Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

A Hybrid OCTA Framework Integrating Quantitative Handcrafted Biomarkers and Deep Learning for Automated Diabetic Retinopathy Grading

Esmat Ramezanzadeh
North Khorasan University of Medical Sciences
Naser Shoeibi
Mashhad University of Medical Sciences
Mohammad Hossein Bahreyni Toosi
Mashhad University of Medical Sciences
Hamidreza sadoughi
North Khorasan University of Medical Sciences
Meysam Tavakoli
meysam.tavakoli@emory.edu
Emory University

Article

Keywords: Oculomics, Tomography, Optical Coherence Tomography Angiography, Machine Learning, Diabetic Retinopathy

Posted Date: September 15th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10572135/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

A Hybrid OCTA Framework Integrating Quantitative Handcrafted Biomarkers and Deep Learning for Automated Diabetic Retinopathy Grading

Esmat Ramezanzadeh[1], Naser Shoeibi[2], Mohammad Hossein Bahreyni Toosi[3], Hamidreza sadoughi[1], Meysam Tavakoli[4]*.

1 Department of Medical Physics and radiology, Faculty of Allied Medical Sciences, North Khorasan University of Medical Sciences, Bojnurd, Iran.
2 Eye Research Center, Mashhad University of Medical Sciences.
3 Medical Physics Research Center, Basic Sciences Research Institute, Mashhad University of Medical Sciences, Mashhad, Iran
4 Department of Radiation Oncology, and Winship Cancer Institute, Emory University, Atlanta, Georgia, USA, Midtwon Emory University Hospital, Atlanta, Georgia.

** Corresponding author's e-mail: Email: meysam.tavakoli@emory.edu*

Abstract

Purpose: This study aimed to develop and evaluate a hybrid OCTA-based framework integrating handcrafted biomarkers, deep-derived features, and stability-guided feature selection for automated diabetic retinopathy (DR) grading.

Methods: The study included a total of 377 patients with DR using OCTA images from the publicly available database (FAZID) and a local cohort one. The OCTA images underwent a dedicated preprocessing pipeline comprising contrast enhancement, denoising, and vessel enhancement. Hybrid feature representations were generated by combining handcrafted OCTA biomarkers with deep features extracted from Vision Transformer (ViT-B16), EfficientNet-B0, and ResNet50. Stable feature selection integrating mutual information, LASSO, and repeated cross-validation was applied before classification using XGBoost, Random Forest, AdaBoost, multilayer perceptron, and a stacking ensemble. Hyperparameters were optimized by nested cross-validation within the training set, and final performance was assessed on a held-out test set with bootstrap confidence intervals.

Results: The proposed feature selection framework reduced feature dimensionality by up to 96.8% while retaining stable discriminative features. Preprocessing consistently improved classification performance, particularly for transformer-based models. The hybrid ViT-B16 + handcrafted features with XGBoost achieved the best performance (AUC = 0.95, balanced accuracy = 0.82, F1-score = 0.84), with robust cross-validation performance and excellent discrimination of moderate NPDR and early PDR.

Conclusion: The proposed hybrid framework demonstrates the potential of combining quantitative OCTA biomarkers with deep learning (DL) representations to achieve robust, accurate, and clinically relevant automated DR grading.

Keywords: Oculomics, Tomography, Optical Coherence Tomography Angiography, Machine Learning, Diabetic Retinopathy

Abbreviation: Vol = Volume; Thk = Thickness; FD = Fractal Dimension; N = Nasal; T = Temporal; S = Superior; I= Inferior; W = Whole; SH = Superior Hemisphere; IH = Inferior Hemisphere; FAZ = Foveal Avascular Zone; IPL = Inner Plexiform Layer; OPL = Outer Plexiform Layer; INL = Inner Nuclear Layer; IOS = Inner Outer Segment; BRM = Bruch's Membrane; RPE= Retinal Pigment Epithelium; FAZ= fovea avascular zone;

1 Introduction

Diabetic Retinopathy (DR) is a major microvascular complication of diabetes, caused by progressive retinal capillary damage, which can lead to blindness if untreated [1, 2]. DR evolves from Non-Proliferative DR (NPDR) to Proliferative stages (PDR), with early PDR carrying a high risk of rapid vision loss [3, 4]. Accurate

staging is thus essential for timely therapeutic intervention [5, 6]. Treatment options vary by stage from routine monitoring and anti-vascular endothelial growth Factor (VEGF) injections before early PDR [7, 8] to laser photocoagulation and retinal surgery afterward [5, 9]. Early detection remains key to preserving vision and improving patients' quality of life [10].

Diagnosis often begins with fundus examination using ophthalmoscopy or a Goldmann lens combined with fundus photography [11-13]. Although fluorescein angiography (FFA) is the gold standard for visualizing neovascularization and ischemic areas [14, 15], its invasive nature limits frequent use. Optical coherence tomography (OCT) and OCT angiography (OCTA) provide complementary, high-resolution and non-invasive 3D insights into retinal microvasculature [16-18], enabling visualization of both NPDR and PDR changes [19]. Structural OCT provides cross-sectional information on retinal morphology and fluid, whereas OCTA noninvasively visualizes perfused retinal and choroidal microvasculature [20]. NPDR typically manifests as microaneurysms, hemorrhages, and ischemia [21], while early PDR exhibits neovascularization and leakage [12, 19]. However, conventional manual assessment is subjective, costly, and inefficient for mass screening.

Among these challenges, the transition from NPDR to PDR represents a critical turning point in disease progression and requires meticulous clinical evaluation, as therapeutic decisions and prognosis are highly stage-dependent. Accurate identification of this shift remains challenging in routine retinal imaging, since neovascularization may be subtle, fine, and easily obscured by hemorrhages or imaging artifacts. Consequently, reliable detection of neovascular changes demands careful assessment and remains a key bottleneck in timely PDR diagnosis [22, 23].

To address these limitations, deep learning (DL) and machine learning (ML) have been employed to enhance DR detection accuracy, reduce inter-observer variability, and automate staging from fundus, FFA, and OCTA images [24-29]. Among these approaches, oculomics has emerged as a promising strategy, transforming medical images into high-dimensional quantitative data by extracting numerical descriptors such as texture, shape, and intensity [29-31]. Combined with ML-based feature selection, oculomics can significantly improve diagnostic and prognostic performance [29, 32].

Recent studies have increasingly leveraged ML models to analyze OCTA and clinical data for DR classification, reporting encouraging results in automated diagnosis and precision medicine [33, 34]. Investigations by Sandhu et al. [33], Xiaoli et al. [34], Aslam et al. [35], Zhiping et al.[1], and Alam et al. [36] demonstrated accuracies exceeding 80 % using algorithms such as random forest, support vector machines, and logistic regression. Commonly used oculomics features included vessel density (VD), retinal layer thickness, and clinical biomarkers. These studies primarily relied on quantitative OCTA biomarkers such as VD, retinal layer thickness, and selected clinical variables, highlighting the diagnostic value of retinal vascular and structural alterations.

More recently, oculomics-based approaches have demonstrated the potential of combining quantitative ocular biomarkers with ML for retinal disease analysis. For example, Meng et al. [8] employed OCT-derived oculomics for predicting persistent diabetic macular edema; Kar et al. [37] analyzed spatial oculomics for anti-VEGF treatment response; Soren et al. [27] applied oculomics on ultra-widefield OCTA for DR grading. Collectively, these studies emphasize the growing relevance of oculomics for retinal disease analysis, yet a unified framework that combines multimodal OCT or OCTA oculomics with clinical data for DR staging remains unexplored.

However, most existing approaches rely either on handcrafted biomarkers or DL representations, while their complementary integration within a unified framework remains relatively unexplored [38, 39]. Furthermore, image preprocessing is frequently regarded as a preliminary enhancement step rather than a key component influencing both handcrafted feature extraction and deep representation learning [40-42]. In addition, feature selection is commonly performed according to feature importance alone, whereas

the stability and reproducibility of the selected biomarkers across different training subsets are rarely investigated. Consequently, although existing models have achieved promising predictive performance, their biological interpretability and robustness remain limited.

