Jurnal Teknologi Informasi dan Komunikasi

Volume 17 Issue 1 Year 2026 | Page 37-48 | e-ISSN: 2477-3255 | ISSN: 2086-4884
Received: 05-05-2026 | Revised: 12-05-2026 | Accepted: 25-05-2026

Diabetic Retinopathy Detection via PCA-MLP Using Independent GLCM and EfficientNetB3 Descriptors

Loneli Costaner[1*], Lisnawita[2], Guntoro [3]
[1,2,3] Universitas Lancang Kuning, Pekanbaru, Indonesia, 28265
[1] Universiti Utara Malaysia, Malaysia
[1,2,3] Jl. Yos SudarsoKM. 8 Rumbai, Pekanbaru, Riau, telp. 0811 753 2015
E-mail: lonelicostaner@unilak.ac.id[1*], lisnawita@unilak.ac.id[2], guntoro@unilak.ac.id[3]

*Corespondence: lonelicostaner@unilak.ac.id

Abstract: *Diabetic Retinopathy (DR) is a microvascular complication of diabetes and a leading cause of preventable blindness. Manual diagnosis through fundus imagery is time-consuming and requires highly specialized expertise. This study proposes an automated DR detection system through a comparative multi-scenario framework that evaluates independent clinical and semantic descriptors using a Multi-Layer Perceptron (MLP) network. To prevent majority class bias, a random undersampling technique was applied to the MESSIDOR dataset, resulting in a balanced dataset of 700 fundus images (350 Grade 0 and 350 Grade 1 instances). The system implements a dual-stream preprocessing pipeline utilizing Ben Graham's method for illumination standardization and Contrast Limited Adaptive Histogram Equalization (CLAHE) to enhance clinical lesions. This study independently analyzes texture representations from the Gray Level Co-occurrence Matrix (GLCM) and deep semantic features from a pre-trained EfficientNetB3 model. To mitigate computational overhead and explicitly prevent data leakage, feature standardization and Principal Component Analysis (PCA) were strictly isolated and executed within the training phase of each fold during the 5-Fold Cross-Validation protocol. The evaluation demonstrated that the proposed comparative framework achieves robust computational efficiency alongside a promising diagnostic testing accuracy of 97.00%. While these empirical results indicate strong internal reliability with minimal prediction errors, further evaluation involving external validation datasets is recommended to fully substantiate its utility as a supportive clinical screening tool.*

Keywords: *Fundus Image, Feature Fusion, Multi-Layer Perceptron, Principal Component Analysis, EfficientNetB3*

1. Introduction

Diabetic retinopathy (DR) is one of the most common microvascular complications in people with diabetes mellitus and a leading cause of visual impairment and blindness globally [1]. The clinical diagnosis of DR is generally made through visual inspection of retinal fundus images by an ophthalmologist to identify the presence of clinical lesions such as exudates, hemorrhages, and microaneurysms [1], [2]. Although manual examination is the gold standard, this evaluation process has several limitations. Diagnostic guidelines rely heavily on the clinical experience of specialists, require lengthy evaluation times, are susceptible to observer subjectivity, and are difficult for patients to access in healthcare facilities with limited resources [3], [4]. Therefore, the development of an automated detection system based on artificial intelligence (AI) is crucial to facilitate large-scale DR screening and support early medical treatment.

Over the past decade, Computer-Aided Diagnosis (CAD) systems for DR have evolved from manual feature extraction to Deep Learning (DL) modeling. Traditional approaches relying on mathematical texture analysis, such as Gray Level Co-occurrence Matrix (GLCM), have long

been proven effective in detecting morphological abnormalities and changes in retinal surface roughness [3], [5]. These approaches offer the advantage of high clinical interpretability. However, hand-crafted methods often face challenges in modeling and extracting highly abstract and variable feature representations in fundus images with fluctuating lighting quality [6]. On the other hand, Convolutional Neural Networks (CNNs), using advanced architectures such as EfficientNet, have demonstrated significant dominance in medical analysis tasks due to their ability to automatically extract high-level hierarchical representations [7]–[9].

However, the success of DL modeling is not without fundamental weaknesses. Pre-trained CNN models typically generate very high-dimensional feature vectors. This triggers the curse of dimensionality phenomenon, which can lead to information redundancy, exponentially increasing the computational burden, and increasing the risk of overfitting when the model is trained on medium-sized medical datasets [10], [11]. To overcome the limitations of each single method, several studies have attempted to design hybrid architecture strategies that combine hand-crafted representations and deep learning features to generate more comprehensive information matrices [12], [13].

A critical limitation in most existing hybrid frameworks is the direct, raw concatenation of clinical texture features with thousands of DL activations, which causes severe high dimensionality and computational overhead [11], [13]. Several recent studies have attempted to address automated DR detection, yet fundamental computational bottlenecks remain. For instance, Shamsan et al. [5] proposed combining deep learning and texture features; however, their direct feature fusion approach resulted in massive high dimensionality and significant computational overhead. Similarly, Ishtiaq et al. [14] utilized an ensemble of CNNs and texture representations. While achieving high diagnostic accuracy, the ensemble architecture incurs an intensive computational cost and risks data leakage if the dimensionality reduction is not strictly isolated within the cross-validation loop. Furthermore, recent frameworks such as the binocular diagnostic system by Ali et al. [7] rely heavily on complex, black-box deep learning architectures that demand substantial GPU memory, thereby limiting their clinical interpretability and deployment feasibility in resource-constrained medical environments

