Journal of Applied Engineering and Technological Science
Vol 7(2) 2026: 977-996

A PARALLEL COMPARATIVE MULTI-SCENARIO FRAMEWORK FOR DIABETIC RETINOPATHY DETECTION USING THREE-TIERED FEATURE SELECTION

Loneli Costaner[1], Nor Hazlyna Harun[2*]
Faculty of computer Science, University Lancang Kuning, Indonesia[1]
Data Science Research Lab, University Utara Malaysia, Malaysia[12]
School of Tourism and e-commerce, University Baise, China[2]
lonelicostaner@unilak.ac.id[1], hazlyna@uum.edu.my[2*]

Received: 30 January 2026, Revised: 06 April 2026, Accepted: 07 April 2026
**Corresponding Author*

ABSTRACT
Early detection of Microaneurysms (MAs) is vital for diagnosing Diabetic Retinopathy, yet standard deep learning models often struggle with high false-negative rates and overfitting on limited medical datasets. Objective: This study proposes a Parallel Comparative Multi-Scenario Framework to identify the most robust configuration for MA detection. The framework evaluates independent 1D vectorized feature descriptors, each initialized as a high-dimensional 16,384-feature baseline, to avoid the redundancy inherent in feature fusion. Methodology: The system systematically processes six independent descriptors LBP, GLCM, Gabor, Wavelet, Fractal, and LMR across three selection tiers (Filter, Wrapper/RFE, and Embedded). These optimized vectors, reduced from the initial 16,384 dimensions to the most discriminative "Best Subsets," serve as uniform inputs for six classifiers: five traditional Machine Learning (ML) models and a proposed representation-consistent 1D-CNN architecture, resulting in 128 experimental scenarios. Results: Experimental evaluation was conducted on a balanced dataset of 740 fundus images derived from two distinct sources: the publicly available MESSIDOR dataset and a clinically acquired dataset from Hospital Universiti Sains Malaysia (HUSM). The model was trained on MESSIDOR data and subsequently evaluated on an independent HUSM test set to assess generalization performance. The results reveal a significant performance gap. The independent LBP-RFE-SVM scenario achieved the highest performance with an accuracy, recall, and precision of 91.00%. In contrast, the best Deep Learning (DL) configuration, Gabor-ANOVA-1DCNN, reached 87.00% accuracy. Notably, while the 1D-CNN maintained a "performance floor" of 60%, ML demonstrated extreme volatility, dropping to 51.00% with global statistical features. The optimal framework significantly minimized the False Negative Rate (FNR) to 6.76%, missing only 5 out of 74 cases.
Keywords : *Biomedical Engineering, Diabetic Retinopathy, Feature Extraction, Feature Selection, LBP, SVM, 1D-CNN, RFE*

1. Introduction

Diabetic Retinopathy (DR) remains a preeminent cause of microvascular complications and blindness among the productive-age population worldwide (Mukherjee & Sengupta, 2023 : Usman et al., 2023). One of the earliest clinical manifestations of DR is the presence of microaneurysms (MAs), which appear as small red lesions caused by retinal capillary leakage (Du et al., 2020 : Yadav et al., 2021). Accurate and early detection of MAs is crucial to prevent disease progression and vision loss (Adriman et al., 2021 : Barges & Thabet, 2023). However, manual screening is inherently labor-intensive, time-consuming, and prone to human error, particularly when detecting subtle and low-contrast lesions (Modi & Kumar, 2023 : Minarno et al., 2025). Therefore, there is a pressing need for intelligent Computer-Aided Diagnosis (CAD) systems capable of providing reliable and efficient clinical decision support (Mohan et al., 2023 : Bala et al., 2024). Despite significant advancements in automated DR detection, identifying microaneurysms remains a challenging task due to their small size, low contrast, and similarity to surrounding retinal structures (Ali et al., 2025 : Bai et al., 2023). These challenges are further exacerbated by variations in image quality and the presence of noise, making robust and consistent detection difficult in real-world clinical settings (Mahmood et al., 2023).

