	Method / Model	Dataset	Results	Limitations
Lin and Wu [6]	2023	Revised ResNet50	Kaggle DR images	74.32% accuracy	Remaining train-test gap
Das et al. [7]	2023	DRFEC: 26 pretrained CNNs	EyePACS	79.11% accuracy	Overfitting; high train-validation gap
Jiwani et al. [8]	2022	CNN classifier	IDRiD	~77.6% accuracy	Limited design detail
Mehboob et al. [9]	2022	CNN-based approach + augmentation	EyePACS	~77.6% accuracy	Single dataset; limited generalization

Identified Gaps and Our Solutions

 Gap Category	 Current Limitations	 Proposed Solution
 Representation	Single-backbone approaches may emphasize either local or contextual features.	Fuse ConvNeXt-Tiny and Swin-V2-Tiny embeddings into a 1,536-dimensional hybrid vector.
 Imbalanced Multiclass Evaluation	Class imbalance may bias the model toward majority severity classes.	Apply class-weighted loss to give minority classes higher training importance.
 Model Selection	Some studies report strong within-dataset results without strict testing separation.	Use validation-based RF selection and reserve the testing subset for final reporting.
 Reproducibility	Complex pipelines can be difficult to rebuild after training.	Save fixed embeddings and use a controlled preprocessing and evaluation pipeline.

Proposed Framework Architecture



Proposed hybrid framework for diabetic retinopathy classification

Dataset Description

APTOS 2019 Blindness Detection Dataset [1]



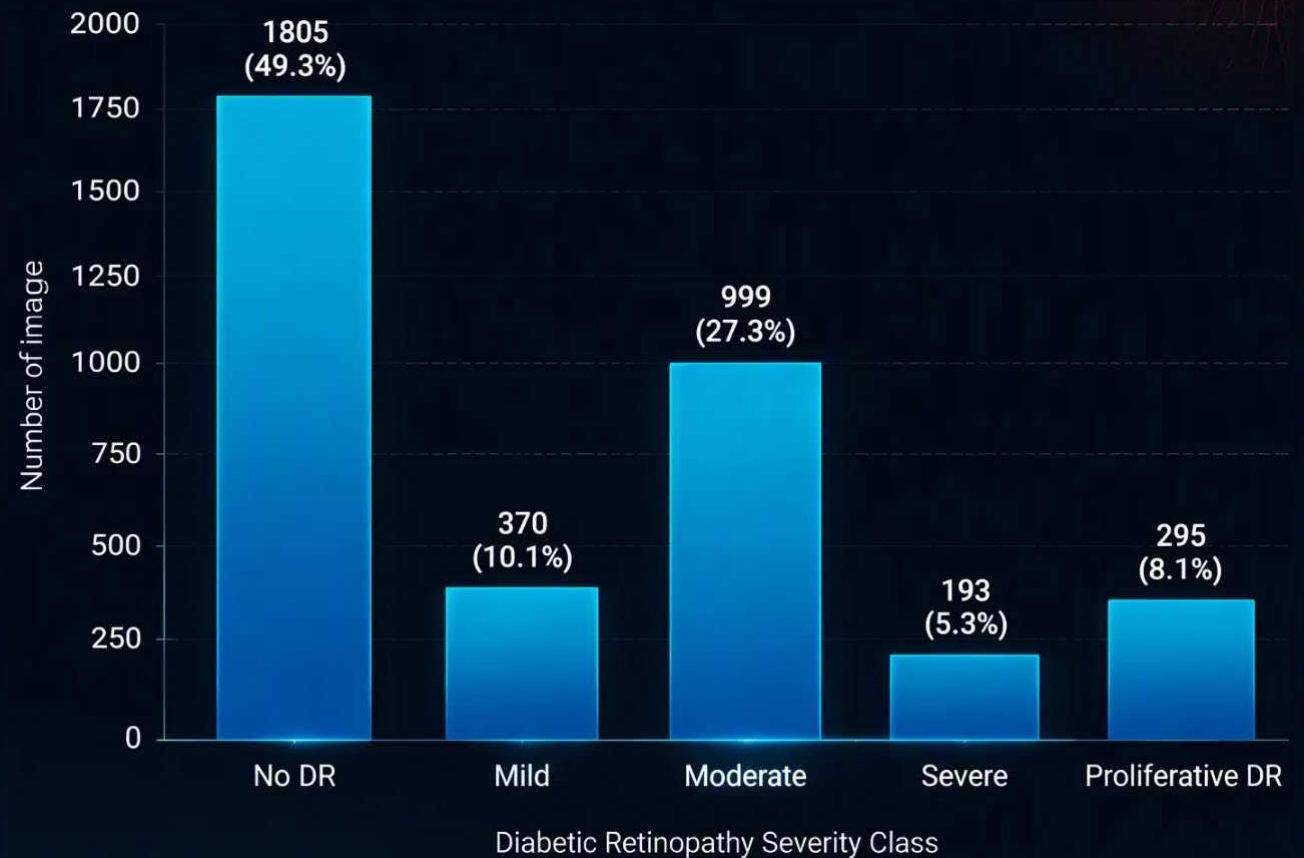
3,662 retinal fundus images.