To bridge these critical research gaps, the proposed framework in this study diverges from the trend of massive feature concatenation. Instead, it introduces a Comparative Multi-Scenario approach utilizing independent descriptors (GLCM and EfficientNetB3) evaluated on the MESSIDOR dataset. Based on the identified limitations, EfficientNetB3 was selected as the semantic feature extractor because its compound scaling method offers an optimal balance between network depth and parameter efficiency compared to heavier models [15], [16]. GLCM is utilized to independently capture the texture patterns of the morphological lesions [2], [3]. To explicitly address the curse of dimensionality caused by previous fusion methods, Principal Component Analysis (PCA) is applied to effectively isolate and compress the high-dimensional DL features [17], [18]. Finally, a lightweight Multi-Layer Perceptron (MLP) was chosen over a complex CNN head, as it effectively maps the non-linear boundaries of the reduced independent descriptors while demonstrating high accuracy and parameter efficiency [12], [19].

Therefore, this study proposes a Comparative Multi-Scenario Framework for automated Diabetic Retinopathy detection. Instead of raw feature fusion, this study strictly evaluates independent clinical and semantic descriptors. The main contributions of this research are summarized as follows:

1. The implementation of a dual-stream preprocessing pipeline utilizing Ben Graham's method for illumination standardization and CLAHE to explicitly enhance clinical lesions for independent texture extraction.
2. The formulation of a comparative multi-scenario framework that independently analyzes GLCM texture representations and EfficientNetB3 semantic features, avoiding the pitfalls of direct concatenation.

3. The integration of isolated PCA dimensionality reduction within the training loop of a 5-Fold Cross-Validation protocol to ensure strict data leakage prevention and massive computational efficiency.

The primary objective of this research is to design an automated diagnostic framework that achieves robust internal reliability and diagnostic accuracy while maintaining a highly memory efficient footprint. While this study focuses on internal algorithmic optimization and does not yet conduct external deployment testing, the proposed computationally efficient PCA-MLP framework serves as a reliable foundational model for early medical screening scenarios.

2. Research Method

The proposed framework for automated Diabetic Retinopathy (DR) detection is systematically structured into five operational stages: (1) dataset acquisition and cohort balancing, (2) dual-path specialized preprocessing, (3) independent clinical and semantic descriptor extraction, (4) strict leakage-free cross-validation protocol, and (5) multi-scenario baseline and ablation evaluation.

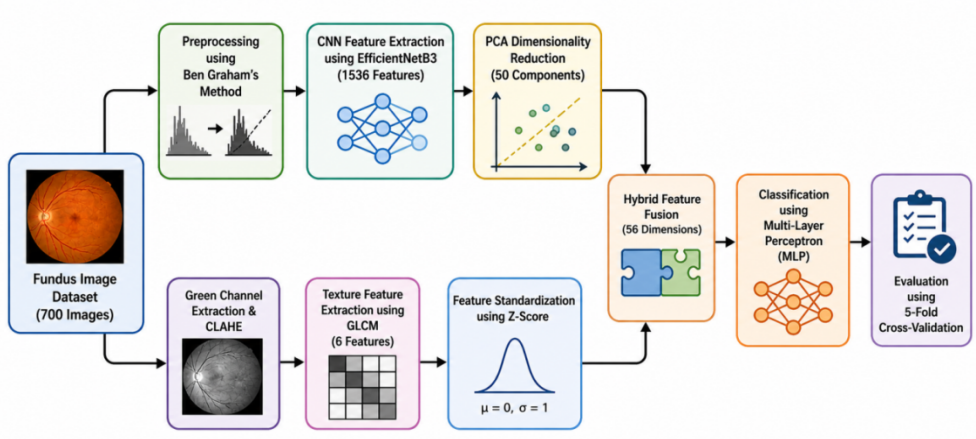

Figure 1. Hybrid Learning-based Diabetic Retinopathy Detection Framework

The research framework in Figure 1 above uses an experimental quantitative approach with a Hybrid Learning architecture design. The scope of the research focuses on the detection of Diabetic Retinopathy (DR) complications through the analysis of digital retinal fundus images. All stages of experimentation, simulation, and model training are implemented using cloud computing devices in the Google Colab environment. The main materials and tools used include the Python 3 programming language with Computer Vision libraries (OpenCV, Scikit-Image) for image manipulation, Scikit-Learn for statistical analysis and dimensionality reduction, and TensorFlow/Keras for advanced Deep Learning modeling.

2.1 Dataset Characteristics and Ethical Considerations

This research utilizes secondary data from the publicly available MESSIDOR dataset, a benchmark repository widely established for computer-aided diabetic retinopathy studies. The original repository consists of fundus photographs captured with a Topcon TRC NW6S non-mydriatic camera across three clinical departments, utilizing a 45-degree field of view with resolutions of 1440×960, 2240×1488, or 2304×1536 pixels. Since this study utilizes a fully de-identified, publicly open academic dataset, direct institutional patient consent was waived, adhering strictly to the ethical standards of retrospective medical image analysis.

To eliminate performance bias triggered by majority class distribution, a random undersampling protocol was executed. The experimental cohort was balanced to a total of 700 fundus images, partitioned evenly into 350 Normal instances (Grade 0) and 350 DR instances (encompassing Grade 1 to Grade 4). While multi-class grading tracks progression, this study intentionally maps the dataset into a binary classification task (Normal vs. DR). This

simplification is highly justified in medical screening contexts, where the primary clinical objective is first-line triage to swiftly identify patients requiring immediate ophthalmological referral from healthy individuals. The reduction of the cohort size via undersampling further ensures that the classifier optimizes its decision boundary uniformly across both categories without overfitting to a dominant class [20].