To address these limitations, the present study proposes a hybrid OCTA-based framework that integrates handcrafted oculomic biomarkers including: vascular, structural, texture, morphological, statistical, and clinical features with deep representations extracted from three complementary pretrained architectures (ViT-B16, EfficientNet-B0, and ResNet50). The proposed framework further incorporates a dedicated preprocessing pipeline and a stability-guided feature selection strategy to improve the robustness and reproducibility of biomarker selection. By combining biologically meaningful OCTA biomarkers with complementary deep representations, the proposed approach aims to provide a more accurate, robust, and clinically interpretable framework for automated DR grading than either handcrafted or deep features alone.

2 Result

This section presents the classification results obtained using handcrafted features and three DL feature sets (EfficientNet-B0, ResNet50, and ViT-B16), each combined with the handcrafted features.

2.1 Study Population and Dataset Characteristics

Table 1 summarizes the image distribution and demographic characteristics of the FAZID and local cohort. The FAZID dataset included superficial OCTA for 75 patients, whereas the local cohort comprised 302 patients spanning three DR stages (moderate NPDR, severe NPDR, and early PDR). The distribution of right and left eyes, as well as the female-to-male ratio, was comparable across the disease stages, providing a balanced dataset for subsequent model development and evaluation.

Table 1. Image Distribution and Demographic Characteristics of the FAZID and the Persian cohort Datasets.

Dataset	Stage	Number of patients	OD: OS	Gender ratio (Female: Male)
FAZID Dataset	***Moderate***	35	17:18	21:14
	Severe NPDR	25	14:11	6:19
	Early PDR	15	6:9	2:13
Our Cohort Dataset	***Moderate***	102	53:49	58:44
	Severe NPDR	88	47:41	64:24
	Early PDR	112	57:55	59:53

2.2 Feature Selection Results

Table 2 summarizes the dimensionality reduction achieved by the proposed feature selection framework for both original and pre-processed image feature sets. Across all feature combinations, the number of retained features was markedly reduced, indicating that the proposed strategy effectively removed redundant and less informative variables while preserving the most stable and discriminative features for subsequent classification.

Table 2. Dimensionality reduction achieved by the proposed feature selection framework for handcrafted and hybrid feature sets extracted from original and pre-processed images.

		Original Features	Selected Features	Reduction (%)
Pre-processed Image	Handcrafted	618	36	94.1
	EfficientNet-B0 + Handcrafted	1896	76	95.9
	ViT-B16 + Handcrafted	1384	71	94.8

	ResNet50 + Handcrafted	2664	83	96.8
Original Images	Handcrafted	618	38	93.8
	EfficientNet-B0 + Handcrafted	1896	72	94.7
	ViT-B16 + Handcrafted	1384	73	96.2
	ResNet50 + Handcrafted	2664	89	96.6

Figure 2 shows the distribution of selected features based on MI Score (x-axis) and LASSO (y-axis). Each bubble represents a feature, with larger bubbles indicating higher relative importance/stability in the final selection. Panels A-D correspond to handcrafted features only, handcrafted + EfficientNet-B0, handcrafted + ResNet50, and handcrafted + ViT-B16, respectively. Overall, features in the upper-right region contribute most strongly to class discrimination. In panel A, VD, Kurtosis, L2-G22, FD-wavelet, and VT-Wavelet were the most relevant handcrafted features. In panel B, VD, center of thickness ILM-RPE, LBP-mean and FD-Wavelet remained prominent, together with EfficientNetB0-derived features. In panel C, Gabor, VD, Kurtosis, and ResNet50-derived features were dominant. In panel D, VD was the strongest overall feature, followed by Width Variation, Kurtosis, VD deep layer and G22 section(L2-G22), ViT-B16-derived features, and Circularity. These results indicate that handcrafted descriptors remain highly informative, while deep features provide complementary discriminative value.

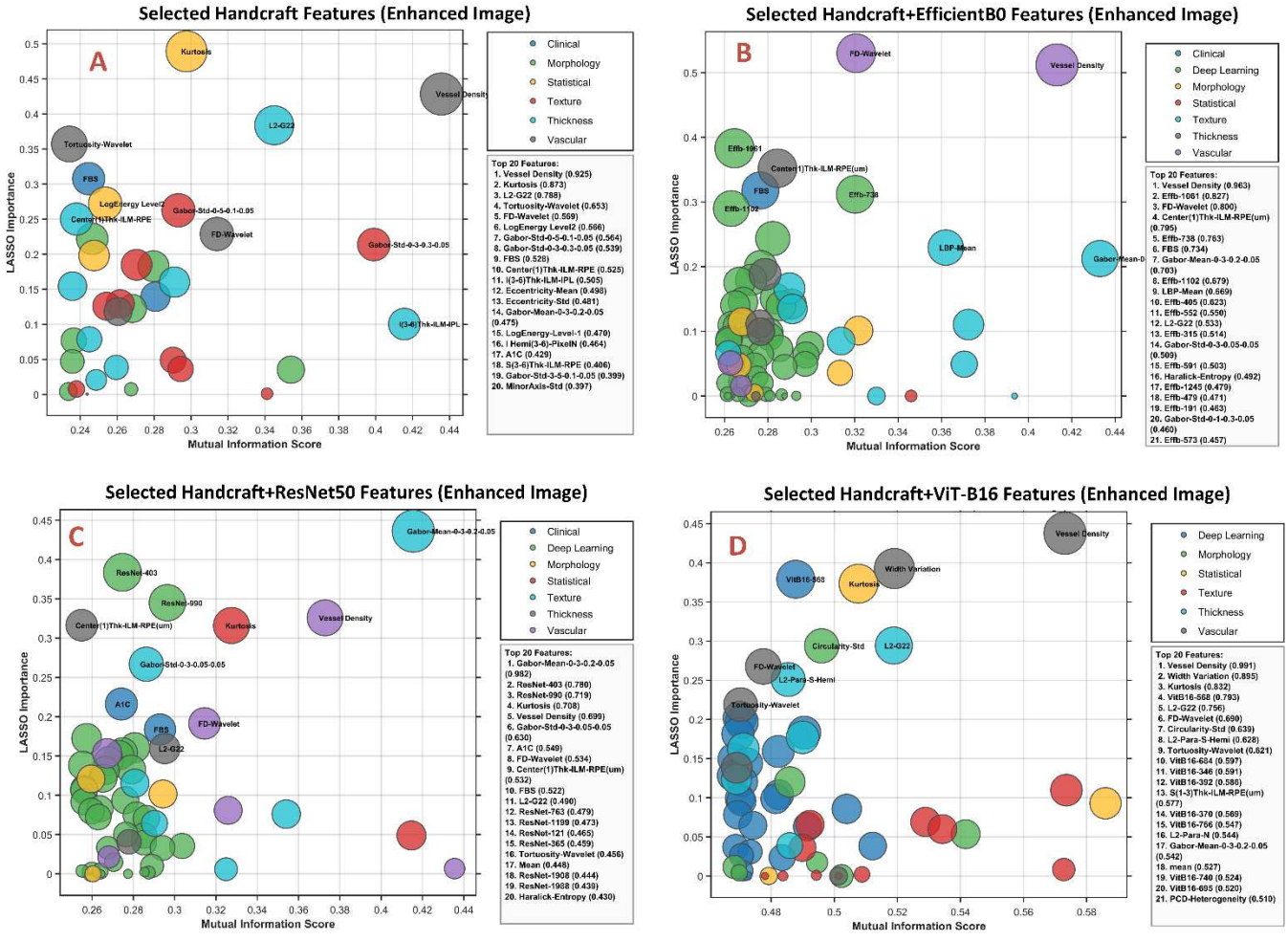


Figure 2. Evaluation of feature importance using the hybrid MI-LASSO model for the used deep learning architectures: A) Handcrafted, B) Handcrafted+EfficientB0, C) Handcrafted+ResNet50, and D) Handcrafted+ViT-B16

features. The bubble charts illustrate the distribution of handcrafted and deep features across seven categories, where the x-axis represents Mutual Information and the y-axis indicates LASSO importance. Bubble size reflects the combined importance score, with the top-ranked features for each model summarized in the adjacent sidebar.

2.3 Comprehensive Performance, Diagnostic, and Robustness Evaluation

To evaluate the proposed DR grading framework, we performed a comprehensive performance analysis. First, we compared all feature-classifier combinations, then examined the class-wise diagnostic performance of the best models, and finally evaluated the robustness and stability of the classification pipeline using cross-validation.