Five clinical severity classes: No DR, Mild, Moderate, Severe, and Proliferative DR.

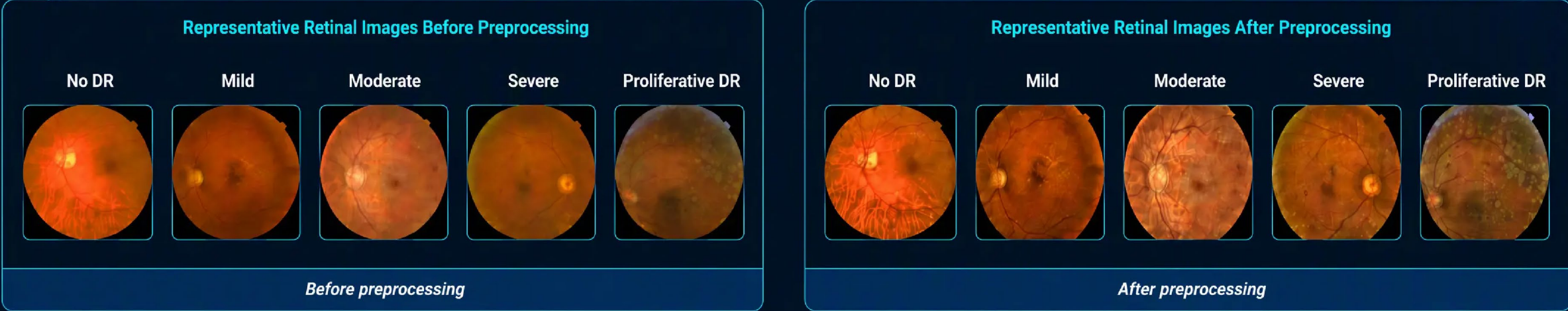
Class distribution is highly imbalanced, with No DR as the majority class and Severe as the smallest class.



Original Class Distribution in the APTOS 2019 Dataset



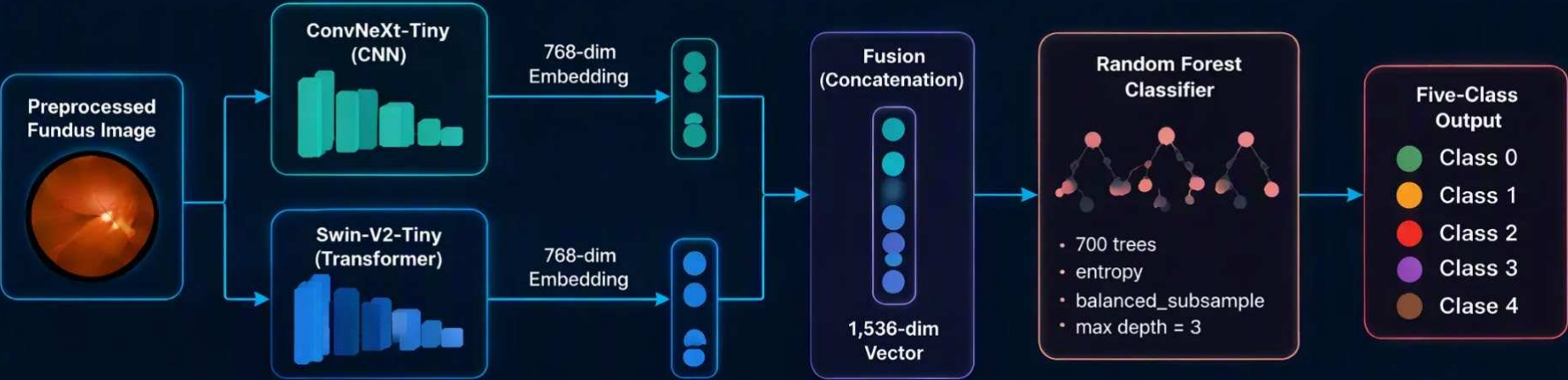
Original class distribution in APTOS 2019