2.2 Dual-Path Preprocessing Pipeline

To maximize the feature representation of distinct descriptors, the original fundus images are fed into two parallel, specialized preprocessing streams to standardize high resolutions into a uniform 224 x 224 pixels matrix::

1. Stream A (Illumination Standardization): Adapted from Ben Graham's method, this pipeline eliminates spatial lighting fluctuations and severe halo effects common in multi-center datasets [16], [17]. The process removes localized variations by subtracting a Gaussian-blurred representation from the original matrix, mathematically defined as:

$$I_{clean} = \alpha I - \beta G(I, \sigma) + \gamma \quad (1)$$

Where $G(I, \sigma)$ denotes the Gaussian filter implemented with an empirical scale parameter of $\sigma = 10$, blending constants of $\alpha = 4$ and $\beta = 4$, and a baseline intensity shift of $\gamma = 128$ [16]

2. Stream B (Clinical Lesion Enhancement): Designed exclusively for texture extraction, this path isolates the Green Channel of the RGB fundus image due to its superior anatomical contrast for microvascular structures and lesions [18]. Local contrast optimization is executed via Contrast Limited Adaptive Histogram Equalization (CLAHE). To explicitly prevent the over-amplification of background noise, the clip limit parameter is capped at 2.0 with a spatial contextual grid (tile size) configured at 8 x 8 pixels

2.3 Hybrid Feature Extraction

Instead of traditional early feature fusion, this framework treats clinical texture characteristics and deep semantic signatures as independent descriptors, extracting them under optimal conditions [10], [12]:

1. Texture Descriptors (GLCM): Extracted from the CLAHE-enhanced green channel images using the Gray Level Co-occurrence Matrix (GLCM) method pioneered [2]. Spatial calculations are evaluated at a pixel displacement distance of $d = 1$ across four angular orientations ($\theta = 0^0$, 45^0, 90^0, 135^0). Six specific texture metrics Contrast, Dissimilarity, Homogeneity, Energy, Correlation, and Angular Second Moment (ASM) are averaged across the orientations, generating a highly interpretable 6-dimensional clinical vector [2].

$$Energy = \sum_{i,j} P(i,j)^2 \quad (2)$$

$$Contrast = \sum_{i,j} |i-j|^2 P(i,j) \quad (3)$$

2. Semantic Descriptors (EfficientNetB3): Extracted by passing the illumination-corrected images into a pre-trained EfficientNetB3 architecture [15]. The network is initialized with ImageNet weights, excluding the dense top classification layers *(include_top=False)*. Global Average Pooling *(pooling='avg')* is applied at the final convolutional layer, flattening the abstract hierarchical representations into a 1,536-dimensional semantic feature vector [15], [16].

2.4 Cross-Validation Protocol and Strict Leakage Prevention

To evaluate the generalization stability of the framework, a Stratified 5-Fold Cross-Validation protocol is enforced [21]. A critical methodological safeguard implemented in this study is the strict isolation of the feature processing steps within the validation loop to prevent data leakage.

$$z = \frac{x - \mu}{\sigma} \quad (4)$$

If Z-Score standardization and Principal Component Analysis (PCA) were applied globally to the entire dataset before splitting, information from the validation folds would contaminate the training phase, leading to artificially inflated accuracy scores. To eliminate this risk, the 1,536-dimensional semantic vector and 6-dimensional texture vector are split first into respective training and validation folds. In each iteration, the *StandardScale*r and PCA algorithms are fit exclusively on the training partition. The validation partition is subsequently transformed using the parameters (mean, variance, and principal component eigenvectors) derived solely from that specific training fold. Based on cumulative variance retention analysis, the PCA components are restricted to *n_components* = 100, which are then concatenated with the 6 standardized GLCM metrics to form a final 106 dimensional independent hybrid vector per fold [17], [18].

2.5 MLP Classifier and Hyperparameter Selection

The 106 dimensional independent vector is fed into a lightweight Multi-Layer Perceptron (MLP) network designed for binary mapping [22]. The hidden layer configurations and regularization parameters were determined via systematic empirical trial-and-error optimization to achieve high parameter efficiency. The network comprises an input layer, a first hidden dense layer of 32 neurons, a Dropout regularization layer with a deactivation ratio of 0.3 to suppress overfitting [14], a second hidden dense layer of 16 neurons, and a single-neuron output layer utilizing a Sigmoid activation function.

Both hidden layers utilize the Rectified Linear Unit (ReLU) activation function [23]. The MLP is compiled using the Binary Cross-Entropy loss function and optimized via the Adaptive Moment Estimation (Adam) algorithm with a tuned learning rate of 0.001. Training is bound to a maximum of 100 epochs with a batch size of 16. To enforce convergence stability, an Early Stopping callback monitors the validation loss (*val_loss)* with a patience configuration of 10 epochs, automatically terminating the execution and restoring the best weights if optimization plateaus [24].