2.3.1 Ablation Study: Impact of Feature Representation, Classifier Selection, and Preprocessing Impact

To quantify the contribution of each component in the proposed framework, we performed a comprehensive ablation study by systematically comparing different feature representations (handcrafted alone and hybrid combinations with ViT-B16, EfficientNet-B0, and ResNet50), multiple classifiers, and the effect of the proposed preprocessing pipeline. Performance was evaluated using AUC, Balanced Accuracy (B-ACC), and F1-score. The results for original and preprocessed images, together with the absolute performance gain (Δ), are summarized in Figure 3.

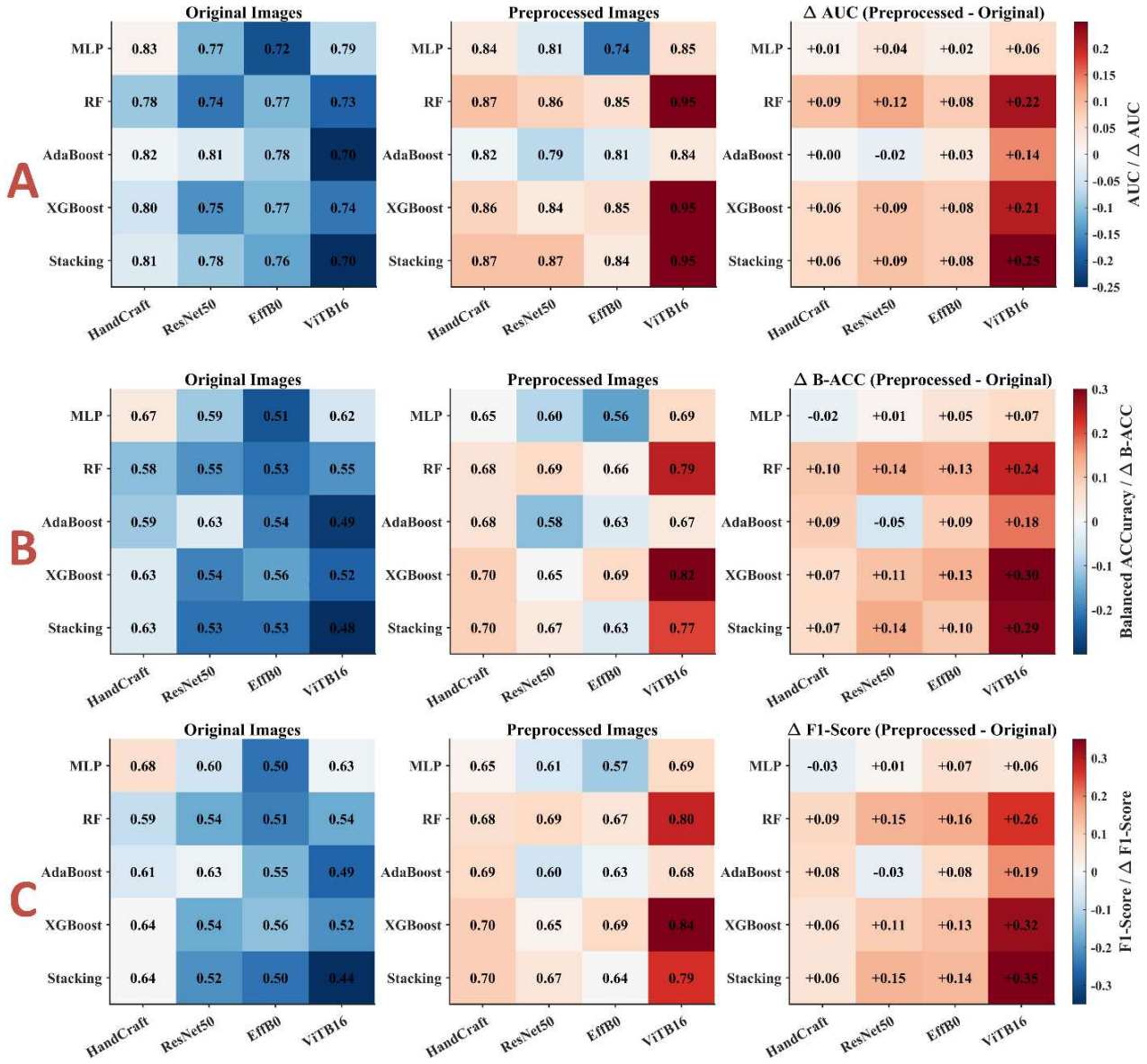


Figure 3. Comparative performance heatmaps across various combinations of handcrafted, handcrafted combined ViT-B16 (ViT-B16), handcrafted combined ResNet50(ResNet50), and handcrafted combined efficientNetB0 (EffB0) sets with different classifiers. The third column (Δ) highlights the absolute performance improvements yielded by the proposed image preprocessing pipeline.

Figure 3 demonstrates that classification performance on original, unprocessed images was generally limited, particularly among deep feature extractors, which struggled to reach satisfactory benchmarks. Preprocessing emerged as a critical corrective measure, substantially elevating these baseline metrics across all models. ViT-B16 exhibited the highest sensitivity to this enhancement, yielding the most significant performance gains. Notably, the Stacking ensemble with ViT-B16 achieved the highest improvements in AUC (+0.25) and F1-score (+0.35), while the XGBoost–ViT-B16 configuration provided the maximum gain in B-ACC (+0.30). Consequently, this preprocessing strategy transformed constrained baseline results into robust performance, with the XGBoost–ViT-B16 combination attaining a peak AUC of 0.95, B-ACC of 0.82, and F1-score of 0.84.

2.3.2 Diagnosis and Class-Wise Performance

Figure 4 summarizes the ROC-AUC curves and normalized confusion matrices for the best-performing DR classifiers under original and enhanced imaging conditions. On original images, MLP with handcrafted features yields the most reliable baseline (AUCs of 0.91, 0.75, and 0.82 for Classes moderate-NPDR, severe-NPDR, and early-PDR, respectively). However, the overall strongest performance is achieved on enhanced

images by XGBoost with hybrid ViT-B16 + handcrafted features, which secures AUCs of 0.96, 0.90, and 0.96 for Classes moderate-NPDR, severe-NPDR, and early-PDR. The transition from original to enhanced images significantly improves class separability and reduces misclassification across all architectures, although the diagnostic bottleneck remains the differentiation between classes severe-NPDR and early-PDR.

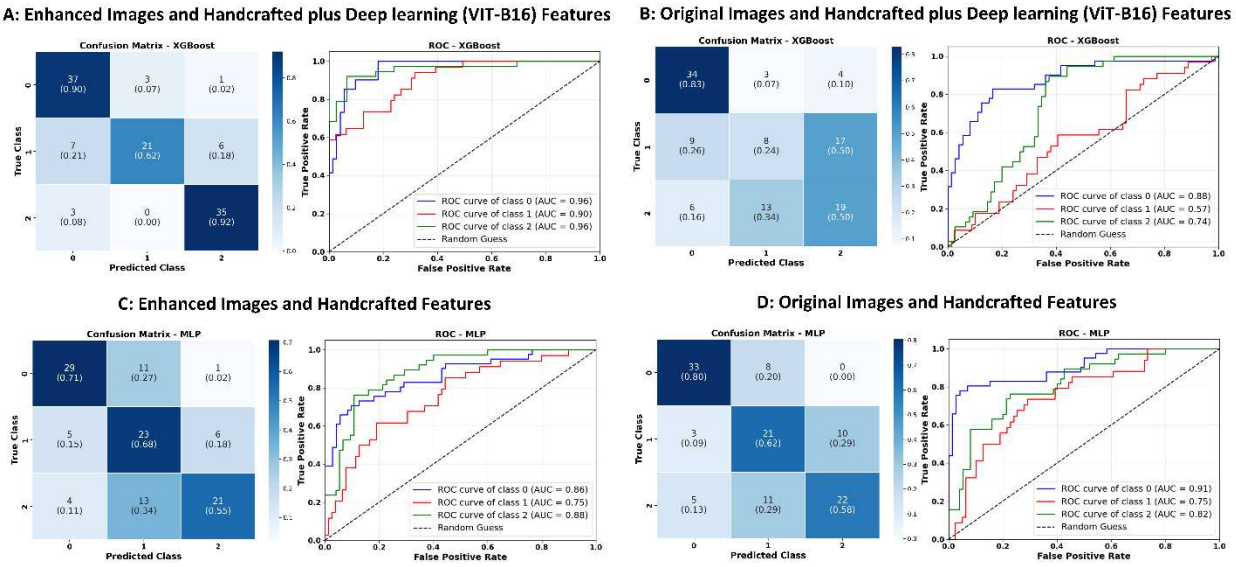


Figure 4. Diagnostic performance breakdown. Panels (A) and (C) depict the normalized confusion matrices and class-wise ROC curves for the top-performing configurations (Hybrid features with XGBoost/MLP) utilizing preprocessed images. Panels (B) and (D) illustrate the corresponding performance on original, non-preprocessed images.