Stage	Applied Procedure
 Integrity validation	Confirmed image-label consistency and removed missing or corrupted entries.
 Preprocessing	Dark-border cropping, CLAHE contrast enhancement, resizing to 300 × 300, tensor conversion, and ImageNet normalization.
 Training-only augmentation	Horizontal flip, rotation, brightness/contrast adjustment, saturation jitter, and random erasing only on training data.

Base Models Architecture

 Component	 Role	 Output / Setting
 ConvNeXt-Tiny	Convolutional feature extractor for local lesion texture, vessels, and hierarchical retinal patterns.	 768-dimensional embedding; best checkpoint at epoch 15
 Swin-V2-Tiny	Transformer-based feature extractor for contextual spatial relationships using shifted-window attention.	 768-dimensional embedding; best checkpoint at epoch 10
 Random Forest	Final five-class classifier trained on the fused 1,536-dimensional vector.	 700 trees; entropy; balanced_subsample; max depth = 3



Feature-level fusion and Random Forest classification

Evaluation Metrics

Comprehensive Performance Assessment

Classification Metrics:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

$$\text{Precision} = \frac{TP}{TP + FP}$$

$$\text{Recall} = \frac{TP}{TP + FN}$$

$$F_1 = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

$$\text{Specificity} = \frac{TN}{TN + FP}$$

Where: TP = True Positives

TN = True Negatives

FP = False Positives

FN = False Negatives

Confusion matrices and one-vs-rest ROC-AUC values are also used to examine class-level errors and threshold-based discrimination across the five DR grades.

Experimental Design

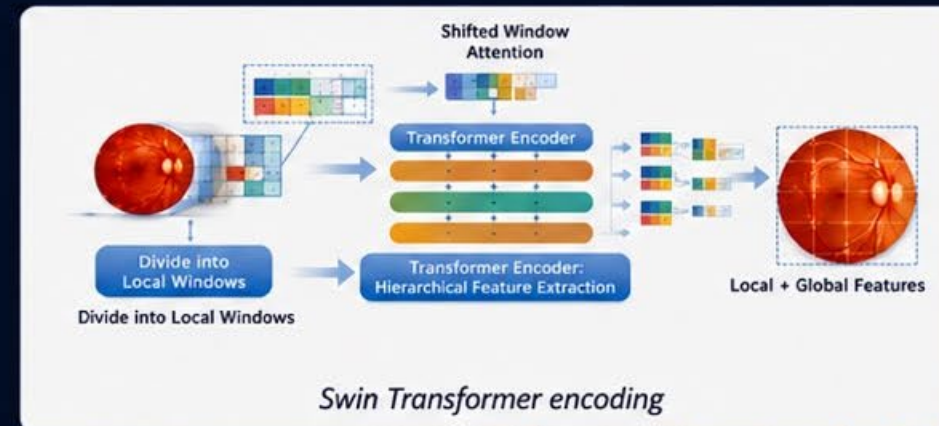
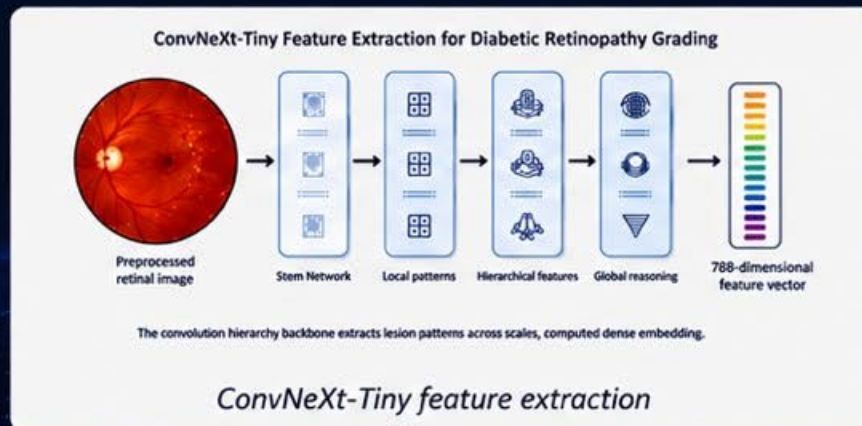
Experimental Step	Implementation
Dataset split	Stratified 70:15:15 split into training, validation, and testing subsets.
Deep-feature extraction	Fine-tuned ConvNeXt-Tiny and Swin-V2-Tiny generated fixed 768-dimensional embeddings.
Fusion and classification	Concatenated embeddings formed a 1,536-dimensional vector classified by Random Forest.
Final evaluation	The held-out test subset was used only after model selection was complete.