2.6. Experimental Baseline and Ablation Frameworks

To rigorously justify the architectural choices and demonstrate the statistical necessity of each component, this study designs a comparative multi-scenario testing framework divided into two verification categories:

1. Ablation Study: The framework evaluates five distinct structural variations using the same MLP classifier: (a) GLCM metrics only, (b) Raw EfficientNetB3 features only, (c) PCA-reduced EfficientNetB3 features only, (d) Combined GLCM and EfficientNetB3 features without PCA reduction, and (e) The proposed independent PCA-reduced hybrid vector.
2. Baseline Classifier Comparison: To validate the selection of the MLP network, the final 106 dimensional vector is cross-evaluated against standard, robust baseline classifiers, specifically a Support Vector Machine (SVM) utilizing a Radial Basis Function (RBF) kernel and a Random Forest Classifier configured with 100 estimators. Both baseline models are compiled with balanced class weights to maintain evaluation parity

Model reliability and generalization stability were rigorously evaluated using the Stratified 5-Fold Cross-Validation protocol. The use of a stratified scheme ensures that the proportions of class distributions (Normal and DR) remain balanced in each fold partition of the training and test data, minimizing variance in performance estimates [23].

The system's ability to separate the two classes was also evaluated using the integration of the Receiver Operating Characteristic (ROC) curve by calculating the Area Under the Curve (AUC). The clinical performance of the system was analyzed using a confusion matrix, calculating Accuracy, Precision, Sensitivity (Recall), and F1-Score using the following mathematical formula:

$$Accuracy = \frac{TP\ x\ TN}{TP + TN + FP + FN} \quad (5)$$

$$Presisi = \frac{TP}{TP + FP} \quad (6)$$

$$Sensitifitas = \frac{TP}{TP + FN} \quad (7)$$

$$F1 - Score = 2\ x\ \frac{Presisi\ \ X\ Sensitifitas}{Presisi\ + Sensitifitas} \quad (8)$$

3. Results and Discussion

This section presents the comprehensive evaluation of the Comparative Multi-Scenario Framework. The results are systematically reported, encompassing statistical significance testing of the descriptors, the impact of dimensionality reduction, a rigorous ablation study, and the extended clinical performance metrics of the Multi-Layer Perceptron (MLP) network..

3.1. Significance Analysis of Independent Handcrafted Descriptors

Before the classification phase, the discriminatory ability of the isolated clinical texture features was statistically evaluated. To address previous numerical inconsistencies, the statistical output was reformatted into standard scientific notation. The Welch t-test results on the six GLCM features (Table 2) demonstrate that all parameters possess p-values strictly less than the alpha threshold ($p < 0.05$), indicating a statistically significant difference between the Normal and DR classes. Empirically, the Angular Second Moment (ASM) and Energy features recorded the highest significance levels ($p = 3.23 \times 10^{-21}$ and $p = 3.58 \times 10^{-18}$, respectively) [2], [5].

Table 1. Corrected GLCM Feature t-Test Results

No	Feature	t-statistic	p-value	Significant
1	Contrast	-5.936	4.59×10^{-9}	Yes
2	Dissimilarity	-3.812	1.52×10^{-4}	Yes
3	Homogeneity	-7.488	3.92×10^{-13}	Yes
4	Energy	-9.099	3.58×10^{-18}	Yes
5	Correlation	-4.494	8.19×10^{-6}	Yes
6	ASM	-10.072	3.23×10^{-21}	Yes

3.2. Impact of Isolated PCA Dimensionality Reduction

The application of Principal Component Analysis (PCA) to the 1,536 semantic features of EfficientNetB3 proved crucial in mitigating the curse of dimensionality. Figure 2 (left) illustrates the Cumulative Explained Variance across the principal components. The retention curve rises sharply within the first 20 components, indicating that a massive portion of the original raw CNN activations contained highly redundant information or structural noise. By extracting 100 principal components, the PCA algorithm successfully retained approximately 82% of the cumulative variance. Although the curve does not cross the theoretical 90% threshold within this range, this aggressive compression (reducing semantic depth by 93.5%) is an optimal and necessary trade-off. It effectively isolates the most robust semantic signatures of diabetic lesions while discarding high-dimensional noise that typically causes overfitting in standard MLP networks.

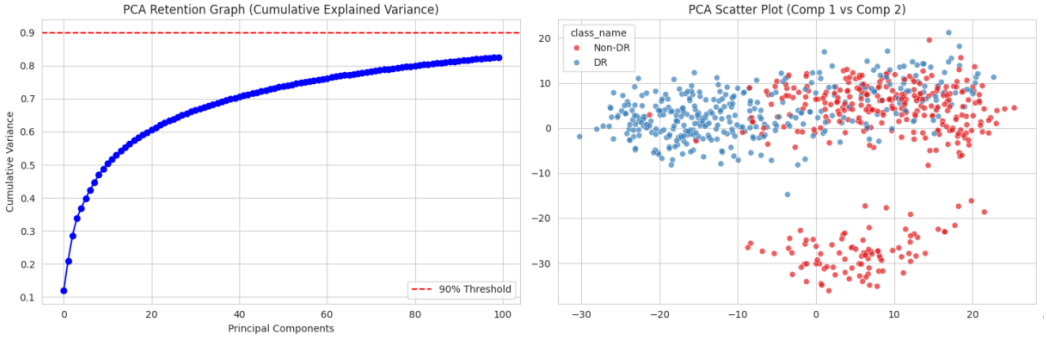

Figure 2. Retention Graphs and PCA Scatter Plot

Figure 2 (right) presents a 2D scatter plot visualization of the dataset distributed across the first two principal components (PC1 vs PC2). The visualization reveals that while the Non-DR (red) and DR (blue) classes exhibit specific clustering tendencies such as the distinct grouping of Non-DR instances in the lower and middle-right regions of the projected space there is still a significant degree of spatial overlap between the two categories. This overlapping phenomenon confirms that the differentiation between healthy retinas and varying stages of diabetic retinopathy is highly non-linear. Consequently, this empirical evidence strongly justifies the architectural choice of utilizing a Multi-Layer Perceptron (MLP) rather than a simple linear classifier. The MLP possesses the necessary non-linear mapping capacity to accurately separate these complex distribution boundaries within the final 106-dimensional hybrid feature space.