2.3.3 Multi-Metric Robustness and Stability Analysis

To ensure that the performance of the proposed pipeline is statistically stable and not overfitted, we perform a sensitivity analysis using 5-fold cross-validation. Figure 5 presents a comparative breakdown across eight distinct classification metrics, accompanied by standard deviation error bars.

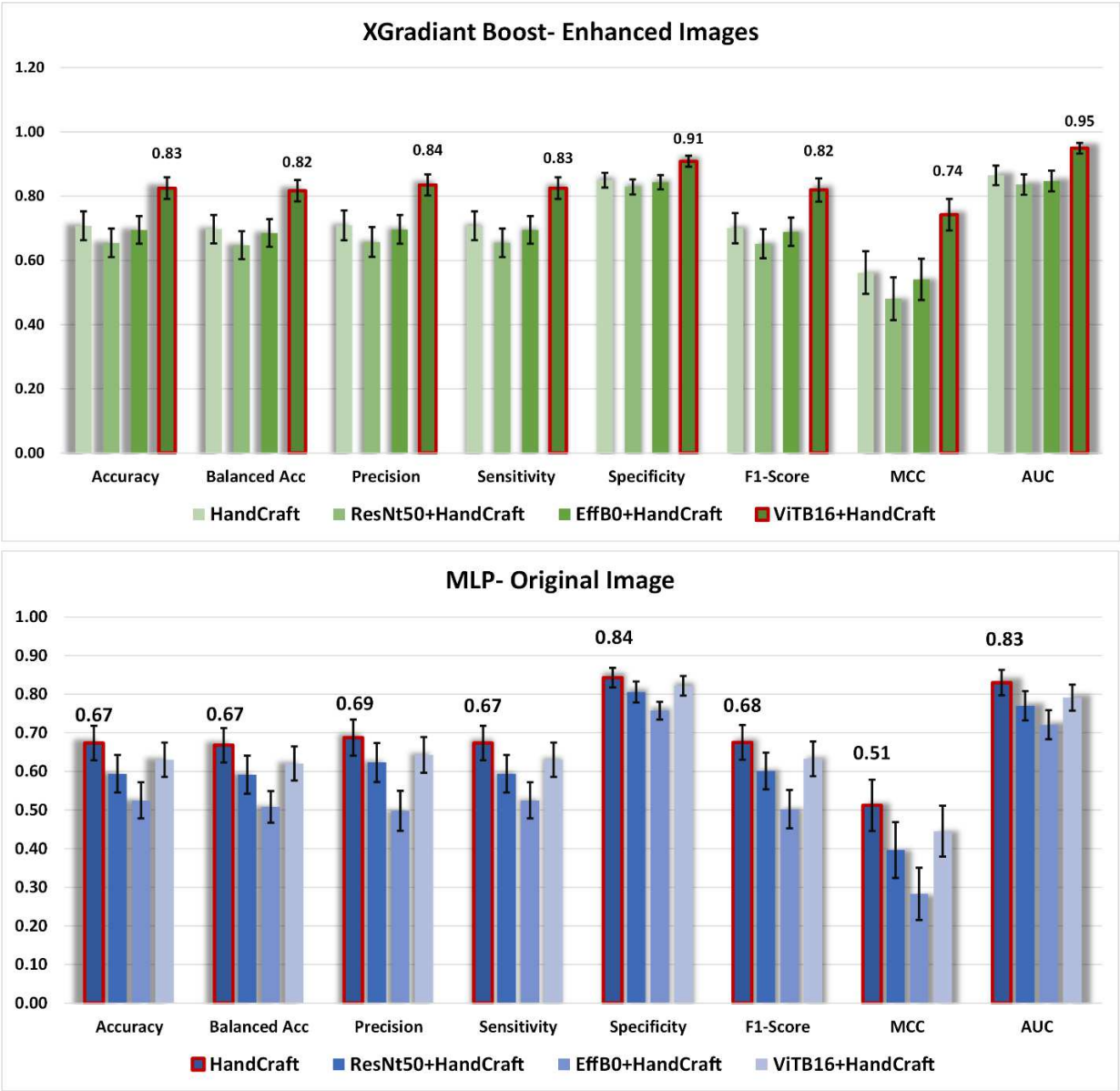


Figure 5. Multi-metric sensitivity and validation stability analysis. The bar charts compare eight evaluation metrics across different feature combinations for (Top) the XGBoost model on preprocessed images, and (Bottom) the MLP model on original images. Error bars denote the standard deviation obtained across 5-fold cross-validation.

The validation results in Figure 5 confirm the robustness of the proposed framework. When using preprocessed images, the XGBoost model utilizing the hybrid ViT-B16 + Handcrafted feature set outperforms all other combinations, maintaining narrow error bars across crucial metrics such as Matthew's Correlation Coefficient (MCC = 0.74 ±0.02) and Sensitivity (0.83 ±0.03). This narrow variance demonstrates high generalization stability. In contrast, the MLP model evaluated on original images (Figure 5, Bottom) exhibits not only significantly lower mean scores across all metrics (e.g., F1-score of 0.68, MCC of 0.51) but also considerably wider error bars, highlighting its vulnerability to noise and variations in raw input images.

3 Discussion

In this study, we developed and evaluated a hybrid OCTA-based framework for automated DR grading by integrating handcrafted vascular biomarkers, deep features extracted from pretrained networks, and

stability-guided feature selection. The principal finding of this study is that integrating handcrafted OCTA biomarkers with deep representations provides a more effective framework for DR grading than either feature family alone. The superior performance of the hybrid framework suggests that handcrafted descriptors and deep features capture complementary aspects of retinal pathology, enabling a more comprehensive characterization of disease-related changes than can be achieved using either representation independently. This complementary behavior is further reflected in the characteristics of the selected feature subset.

The consistent selection of vascular descriptors, including VD, FD, and VT, suggests that retinal microvascular alterations remain the dominant source of discriminative information for DR staging. This observation is biologically plausible because DR progression is characterized by progressive capillary dropout, vascular remodeling, and increasing disruption of the retinal microvascular network, all of which directly influence VD and branching complexity. One possible explanation for the strong contribution of FD and VT is that their wavelet-based computation may better capture multiscale vascular organization and subtle structural alterations that are less apparent using conventional measurements [22]. Together, these findings indicate that vascular biomarkers derived from OCTA provide biologically relevant information that complements deep image representations and contributes to more reliable disease characterization.

Beyond identifying biologically relevant biomarkers, the proposed feature selection pipeline substantially reduced feature dimensionality while retaining a compact subset of informative variables. This finding suggests that a large proportion of the original descriptors contained redundant or overlapping information, whereas the retained features captured the essential characteristics required for robust DR classification. Reducing feature redundancy is also expected to improve model generalization and decrease the risk of overfitting, particularly in relatively small medical imaging datasets. Importantly, incorporating stability selection into the combined MI–LASSO ranking enabled the consistent identification of informative features across repeated cross-validation iterations, reducing the likelihood of selecting unstable or dataset-specific variables and increasing confidence in the reproducibility of the selected biomarker set.

The benefits of hybrid feature integration were further enhanced by the preprocessing strategy, which consistently improved classification performance across multiple experimental settings. Consistent with the findings of Grace et al. [40] and Sara et al. [41], preprocessing contributed to higher classification accuracy and improved overall evaluation metrics, supporting its role as more than a conventional image enhancement step. Rather than simply improving visual appearance, preprocessing is likely to increase the fidelity of disease-relevant vascular and textural information while reducing image variability caused by noise, illumination inconsistencies, and OCTA-specific artifacts. Consequently, both handcrafted descriptors and deep representations can be extracted more reliably, leading to improved discrimination between DR stages. Unlike the image-channel fusion strategy reported by Grace et al. [40], our findings further suggest that preprocessing also facilitates more effective integration of handcrafted OCTA biomarkers with transformer-derived representations, thereby strengthening the complementary information exploited by the hybrid framework.