Subset		Images	Percentage
 Training		2,563	70%
 Validation		549	15%
 Testing		550	15%

Individual Feature-Extractor Results

The individual backbones were used as complementary feature extractors rather than as the final classifier.

Parameter	ConvNeXt-Tiny	Swin-V2-Tiny
Pretrained weights	ImageNet-1K V1	ImageNet-1K V1
Batch size	32	32
Classifier head	Dropout (0.45) + Linear (768→5)	Dropout (0.20) + Linear (768→5)
Loss function	Class-weighted cross-entropy	Class-weighted cross-entropy
Label smoothing	0.06	0.02
Learning rate	Single (not differentiated)	Backbone 1×10^{-5} / Head 2×10^{-4}
Optimizer	AdamW	AdamW

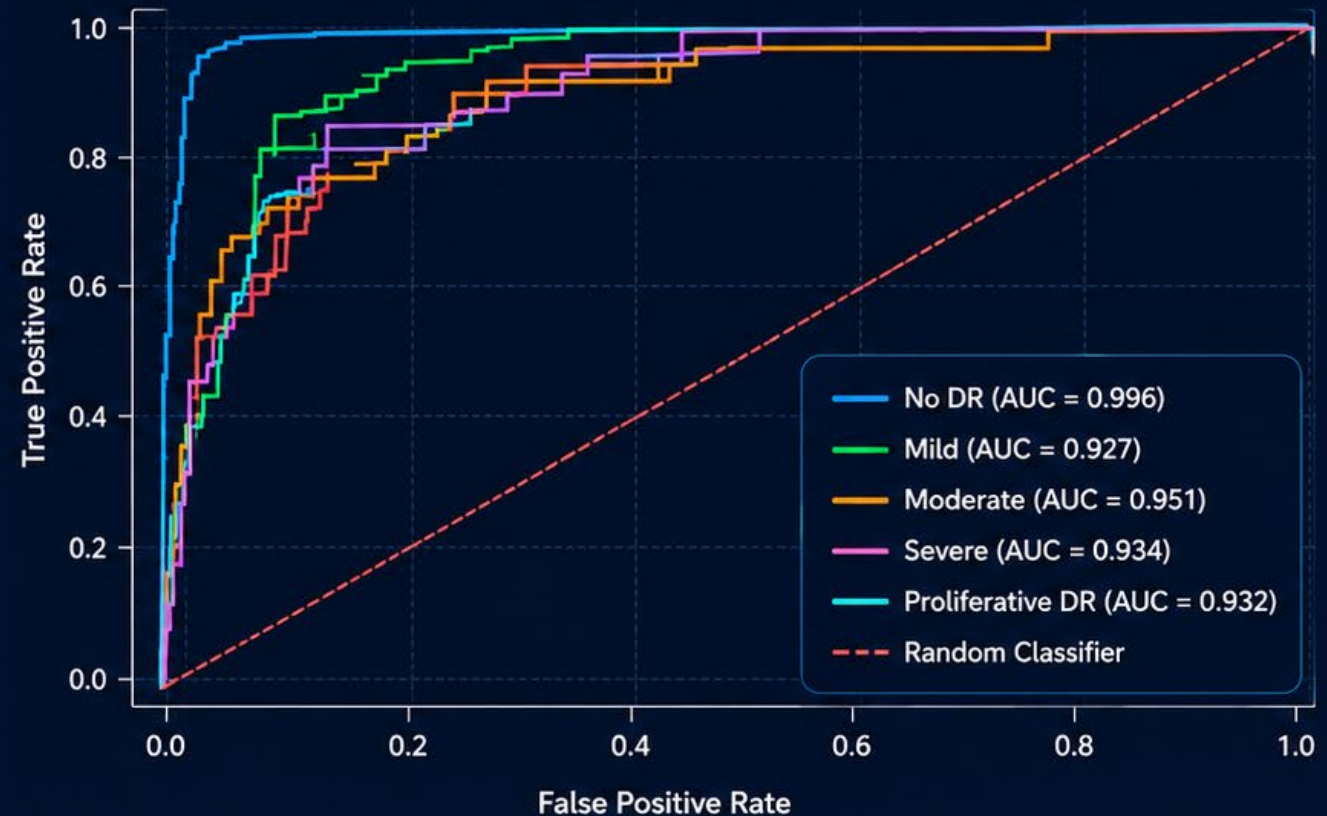


Final Hybrid Framework: Results & ROC Curve

Evaluation Metrics

Metric	Value (%)
 Testing Accuracy	84.91%
 Macro Precision	73.93%
 Macro Recall	68.99%
 Macro F1 Score	70.98%

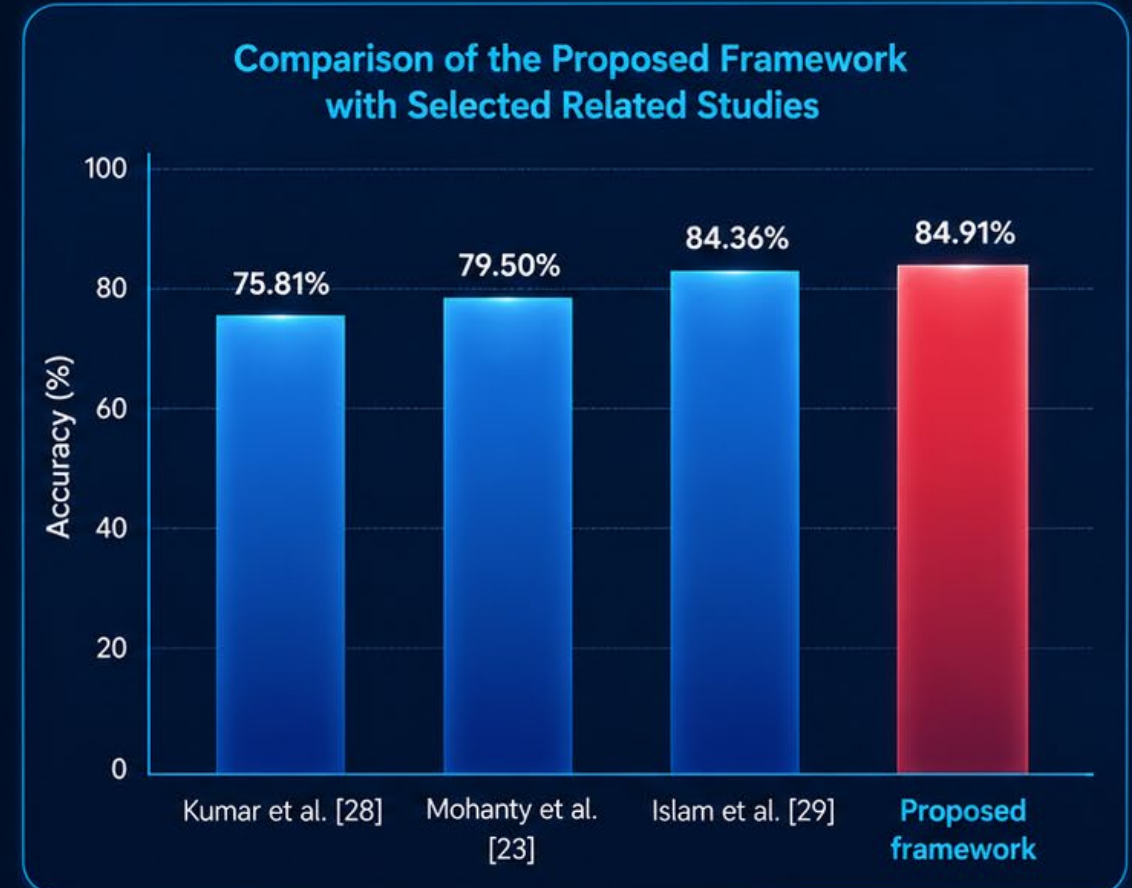
One-vs-Rest ROC Curves of the Final Hybrid Random Forest Model