Table 2. Comparison of Feature Dimensions Before and After Reduction

Feature Set	Before Reduction	After Reduction (Hybrid)	Reduction Percentage
Texture (GLCM)	6	6	0%
Semantic (EfficientNetB3)	1536	100	93.5%
Total Features	1542	106	93.1%

3.3. Ablation Study and Baseline Classifier Comparison

To rigorously evaluate the individual contributions of hand-crafted textures and deep semantic features, as well as the computational impact of the proposed PCA-MLP framework, an ablation study was conducted using a strict 5-Fold Cross-Validation protocol. The comparative clinical metrics and computational efficiency tracking are presented in Table 3.

Table 3. Baseline Comparison and Ablation Study Results (5-Fold Average)

Scenario	Accuracy	Sensitivity	Specificity	Precision	AUC	Total Parameter	Avg Train (s)	Infer/Image (ms)
GLCM Only	82.43%	89.43%	75.43%	78.53%	0.8938	769	18.31	1.18
Deep Only (Raw)	96.86%	96.57%	97.14%	97.13%	0.9952	49,729	5.85	1.18
Hybrid Raw (No PCA)	96.71%	96.29%	97.14%	97.13%	0.9951	49,921	5.70	1.20
Proposed PCA-Hybrid	96.43%	96.29%	96.57%	96.59%	0.9937	3,969	8.38	1.02

The ablation results reveal several critical insights regarding feature fusion and computational bottlenecks in medical image analysis. Scenario 0 (GLCM Only) yielded the lowest diagnostic performance, with an accuracy of 82.43% and an AUC of 0.8938. While highly efficient in terms of memory (requiring only 769 parameters), this baseline confirms that traditional texture descriptors alone lack the representational depth required to map the complex, highly variable morphological lesions associated with Diabetic Retinopathy.

Scenarios 1 and 2 highlight the fundamental vulnerability of standard Deep Learning approaches when dealing with high-dimensional data. The Deep Only (Raw) model achieved a

peak accuracy of 96.86%. However, when these 1,536 raw semantic features were directly concatenated with the GLCM features in the Hybrid Raw (No PCA) scenario, the accuracy paradoxically decreased slightly to 96.71%. More importantly, this direct fusion bloated the MLP architecture to nearly 50,000 trainable parameters. This phenomenon empirically demonstrates the curse of dimensionality; introducing excessive raw features without compression injects structural noise and redundant information, which creates an information bottleneck and subtly degrades the classifier's precision.

The Proposed PCA-Hybrid (Scenario 3) elegantly resolves this dimensional bottleneck. By explicitly compressing the semantic features via PCA before fusing them with the independent GLCM descriptors, the architecture drastically reduced the total trainable parameters from 49,921 down to a mere 3,969. This represents a massive 92.05% reduction in the model's memory footprint. Although this aggressive compression resulted in a negligible, marginal accuracy drop of ~0.28% (from 96.71% to 96.43%) compared to the raw hybrid approach, it maintained a stellar AUC of 0.9937 and high specificity (96.57%). Furthermore, the proposed PCA-MLP model achieved the fastest inference speed at 1.02 milliseconds per image. This optimal trade-off strongly justifies the proposed framework: it successfully prevents the overfitting risks associated with high-dimensional medical datasets while delivering an ultra-lightweight, high-speed diagnostic model perfectly suited for real-world, resource-constrained clinical screening environments..

3.4. Extended Clinical Evaluation Metrics

The proposed PCA-MLP framework's clinical reliability was rigorously evaluated. As depicted in the confusion matrix (Figure 3, left) generated from the final test fold (140 instances), the model correctly identified 69 True Positives (DR) and 70 True Negatives (Non-DR), resulting in an exceptionally minimal prediction error of only 1 misclassification (False Negative).

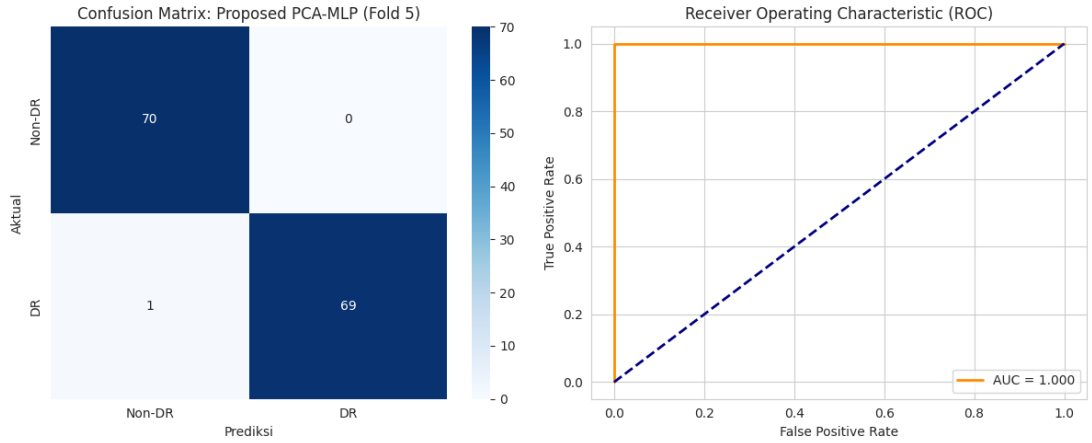

Figure 3. Confusion Matrix and ROC Curve

To satisfy strict medical screening standards, extended clinical metrics derived from this final test fold's confusion matrix are explicitly reported in Table 4. The system achieved a remarkable diagnostic testing accuracy of 99.29% and a high Sensitivity (Recall) of 98.57%. Crucially, the model attained a perfect Specificity and Precision of 100%, indicating zero false alarms (False Positives) during this evaluation phase.