An additional observation was that different deep backbones consistently selected distinct subsets of informative features, suggesting that each architecture captures different characteristics of retinal pathology. This behavior is consistent with the inherent differences between feature extraction mechanisms, where convolutional networks primarily emphasize local structural and textural patterns, whereas transformer-based models are more capable of modeling long-range spatial relationships across the retinal vasculature. Consequently, the complementary nature of these learned representations provides a plausible explanation for the improved performance observed when they are integrated with

handcrafted OCTA biomarkers. Similar benefits of combining handcrafted and deep features have recently been reported for fundus image analysis [42]. However, unlike fundus photography, OCTA directly visualizes the retinal microvasculature, allowing the hybrid framework to exploit vascular biomarkers that are more closely associated with the underlying pathophysiology of DR. This distinction may explain why hybrid feature integration provides not only improved discriminative performance but also greater biological relevance for disease staging.

The improvement observed after preprocessing further suggests that its impact extends beyond conventional image enhancement and depends on both the feature extraction architecture and the downstream classifier. The more pronounced performance gains achieved by the ViT-B16-based models indicate that transformer-derived representations are particularly sensitive to image quality, likely because self-attention mechanisms rely on accurate global structural information that can be degraded by noise, illumination variability, and OCTA-specific artifacts. In contrast, the consistently strong performance of XGBoost suggests that boosted decision trees are well suited to exploiting the heterogeneous information contained in both handcrafted descriptors and transformer-derived embeddings. Similar observations have recently been reported in fundus-based DR grading, where Adaptive Sigmoid Enhancement combined with hybrid handcrafted and deep features achieved the highest performance[42]. Collectively, these findings suggest that preprocessing should be considered an integral component of hybrid OCTA analysis pipelines, as it enhances the quality of learned representations and enables more reliable disease discrimination.

The proposed framework may have important clinical implications for OCTA-based DR assessment. By integrating quantitative vascular biomarkers with deep image representations, the model provides an objective approach for disease staging while reducing reliance on subjective image interpretation. Because OCTA is a non-invasive imaging modality capable of visualizing retinal microvasculature without dye injection, the proposed strategy may facilitate automated screening and longitudinal monitoring of DR progression in routine clinical practice.

The proposed framework has several methodological strengths, including hybrid integration of handcrafted OCTA biomarkers and transformer-derived features, dedicated preprocessing, robust feature selection, and comprehensive evaluation across multiple deep architectures and classifiers. Nevertheless, several limitations should be acknowledged. The study did not encompass the full spectrum of DR severity, and independent external validation was constrained by the limited availability of publicly accessible OCT/OCTA datasets. In addition, only SCP images were analyzed to maintain consistency with the public FAZID dataset and minimize potential confounding effects arising from layer-specific vascular differences. Finally, the relatively limited sample size and use of a single OCTA device may restrict the generalizability of the findings. Future multicenter studies incorporating larger cohorts, multiple imaging devices, and multimodal retinal imaging are warranted to further validate the proposed framework.

4 Conclusion

This study presents a hybrid framework for DR grading that integrates handcrafted OCTA biomarkers with deep learning representations, demonstrating that the combination of complementary feature types provides superior performance compared with either approach alone. The results further highlight the importance of dedicated image preprocessing, which enhanced the quality of both handcrafted and transformer-derived features and consistently improved classification performance across different architectures and classifiers. The proposed stability-guided feature selection strategy effectively reduced feature dimensionality while identifying reproducible and biologically relevant biomarkers, with vascular descriptors emerging as the most informative indicators of DR severity. Collectively, these findings demonstrate that combining quantitative OCTA biomarkers with deep representations offers a robust and

interpretable approach for automated DR assessment. Although further validation on larger multicenter cohorts and multimodal retinal imaging datasets is warranted, the proposed framework represents a promising step toward reliable OCTA-based computer-aided DR grading and has the potential to support objective disease staging and clinical decision-making in routine ophthalmic practice.

5 Acknowledgements

The authors would like to thank the Vice-Chancellor of the Research Department at North Khorasan University of Medical Sciences for supporting this project.

6 Funding Sources

This work was funded by the Vice-Chancellor for Research at North Khorasan University of Medical Sciences (Grant number: 4050116).

7 Methodology

Figure 1 presents a comprehensive block diagram illustrating the complete workflow for preprocessing, feature extraction and selection, and classification of DR using OCTA images.

7.1 Study subjects

This prospective study included 377 patients diagnosed with DR, comprising 302 patients from local cohort and 75 ones from the publicly available FAZID dataset [43]. The local cohort component of the study was conducted in accordance with the Declaration of Helsinki and all relevant guidelines and regulations. The study protocol and OCTA examinations were approved by the Ethics Committee of Mashhad University of Medical Sciences (MUMS), Mashhad, Iran (Ethics code: IR.MUMS.REC.1401.669). Written informed consent was obtained from all participants in the local cohort prior to their inclusion in the study.

The participants ranged in age from 29 to 79 years, with a mean age of 59 years for men and 57 years for women. Three expert ophthalmologists with different experience confirmed the diagnosis and severity of DR through dilated posterior segment examinations using slit-lamp biomicroscopy [44]. Before grading, all evaluators reviewed and agreed upon the DR staging criteria defined by the senior ophthalmologist. In cases of disagreement (especially between severe-NPDR, and early PDR) FFA imaging was used, and the final stage was decided by consensus among all three ophthalmologists. Lens clarity was graded during slit-lamp assessment using four predefined categories (completely clear, clear, slightly turbid, turbid). For stage-based comparisons, only the proportion of eyes classified as completely clear was used.

Exclusion criteria included a history of vitrectomy or laser photocoagulation, the presence of age-related macular degeneration, significant media opacities, high myopia, or a history of intraocular infection, uveitis, or retinitis. Additionally, patients with macular edema associated with vein occlusions or vitreomacular traction syndrome were excluded from the study.

7.2 Data Acquisition

OCTA scans for all the local participants were conducted using the AngioVue system (RTVue XR Avanti, Optovue, Fremont, CA, USA; Software version 2018.0.0.14). The study utilized the AngioRetina protocol [45-47], which captures a 6 × 6 mm scan grid consisting of 400 × 400 A-scans. In addition to the primary cohort, superficial capillary plexus (SCP) OCTA images from the publicly available FAZID database were included in the study [43].

7.3 Methodology

OCTA images were acquired and enhanced using MATLAB R2020b (Intel Core i5-10300H, 8 GB RAM). Classification was conducted in Python 3.11 on an Ubuntu 2020 system (Intel Xeon E5-2620 v3, RTX-3060,

12 GB RAM). The overall workflow of the proposed framework is illustrated in Figure 1, including the preprocessing, feature extraction, and classification stages. Each of these stages is discussed in detail below.