Comparison of the Proposed Framework with Selected Related Studies



 Study	 Method	 Dataset / Task	 Reported Accuracy
Kumar et al. [10]	DRISTI: VGG16 + Capsule Network	APTOS 2019; five-class grading	75.81%
Mohanty et al. [11]	VGG16 feature extraction + XGBoost	APTOS 2019; five-class grading	79.50%
Islam et al. [12]	Xception + supervised contrastive learning	APTOS 2019; five-class grading	84.36%
Proposed framework	ConvNeXt-Tiny + Swin-V2-Tiny + Random Forest	APTOS 2019; five-class grading	84.91%



Conclusion & Future Work



Conclusion

This thesis presents a hybrid deep learning framework that effectively classifies diabetic retinopathy into five severity grades using fundus images.

By combining ConvNeXt-Tiny and Swin-V2-Tiny features with a Random Forest classifier, the model achieved 84.91% test accuracy and 94.80% macro ROC-AUC, outperforming existing approaches.



Future Work

Future work will focus on further improving performance through advanced hyperparameter optimization, better handling of class imbalance, and exploration of explainable AI techniques.

Additionally, reducing computational cost and validating the model on larger multi-center datasets will enhance its real-world applicability.

References (1/2)

[1] Karthik, M. & Dane, S. (2019). APTOS 2019 Blindness Detection. Kaggle. Dataset:

<https://www.kaggle.com/competitions/aptos-2019-blindness-detection>

[2] Liu, Z. et al. (2022). A ConvNet for the 2020 s. CVPR.

[3] Liu, Z. et al. (2022). Swin Transformer V 2: Scaling Up Capacity and Resolution. CVPR.

[4] Breiman, L. (2001). Random Forests. Machine Learning, 45, 5–32.

[6] Lin & Wu (2023) — Revised ResNet-50 for Diabetic Retinopathy Detection

[7] Das et al. (2022) — DRFEC: CNNs for Feature Extraction and Classification

[8] Jiwani et al. (2023) — CNN-based Diabetic Retinopathy Classification

[9] Mehboob et al. (2022) — A Deep Learning Based Approach for Grading of Diabetic Retinopathy Using Large Fundus Image Dataset

References (2/2)

[10] Kumar, G. et al. (2021). DRISTI: A Hybrid Deep Neural Network for Diabetic Retinopathy Diagnosis. Signal, Image and Video Processing.

[11] Mohanty, C. et al. (2023). Using Deep Learning Architectures for Detection and Classification of Diabetic Retinopathy. Sensors.

[12] Islam, M. R. et al. (2022). Supervised Contrastive Learning for Diabetic Retinopathy Detection. Computers in Biology and Medicine.

[13] Guefrachi, S. et al. (2025). Diabetic Retinopathy Detection Using Deep Learning Multistage Training. Arabian Journal for Science and Engineering.

[14] Akhtar, S. et al. (2025). A Deep Learning Based Model for Diabetic Retinopathy Grading. Scientific Reports.

Thank You

Questions & Discussion