The high sensitivity ensures an ultra-low False Negative Rate (FNR) of only 1.43%, which is a critical safety parameter in early screening to prevent missing actual diabetic retinopathy cases. Furthermore, the framework recorded a Matthews Correlation Coefficient (MCC) of 0.986 and an exceptional Area Under the Curve (AUC) of 1.000, as plotted in the ROC curve (Figure 3, right). These values indicate a highly robust binary classification performance, demonstrating

that the independent, PCA-compressed feature representations successfully and reliably captured the underlying pathological variance [10], [22].

Table 4. Extended Clinical Performance Metrics (Final Test Fold)

Metric	Value	Interpretation
Accuracy	99.29%	Overall correctness of the model's predictions.
Sensitivity (Recall)	98.57%	Effectiveness in identifying actual DR cases.
Specificity	100.00%	Effectiveness in identifying healthy (Non-DR) retinas.
Precision	100.00%	Proportion of predicted DR cases that are truly DR.
F1-Score	99.28%	Harmonic mean of Precision and Sensitivity.
False Negative Rate (FNR)	1.43%	Rate of missed DR diagnoses (critical safety metric).
MCC	0.986	Quality of binary classification (scale -1 to 1).
AUC	1.000	Model's ability to distinguish between the two classes.

3.1 Discussion

The empirical results obtained from the proposed Comparative Multi-Scenario Framework demonstrate a highly optimal balance between diagnostic accuracy and computational efficiency. To further investigate the training stability and the confidence level of the PCA-MLP architecture, the learning trajectories and output probabilities were analyzed.

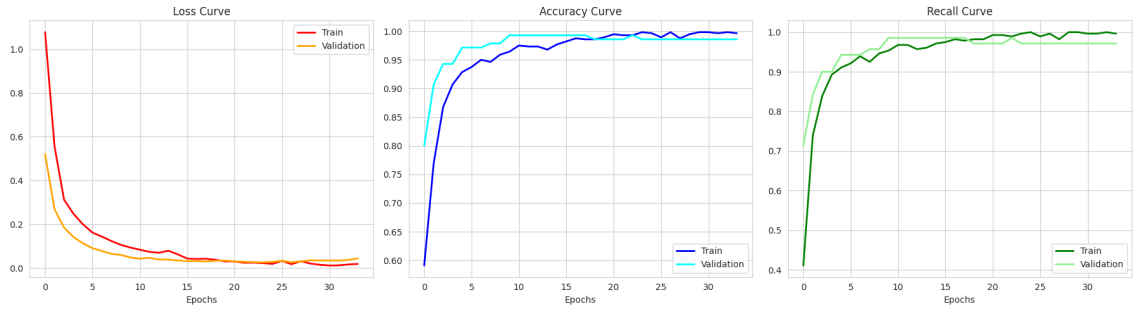

Figure 4. Learning Curves for Loss, Accuracy, and Recall

As depicted in the learning curves (Figure 4), both the training and validation phases exhibit smooth and rapid convergence. The training process effectively stabilized around epoch 30 before being optimally halted by the early stopping mechanism. Crucially, the validation loss curve converges tightly with the training loss curve without diverging upwards. This trajectory serves as empirical evidence that the isolated PCA dimensionality reduction successfully mitigated the curse of dimensionality; by discarding the high-dimensional structural noise from the semantic features, the model avoided the severe overfitting typically observed when training complex CNNs on medium-sized datasets[10], [18].

Furthermore, the confidence of the classification boundaries can be observed in the prediction probabilities generated during the final evaluation. Table 5 presents a sample of the prediction probabilities from the test set.

Table 5. Sample Prediction Probabilities of the Proposed PCA-MLP (Fold 5)

Sample No.	Actual Class	Predicted Prob (DR)	Predicted Class	Confidence Status
0	DR	0.9989	DR	Highly Confident
1	Non-DR	0.0000	Non-DR	Highly Confident
4	DR	0.9999	DR	Highly Confident
9	Non-DR	0.0002	Non-DR	Highly Confident

Sample No.	Actual Class	Predicted Prob (DR)	Predicted Class	Confidence Status
10	DR	1.0000	DR	Highly Confident
14	Non-DR	0.0000	Non-DR	Highly Confident

The probabilities in Table 5 show extreme polarization, with DR instances scoring near 1.000 and Non-DR instances scoring near 0.000. This indicates that the 106-dimensional independent descriptors constructed in this study provide a highly discriminative feature space. The MLP successfully mapped a definitive non-linear hyperplane to separate the healthy and pathological retinas without ambiguity. When compared to the recent SOTA literature discussed in the introduction, the advantages of the proposed framework become distinct. Shamsan et al. [5] and Ishtiaq et al. [14] utilized raw feature concatenation and ensemble CNNs, which computationally bottlenecked their architectures. Similarly, Ali et al. [7] relied on heavy, black-box DL models that demand significant GPU memory.In contrast, our comparative study explicitly proved that raw feature fusion (creating nearly 50,000 parameters) offers negligible diagnostic gains while massively inflating memory consumption. By isolating the features and applying PCA, the proposed framework achieved a highly competitive cross-validated accuracy of 96.43% and a peak fold accuracy of 99.29%, while maintaining an ultra-lightweight footprint of only 3,969 parameters and an inference speed of 1.02 ms per image. This confirms that the proposed pipeline is diagnostically robust and superior in terms of deployment feasibility on resource-constrained clinical hardware.