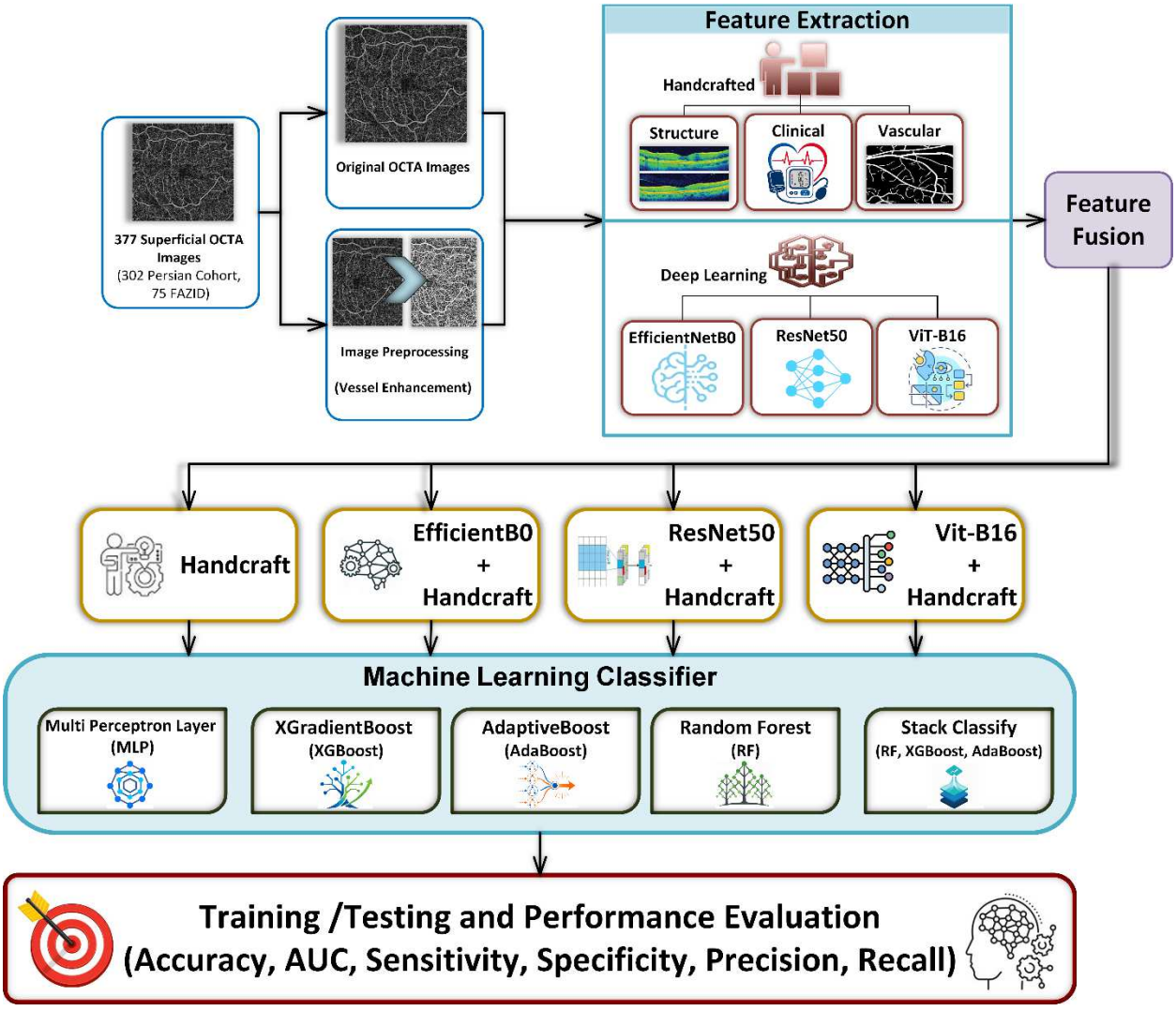


Figure 1. Block diagram of the overall Preprocessing, Feature extraction and classification.

7.3.1 Enhancement procedures

Prior to feature extraction, all superficial OCTA images underwent a standardized preprocessing pipeline to improve image quality and enhance vascular structures. First, contrast enhancement was performed using a combination of Recursive Mean-Separate Histogram Equalization (RMSHE, $n = 1$) and Contrast Limited Adaptive Histogram Equalization (CLAHE, clipLimit = 0.05, numTiles = 5 × 5, nBins = 128) [48, 49], followed by gamma correction to improve vessel visibility. RMSHE enhances images by recursively segmenting the histogram based on mean intensity values, ensuring better preservation of the original brightness while minimizing unwanted artifacts [50]. CLAHE further refines contrast enhancement by partitioning the image into small, localized tiles, performing histogram equalization independently within each tile, and employing contrast limiting to prevent noise amplification [51].

The enhanced images were then denoised using a Non-Local Means filter and DL–based denoising using the Deep Denoising Convolutional Neural Network (DnCNN) algorithm. To improve vascular structure visibility while reducing noise, Wavelet-based denoising [52] was employed to decompose the image into

multiple frequency components, enabling the selective removal of high-frequency noise while retaining essential anatomical details. Additionally, a DnCNN was utilized for noise suppression. This pre-trained DL model effectively removes complex noise patterns without requiring clean reference images, outperforming traditional denoising approaches by maintaining fine vascular structures [53]. In fact, the DL-based denoising, DnCNN [54, 55], designed for noise reduction while preserving fine structural details. DnCNN training on OCTA will suppress noise, enhance vessel visibility, and maintain high-resolution features crucial for detecting vascular abnormalities.

Finally, we applied multi-scale, multi-orientation matched filtering and Frangi vesselness filtering to enhance the retinal microvasculature, yielding high-quality OCTA images for downstream feature extraction and analysis.

7.3.2 Feature Extraction Framework

Two feature levels were extracted in this study: handcrafted features (334 vascular, 262 structural, and 22 clinical), and DL features. The three pretrained feature extractors (EfficientNet-B0, ResNet50, and ViT-B16) were selected because they represent three complementary families of state-of-the-art feature extractors with distinct representation mechanisms. ResNet50 captures hierarchical local features through deep residual convolutional learning, EfficientNet-B0 provides an optimized balance between accuracy and computational efficiency via compound scaling, while ViT-B16 models long-range spatial dependencies using self-attention. Comparing these architectures enabled a systematic evaluation of convolutional and transformer-based representations for OCTA image classification without introducing architectural bias.

Handcrafted oculomic features

Vascular-structural (VA-ST): Our comprehensive feature set comprised 618 metrics derived from three categories: vascular, texture, and structural features. First, we computationally extracted 334 vascular features quantifying retinal vascular network characteristics. Notably, Fractal Dimension (FD), calculated using a hybrid method integrating wavelet analysis with box-counting techniques [22, 56], and vessel tortuosity (VT), computed using hybrid methods combining conventional and wavelet-based analyses, enabled robust characterization of both global and local vascular variations. Additional vascular features included macular perfusion (FAZ parameters), vessel diameter, branching, skeletonization, and perfusion density. We also extracted texture descriptors, including GLCM, LBP, Haralick, entropy, Gabor filter banks, geometric properties (area, eccentricity, perimeter, and orientation), and statistical intensity-distribution parameters. Finally, we incorporated 262 structural and volumetric features directly exported from the Optovue OCT system, detailing retinal layer thickness, volume, and structural biomarkers across the standardized ETDRS grid sectors.

Clinical: Furthermore, a comprehensive set of clinical features was collected for each patient. These included fasting blood sugar (FBS), glycated hemoglobin (HbA1C), visual acuity, intraocular pressure (IOP), blood glucose levels two hours postprandial (B2hr), diabetes duration, age, weight, image quality, sex, the presence of macular edema and etc.

Deep learning Features

To complement the handcrafted oculomic descriptors, deep features were extracted using three ImageNet-pretrained architectures: Vision Transformer (ViT-B16), EfficientNet-B0, and ResNet50. In ViT-B16, images were partitioned into 16×16 patches and processed by a Transformer encoder, with the final CLS token used as the image representation. For the CNN-based architectures (EfficientNet-B0 and ResNet50), feature maps were refined using a multi-attention mechanism consisting of Squeeze-and-Excitation (SE), Channel Attention (CA), and Spatial Attention (SA), followed by global average pooling to obtain discriminative feature vectors. A total of 1,384, 1,896, and 2,664 features were extracted from ViT-

B16, EfficientNet-B0, and ResNet50, respectively. The resulting deep features were concatenated with the handcrafted oculomic descriptors to form hybrid feature representations.

To assess the contribution of different DL architectures, four feature configurations were evaluated: (1) handcrafted oculomic features only; (2) handcrafted + ViT-B16 features; (3) handcrafted + attention-enhanced EfficientNet-B0 features; and (4) handcrafted + attention-enhanced ResNet50 features. All configurations were evaluated using the same ML pipeline to ensure a fair comparison and identify the most effective hybrid feature representation.

7.3.3 Feature selection

To reduce data dimensionality, eliminate redundant features, and improve model robustness, a multi-stage feature selection framework was employed. First, low-variance features were removed using the variance threshold method, as such features provide limited discriminative information for classification. Next, highly correlated features (correlation coefficient > 0.90) were eliminated to reduce multicollinearity, retaining only single representative feature from each group of highly correlated features.

Subsequently, the relevance of each feature to the class labels was quantified using Mutual Information (MI) [57], and the most informative subset of features was retained for further analysis. Following feature standardization, feature importance was estimated using L1-regularized logistic regression (LASSO), an embedded feature selection method that promotes sparsity by shrinking the coefficients of less informative features toward zero while preserving the most discriminative ones [58, 59].

The normalized MI and LASSO importance scores were then integrated using equal weighting (50% each) to obtain a unified importance score for each feature [59]. To assess feature robustness, the feature selection procedure was performed using Repeated Stratified 5-Fold Cross-Validation with 10 repetitions, and the frequency with which each feature was selected across all iterations was used as a measure of its selection stability.

Finally, the unified importance score of each feature was further refined based on its selection stability, and the highest-ranked stable features were retained as the final subset for subsequent classification. A detailed description of the selection stability procedure is provided in the Supplementary Material (S1).

7.3.4 Machine learning Algorithm

Following feature selection, four ML classifiers, namely multi-layer perceptron (MLP) [60], adaptive boost (AdaBoost) [61], x-gradient boost(XGBoost) [62], and random forest (RF), and a stacking ensemble model [63], were evaluated for DR classification. The dataset was divided into training (70%) and independent testing (30%) sets using stratified sampling to preserve the original class distribution. Feature selection was performed exclusively on the training data to prevent information leakage into the validation and test sets.

To alleviate class imbalance, Borderline-SMOTE followed by ADASYN oversampling was applied exclusively to the training set. The oversampled training data were subsequently standardized using z-score normalization before feature selection and classifier training. optimization was performed using nested stratified cross-validation, consisting of a 5-fold outer loop and a 3-fold inner loop. The optimal hyperparameter configuration for each classifier was determined by majority voting across the outer folds and subsequently used to train the final model on the entire training set.

In the stacking architecture, RF, XGBoost, and MLP were used as heterogeneous base learners, while logistic regression was adopted as the meta-learner to integrate their probabilistic outputs. This hierarchical ensemble strategy was designed to leverage the complementary strengths of the individual classifiers and thereby enhance overall generalization performance.

The optimized models were finally evaluated on the independent hold-out test set. Model performance was quantified using Accuracy, Balanced Accuracy, Precision, Sensitivity, Specificity, F1-score, Matthews Correlation Coefficient (MCC), and the area under the receiver operating characteristic curve (ROC-AUC).

To estimate the robustness of the performance metrics, 1,000 bootstrap resamples were generated to calculate the mean, standard deviation, and 95% confidence intervals.

8 Acknowledgements

The authors would like to thank the Vice-Chancellor of the Research Department at North Khorasan University of Medical Sciences for supporting this project.

9 Disclosure Statement

The authors declare that they have no conflict of interest.

10 Reference

1. Liu Z, Wang C, Cai X, Jiang H, Wang J. Discrimination of diabetic retinopathy from optical coherence tomography angiography images using machine learning methods. IEEE access. 2021;9:51689-94.
2. Becker K, Klein H, Simon E, Viollet C, Haslinger C, Leparc G, et al. In-depth transcriptomic analysis of human retina reveals molecular mechanisms underlying diabetic retinopathy. Scientific Reports. 2021;11(1):10494.
3. Chaudhary S, Zaveri J, Becker N. Proliferative diabetic retinopathy (PDR). Disease-a-Month. 2021;67(5):101140.
4. Sun JK, Aiello LP. Nonproliferative and Proliferative Diabetic Retinopathy. Albert and Jakobiec's Principles and Practice of Ophthalmology: Springer; 2022. p. 2955-93.
5. Simó R, Hernández C. New insights into treating early and advanced stage diabetic retinopathy. International journal of molecular sciences. 2022;23(15):8513.
6. Tavakoli M, Mehdizadeh A, Aghayan A, Shahri RP, Ellis T, Dehmeshki J. Automated microaneurysms detection in retinal images using radon transform and supervised learning: application to mass screening of diabetic retinopathy. IEEE Access. 2021;9:67302-14.
7. Arrigo A, Aragona E, Bandello F. VEGF-targeting drugs for the treatment of retinal neovascularization in diabetic retinopathy. Annals of medicine. 2022;54(1):1089-111.
8. Meng Z, Chen Y, Li H, Zhang Y, Yao X, Meng Y, et al. Machine learning and optical coherence tomography-derived radiomics analysis to predict persistent diabetic macular edema in patients undergoing anti-VEGF intravitreal therapy. Journal of Translational Medicine. 2024;22(1):358.
9. Simo R, Hernandez C. Novel approaches for treating diabetic retinopathy based on recent pathogenic evidence. Progress in Retinal and eye Research. 2015;48:160-80.
10. Pandya M, Banait S, Daigavane S. Insights Into Visual Rehabilitation: Pan-Retinal Photocoagulation for Proliferative Diabetic Retinopathy. Cureus. 2024;16(2).
11. Saleh GA, Batouty NM, Haggag S, Elnakib A, Khalifa F, Taher F, et al. The role of medical image modalities and AI in the early detection, diagnosis and grading of retinal diseases: a survey. Bioengineering. 2022;9(8):366.
12. Kanclerz P, Tuuminen R, Khoramnia R. Imaging modalities employed in diabetic retinopathy screening: a review and meta-analysis. Diagnostics. 2021;11(10):1802.
13. Oh K, Kang HM, Leem D, Lee H, Seo KY, Yoon S. Early detection of diabetic retinopathy based on deep learning and ultra-wide-field fundus images. Scientific reports. 2021;11(1):1897.
14. Chua J, Sim R, Tan B, Wong D, Yao X, Liu X, et al. Optical coherence tomography angiography in diabetes and diabetic retinopathy. Journal of Clinical Medicine. 2020;9(6):1723.
15. Tavakoli M, Shahri RP, Pourreza H, Mehdizadeh A, Banaee T, Toosi MHB. A complementary method for automated detection of microaneurysms in fluorescein angiography fundus images to assess diabetic retinopathy. Pattern Recognition. 2013;46(10):2740-53.

16. La Mantia A, Kurt RA, Mejor S, Egan CA, Tufail A, Keane PA, et al. Comparing fundus fluorescein angiography and swept-source optical coherence tomography angiography in the evaluation of diabetic macular perfusion. Retina. 2019;39(5):926-37.
17. Fenner BJ, Wong RL, Lam W-C, Tan GS, Cheung GC. Advances in retinal imaging and applications in diabetic retinopathy screening: a review. Ophthalmology and therapy. 2018;7:333-46.
18. Liu G, Xu D, Wang F. New insights into diabetic retinopathy by OCT angiography. Diabetes Research and Clinical Practice. 2018;142:243-53.
19. Vaz-Pereira S, Morais-Sarmento T, Engelbert M. Update on optical coherence tomography and optical coherence tomography angiography imaging in proliferative diabetic retinopathy. Diagnostics. 2021;11(10):1869.
20. Courtie E, Kirkpatrick JRM, Taylor M, Faes L, Liu X, Logan A, et al. Optical coherence tomography angiography analysis methods: a systematic review and meta-analysis. Scientific Reports. 2024;14(1):9643.
21. Qiao L, Zhu Y, Zhou H. Diabetic retinopathy detection using prognosis of microaneurysm and early diagnosis system for non-proliferative diabetic retinopathy based on deep learning algorithms. IEEE Access. 2020;8:104292-302.
22. Ramezanzadeh E, Zare H, Shoeibi N, Moattar MH, Shariati MM, Dourandeesh M, et al. Precise staging of diabetic retinopathy through machine learning analysis of leakage source characteristics: A Radiomics-driven approach integrating OCT-A and FFA. Photodiagnosis and Photodynamic Therapy. 2026:105522.
23. Tavakoli M, Ramezanzadeh E. Artificial Intelligence Applications in OCT and OCTA for Diabetic Retinopathy: A Systematic Review. Experimental Eye Research. 2026:111046.
24. Grzybowski A, Jin K, Zhou J, Pan X, Wang M, Ye J, et al. Retina fundus photograph-based artificial intelligence algorithms in medicine: a systematic review. Ophthalmology and Therapy. 2024;13(8):2125-49.
25. Rajarajeshwari G, Selvi GC. Application of Artificial Intelligence for Classification, Segmentation, Early Detection, Early Diagnosis, and Grading of Diabetic Retinopathy From Fundus Retinal Images: A Comprehensive Review. IEEE Access. 2024.
26. Ryu G, Lee K, Park D, Kim I, Park SH, Sagong M. A deep learning algorithm for classifying diabetic retinopathy using optical coherence tomography angiography. Translational Vision Science & Technology. 2022;11(2):39-.
27. Soren VN, Prajwal H, Sundaresan V. Automated grading of diabetic retinopathy and Radiomics analysis on ultra-wide optical coherence tomography angiography scans. Image and Vision Computing. 2024;151:105292.
28. Savastano MC, Vagni M, Carlà MM, Tran HE, Fossataro C, Cestrone V, et al. Radiomic Features Extraction from OCT Angiography of Idiopathic Epiretinal Membranes and Correlation with Visual Acuity: A Pilot Study. Ophthalmology Science. 2025:100716.
29. Zhang H, Zhang H, Jiang M, Li J, Li J, Zhou H, et al. Radiomics in ophthalmology: a systematic review. European Radiology. 2025;35(1):542-57.
30. Khajetash B, Talebi A, Bagherpour Z, Abbaspour S, Tavakoli M. Introducing radiomics model to predict active plaque in multiple sclerosis patients using magnetic resonance images. Biomedical Physics & Engineering Express. 2023;9(5):055004.
31. Zhang W, Guo Y, Jin Q. Radiomics and its feature selection: a review. Symmetry. 2023;15(10):1834.
32. Carrera-Escalé L, Benali A, Rathert A-C, Martín-Pinardel R, Bernal-Morales C, Alé-Chilet A, et al. Radiomics-based assessment of OCT angiography images for diabetic retinopathy diagnosis. Ophthalmology Science. 2023;3(2):100259.
33. Sandhu HS, Elmogy M, Sharafeldeen AT, Elsharkawy M, El-Adawy N, Eltanboly A, et al. Automated diagnosis of diabetic retinopathy using clinical biomarkers, optical coherence tomography, and optical coherence tomography angiography. American journal of ophthalmology. 2020;216:201-6.