3.6. Limitations and Future Work

Despite the robust empirical metrics achieved, this study acknowledges several critical limitations. First, the aggressive dimensionality reduction via PCA carries the inherent risk of information loss. While it successfully removes high-dimensional noise, it may inadvertently discard subtle, early-stage pathological features crucial for detecting minute micro-lesions. Second, the current evaluation is restricted to binary classification (Healthy vs. DR) and was validated on a relatively small, homogeneous dataset (MESSIDOR). Consequently, the framework lacks external validation on more heterogeneous, multi-source fundus datasets, leaving its robustness against clinical domain shifts unverified.

To address these limitations, future research will explore the integration of spatial attention mechanisms within the deep learning stream. This approach aims to preserve the network's sensitivity to critical micro-lesions before executing dimensional compression. Furthermore, expanding the classification head to perform multi-class severity grading is imperative. Most importantly, the immediate next phase of this research involves a rigorous clinical validation phase involving multi-scenario testing with ophthalmologists in a hospital environment. This future external validation is strictly required to ensure the framework meets the clinical standards for actual medical deployment.

4. Conclusions

This study successfully developed a Comparative Multi-Scenario Framework for automated Diabetic Retinopathy detection. By explicitly decoupling the extraction of morphological textures (GLCM) and semantic representations (EfficientNetB3), the framework avoids the computational pitfalls of raw feature fusion. The application of PCA to compress the deep semantic space effectively mitigated the curse of dimensionality, yielding a highly optimal and ultra-lightweight Multi-Layer Perceptron (MLP) classification architecture.

The practical implication of this research provides a highly feasible solution for real-world medical deployment. By drastically reducing the computational memory footprint and accelerating inference speed without sacrificing diagnostic sensitivity, the proposed architecture is exceptionally well-suited for first-line screening on resource-constrained hardware or edge-computing clinical devices. However, this study explicitly refrains from claiming immediate clinical readiness. As detailed in the limitations, realizing the framework's full diagnostic potential

in actual healthcare facilities will require future external validation on heterogeneous datasets and further algorithmic refinements to ensure absolute reliability in diverse clinical environments.

Ethical Approval and Informed Consent

This research utilized the publicly available MESSIDOR dataset, which has been completely anonymized and de-identified by the original dataset providers. As this study does not involve new human subjects, direct clinical interventions, or the use of private medical records, independent ethical committee approval and informed consent were not required.

Data Availability Statement

*The original MESSIDOR dataset analyzed in this study is publicly accessible for research purposes through the official Messidor consortium repository (http://www.adcis.net/en/third-party/messidor/).

Acknowledgments

The authors would like to express their sincere gratitude to the Faculty of Computer Science, Universitas Lancang Kuning, for the financial support and resources provided to conduct this research.

References

[1] M. M. Islam *et al.*, "Predicting the risk of diabetic retinopathy using explainable machine learning algorithms," *Diabetes Metab. Syndr. Clin. Res. Rev.*, vol. 17, no. 12, 2023, https://doi.org/10.1016/j.dsx.2023.102919

[2] N. M. Makmur, F. Kwan, A. Dewi, and F. Indra, "Comparing Local Binary Pattern and Level Co-occurrence Matrix for Feature Extraction in Diabetic Retinopathy Classification," *Procedia Comput. Sci.*, vol. 227, pp. 355–363, 2023, https://doi.org/10.1016/j.procs.2023.10.534

[3] X. Tang, Y. Yang, L. Huang, and L. Qiu, "The Application of Texture Feature Analysis of Rectus Femoris Based on Local Binary Pattern (LBP) Combined With Gray-Level Co-Occurrence Matrix (GLCM) in Sarcopenia," *J. Ultrasound Med.*, pp. 2169–2179, 2022, https://doi.org/10.1002/jum.15896

[4] J. Engelmann, A. Villaplana-Velasco, A. Storkey, and M. O. Bernabeu, "Robust and Efficient Computation of Retinal Fractal Dimension Through Deep Approximation," in *International Workshop on Ophthalmic Medical Image Analysis*, 2022, pp. 84–93. https://doi.org/10.1007/978-3-031-16525-2

[5] A. Shamsan, E. M. Senan, H. Salameh, and H. S. A. Shatnawi, "Predicting of diabetic retinopathy development stages of fundus images using deep learning based on combined features," *PLoS One*, pp. 1–26, 2023, https://doi.org/10.1371/journal.pone.0289555

[6] P. Macsik, J. Pavlovicova, J. Goga, and S. Kajan, "Local Binary CNN for Diabetic Retinopathy Classification on Fundus Images," *Acta Polytech. Hungarica*, vol. 19, no. 7, pp. 27–45, 2022, https://doi.org/10.12700/APH.19.7.2022.7.2