34. Li X, Wen X, Shang X, Liu J, Zhang L, Cui Y, et al. Identification of diabetic retinopathy classification using machine learning algorithms on clinical data and optical coherence tomography angiography. Eye. 2024:1-9.
35. Aslam TM, Hoyle DC, Puri V, Bento G. Differentiation of diabetic status using statistical and machine learning techniques on optical coherence tomography angiography images. Translational vision science & technology. 2020;9(4):2-.
36. Alam M, Zhang Y, Lim JI, Chan RVP, Yang M, Yao X. QUANTITATIVE OPTICAL COHERENCE TOMOGRAPHY ANGIOGRAPHY FEATURES FOR OBJECTIVE CLASSIFICATION AND STAGING OF DIABETIC RETINOPATHY. Retina. 2020;40(2):322-32.
37. Kar SS, Sevgi DD, Dong V, Srivastava SK, Madabhushi A, Ehlers JP. Multi-compartment spatially-derived radiomics from optical coherence tomography predict anti-VEGF treatment durability in macular edema secondary to retinal vascular disease: preliminary findings. IEEE Journal of Translational Engineering in Health and Medicine. 2021;9:1-13.
38. Atwany MZ, Sahyoun AH, Yaqub M. Deep learning techniques for diabetic retinopathy classification: A survey. IEEE Access. 2022;10:28642-55.
39. Hayati A, Abdol Homayuni MR, Sadeghi R, Asadigandomani H, Dashtkoohi M, Eslami S, et al. Advancing diabetic retinopathy screening: a systematic review of artificial intelligence and optical coherence tomography angiography innovations. Diagnostics. 2025;15(6):737.
40. Nneji GU, Cai J, Deng J, Monday HN, Hossin MA, Nahar S. Identification of diabetic retinopathy using weighted fusion deep learning based on dual-channel fundus scans. Diagnostics. 2022;12(2):540.
41. El-Ateif S, Idri A. Single-modality and joint fusion deep learning for diabetic retinopathy diagnosis. Scientific African. 2022;17:e01280.
42. Anupama B, Rao SN, Malini MB, Athreya VV. Comparative analysis of novel preprocessing techniques and deep learning based multi modal feature fusion for diabetic retinopathy grading. Scientific Reports. 2025.
43. Agarwal A, Raman R, Lakshminarayanan V, editors. The foveal avascular zone image database (fazid). Applications of Digital Image Processing XLIII; 2020: SPIE.
44. Russo A, Morescalchi F, Costagliola C, Delcassi L, Semeraro F. Comparison of smartphone ophthalmoscopy with slit-lamp biomicroscopy for grading diabetic retinopathy. American journal of ophthalmology. 2015;159(2):360-4. e1.
45. Abrishami M, Hassanpour K, Hosseini S, Emamverdian Z, Ansari-Astaneh M-R, Zamani G, et al. Macular vessel density reduction in patients recovered from COVID-19: a longitudinal optical coherence tomography angiography study. Graefe's Archive for Clinical and Experimental Ophthalmology. 2022:1-9.
46. Heidarzadeh HR, Abrishami M, Ebrahimi Miandehi E, Shoeibi N, Ansari Astaneh MR, Hosseini SM, et al. The central retina vessel density and foveal avascular zone values of 792 healthy adults using optical coherence tomography angiography. Eye. 2024;38(18):3434-42.
47. Astaneh MA, Gharbi NS, Abrishami M, Bakhtiari E, Shariati MM. Investigating the Macular Vasculature Using Optical Coherence Tomography Angiography Following Pan-Retinal Photocoagulation in Diabetic Retinopathy. Frontiers in Biomedical Technologies. 2025.
48. Naz H, Ahuja NJ. A novel contrast enhancement technique for diabetic retinal image pre-processing and classification. International Ophthalmology. 2024;45(1):11.
49. Ramezanzadeh E, Shoeibi N, Rabbani H, Hosseini SM, Astaneh MA, Tavakoli M, et al. A high-accuracy hybrid method for detecting retinal blood vessel changes across different phases of fluorescein angiography in diabetic retinopathy patients. Biomedical Physics & Engineering Express. 2025;11(5):055016.
50. Chen S-D, Ramli AR. Contrast enhancement using recursive mean-separate histogram equalization for scalable brightness preservation. IEEE Transactions on consumer Electronics. 2003;49(4):1301-9.
51. Patil SB, Patil B. Retinal fundus image enhancement using adaptive CLAHE methods. Journal of Seybold Report ISSN NO. 2020;1533:9211.

52. Khanbabaei H, Mohammadi S, Sadoughi H-R. Applications of wavelet transforms in radiomic feature extraction from medical images: a systematic review. Journal of Research and Health. 2025;15(6):0-.
53. Nagpal D, Alsubaie N, Soufiene BO, Alqahtani MS, Abbas M, Almohiy HM. Automatic detection of diabetic hypertensive retinopathy in fundus images using transfer learning. Applied Sciences. 2023;13(8):4695.
54. Zhang K, Zuo W, Chen Y, Meng D, Zhang L. Beyond a gaussian denoiser: Residual learning of deep cnn for image denoising. IEEE transactions on image processing. 2017;26(7):3142-55.
55. Zhao Y, Li Y, Dong X, Yang B. Low-frequency noise suppression method based on improved DnCNN in desert seismic data. IEEE Geoscience and Remote Sensing Letters. 2018;16(5):811-5.
56. Che Azemin MZ, Ab Hamid F, Wang JJ, Kawasaki R, Kumar DK, editors. Box-counting fractal dimension algorithm variations on retina images. Advanced Computer and Communication Engineering Technology: Proceedings of ICOCOE 2015; 2016: Springer.
57. Alalhareth M, Hong S-C. An improved mutual information feature selection technique for intrusion detection systems in the Internet of Medical Things. Sensors. 2023;23(10):4971.
58. Mohammadi S, Zuckerman N, Goldsmith A, Grama A. A critical survey of deconvolution methods for separating cell types in complex tissues. Proceedings of the IEEE. 2016;105(2):340-66.
59. Ranek JS. Resolving Biological Trajectories in Single-Cell Data Using Feature Selection and Multi-Modal Integration: The University of North Carolina at Chapel Hill; 2023.
60. Khan FA, Agrawal VL. Classification of diabetic retinopathy using multilayer perceptron neural network. International Journal of Electronics, Communication and Soft Computing Science & Engineering (IJECSCSE). 2017:148-52.
61. Washburn PS. Investigation of severity level of diabetic retinopathy using adaboost classifier algorithm. Materials Today: Proceedings. 2020;33:3037-42.
62. Sikder N, Masud M, Bairagi AK, Arif ASM, Nahid A-A, Alhumyani HA. Severity classification of diabetic retinopathy using an ensemble learning algorithm through analyzing retinal images. Symmetry. 2021;13(4):670.
63. Wolpert DH. Stacked generalization. Neural networks. 1992;5(2):241-59.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Table1.docx
- Table2.docx
- S1stability.docx