[7] S. G. Ali, X. Wang, L. Bi, Y. Jung, T. Chen, and H. Zhang, "Deep learning-based binocular system for automated diabetic retinopathy grading with prior clinical knowledge integration," *Vis. Comput.*, vol. 41, no. 8, pp. 5675–5688, 2025, https://doi.org/10.1007/s00371-024-03745-0

[8] N. M. Suhaimi *et al.*, "Comparative Analysis of Superpixel and Gabor Methods for Exudate Feature Extraction in Diabetic Retinopathy Fundus Images," in *Proceedings of the 7th International Conference on Electrical, Control and Computer Engineering*, 2024, vol. 2, pp. 123–136. https://doi.org/10.1007/978-981-97-3851-9

[9] I. Ahmad, V. Prakash, S. Manoj, and M. Gore, "Segmented Fractal and Central Symmetric LBP Based Texture Features for the Detection of Diabetic Retinopathy Using SVM," *SN Comput. Sci.*, vol. 5, pp. 1–14, 2024, https://doi.org/10.1007/s42979-024-02996-x

[10] J. Du, B. Zou, P. Ouyang, and R. Zhao, "Retinal microaneurysm detection based on transformation splicing and multi-context ensemble learning," *Biomed. Signal Process. Control*, vol. 74, no. November 2021, 2022, https://doi.org/10.1016/j.bspc.2022.103536

[11] S. Gupta, S. Thakur, and A. Gupta, *Comparative study of different machine learning models for automatic diabetic retinopathy detection using fundus image*, vol. 83, no. 12. Springer US, 2024. https://doi.org/10.1007/s11042-023-16813-9

[12] N. Gundluru *et al.*, "Enhancement of Detection of Diabetic Retinopathy Using Harris Hawks Optimization with Deep Learning Model," *Hindawi Comput. Intell. Neurosci.*, vol. 2022, 2022, https://doi.org/10.1155/2022/8512469

[13] R. Kumari, S. Kumar, and S. Godara, "An Efficient Diabetic Retinopathy Detection and Classification System Using LRKSA-CNN and," *Int. J. Adv. Comput. Sci. Appl.*, vol. 15, no. 12, pp. 613–627, 2025, https://doi.org/10.14569/IJACSA.2024.0151263

[14] U. Ishtiaq, E. Rahayu, M. Faizal, and Z. Ishtiaque, "A Hybrid Technique for Diabetic Retinopathy Detection Based on Ensemble-Optimized CNN and Texture Features," *Diagnostics MDPI*, vol. 13, pp. 1–21, 2023, https://doi.org/10.3390/diagnostics13101816

[15] H. N. Pham *et al.*, "Automated Grading in Diabetic Retinopathy Using Image Processing and Modified EfficientNet," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 12496 LNAI, pp. 505–515, 2020, https://doi.org/10.1007/978-3-030-63007-2_39

[16] A. Elmoufidi and H. Ammoun, "Diabetic Retinopathy Prevention Using EfficientNetB3 Architecture and Fundus Photography," *SN Comput. Sci.*, vol. 4, no. 1, pp. 1–9, 2023, https://doi.org/10.1007/s42979-022-01482-6

[17] U. Pamula *et al.*, "Optimizing diabetic retinopathy detection with electric fish algorithm and bilinear convolutional networks," *Sci. Rep.*, pp. 1–19, 2025, https://doi.org/10.1038/s41598-025-99228-w

[18] M. A. I. Mahmood, N. Aktar, and M. F. Kader, "A hybrid approach for diagnosing diabetic retinopathy from fundus image exploiting deep features," *Heliyon*, vol. 9, no. 9, pp. 1–14, 2023, https://doi.org/10.1016/j.heliyon.2023.e19625

[19] A. Gupta and V. Kumar, "A Deep Learning Fusion Approach for Automatic Diagnosis and Grading of Diabetic Retinopathy," *2023 14th Int. Conf. Comput. Commun. Netw. Technol.*, pp. 1–6, 2023, https://doi.org/10.1109/ICCCNT56998.2023.10306810

[20] A. Karsaz, "Diabetic retinopathy screening using improved support vector domain description : a clinical study," *Soft Comput.*, vol. 26, no. 19, pp. 10085–10101, 2022, https://doi.org/10.1007/s00500-022-07387-z

[21] R. K. Patel and M. Kashyap, "Automated screening of glaucoma stages from retinal fundus images using BPS and LBP based GLCM features," *Int. J. Imaging Syst. Technol.*, vol. 33, no. 1, pp. 246–261, 2022, https://doi.org/10.1002/ima.22797

[22] S. Gupta, "Microaneurysm detection by multiple feature subset extraction and selection based on SVM-weights and Genetic Algorithm-Neural Network," *IEEE ICACCS*, pp. 129–134, 2021, https://doi.org/10.1109/ICACCS51430.2021.9441746

[23] S. Gayathri, V. P. Gopi, and P. Palanisamy, "Automated classification of diabetic retinopathy through reliable feature selection," *Phys. Eng. Sci. Med.*, vol. 43, no. 3, pp. 927–945, 2020, https://doi.org/10.1007/s13246-020-00890-3.

[24] P. Macsik, J. Pavlovicova, S. Kajan, and J. Goga, "Image preprocessing-based ensemble deep learning classification of diabetic," *IET Image Process*, vol. 18, pp. 807–822, 2024, https://doi.org/10.1049/ipr2.12987