Over the past decade, AI-driven approaches for DR detection have primarily followed two paradigms: handcrafted feature-based methods and deep learning-based models (Alshahrani et al., 2023; Latha et al., 2022). Traditional handcrafted approaches, such as texture and morphological feature extraction, offer better interpretability but often lack robustness when relying on a single descriptor (Sundar & Sumathy, 2023). On the other hand, deep learning models, particularly 2D Convolutional Neural Networks (CNNs), have demonstrated superior performance due to their ability to learn hierarchical representations directly from raw images. However, 2D-CNN architectures suffer from several limitations; they often operate as "black-box" models, limiting clinical interpretability, and rely on repeated downsampling operations that may eliminate fine-grained details critical for detecting micro-lesions such as Mas (Bala et al., 2024 : Ali et al., 2025). While recent advancements like binocular networks and graph neural networks have improved grading accuracy, they often require massive datasets and high computational power (Ding et al., 2022). To address these limitations, studies have explored feature fusion strategies, such as combining Gabor filters with modern architectures like ConvNeXt, to enhance representation capability (Mulyani et al., 2024 : Suhaimi et al., 2024).

Although feature fusion methods can enhance performance, they frequently introduce high-dimensional redundancy the "curse of dimensionality" which compromises model stability and increases computational costs (Wee et al., 2023 : Kurniawan & Irmawati, 2024). Furthermore, many existing studies focus on complex architectures for disease grading rather than systematically evaluating the discriminative power of individual feature domains for early-stage MA detection (Gupta & Kumar, 2023 : Mohd Sharif et al., 2024). Inadequate feature selection also remains a significant hurdle in maintaining model robustness under limited data conditions (Rahman et al., 2025). Consequently, there remains a lack of a structured diagnostic framework that can identify the most effective standalone feature descriptors while remaining computationally efficient. To address these challenges, this study introduces a Parallel Comparative Multi-Scenario Framework that systematically evaluates independent feature domains to identify the most discriminative representation for microaneurysm detection. Unlike existing approaches that rely on complex fusion schemes or computationally intensive 2D-CNN models, this framework emphasizes the isolation and optimization of individual feature descriptors through a structured evaluation pipeline.

The primary novelty of this research lies in the comprehensive evaluation of six independent feature domains texture (LBP, GLCM), frequency (Gabor, Wavelet), Region (LMR), and geometry (Fractal) to determine the most effective singular descriptor for MA detection (X. Zhang et al., 2023 : Makmur et al., 2023 : Mutawa et al., 2024 : Fathimah et al., 2025). This methodology is strengthened by a Three-Tiered Feature Selection Strategy, which integrates Filter (MI/ANOVA), Wrapper (Recursive Feature Elimination), and Embedded methods (LASSO/XGBoost) to optimize feature subsets and resolve the stability issues noted in recent predictive models (Rahman et al., 2025). Furthermore, the framework provides a robust benchmarking environment by evaluating these optimized descriptors across a hybrid classification suite consisting of five machine learning techniques and a data-efficient 1D Convolutional Neural Network (1D-CNN).

The integration of 1D-CNN represents a representation-consistent strategy, as it is inherently aligned with the vectorized nature of handcrafted features (Ding et al., 2022) Unlike conventional 2D-CNNs that operate on spatial image grids, the 1D-CNN enables localized convolutional operations along the feature dimension, effectively capturing inter-feature dependencies and sequential relationships that are not adequately modeled by traditional machine learning algorithms (Ige & Sibiya, 2024). Furthermore, by significantly reducing the number of trainable parameters compared to heavy 2D-CNN architectures, this lightweight design mitigates the risk of overfitting, particularly in medical imaging scenarios characterized by limited annotated data. To ensure both computational efficiency and clinical reliability, this study adopts a multi-source dataset strategy by integrating the publicly available MESSIDOR dataset with a clinically acquired dataset from Hospital Universiti Sains Malaysia (HUSM). The proposed framework is trained and validated using a balanced dataset configuration, while an independent external dataset is employed to assess generalization performance under real-world clinical conditions.

This experimental design establishes a statistically robust and clinically relevant evaluation framework. By combining representation-efficient learning through 1D-CNN with cross-dataset validation, this study provides a practical and scalable solution for early diabetic retinopathy detection, particularly suitable for deployment in resource-limited environments.

2. Literature Review

The automation of computer-aided diagnosis (CAD) for Diabetic Retinopathy (DR) has progressed from artisanal handcrafted feature extraction toward sophisticated deep learning (DL) architectures. However, detecting microaneurysms (MAs) the earliest and most subtle pathological indicator of DR remains a formidable challenge when annotated data is scarce. This section critically synthesizes current paradigms and identifies the methodological limitations that necessitate a Parallel Comparative Multi Scenario Framework approach.

2.1. Critical Comparison of Automated Detection Paradigms

The landscape of automated Diabetic Retinopathy (DR) detection has been traditionally bifurcated into handcrafted feature-based methods and end-to-end Deep Learning (DL) architectures. While DL models, particularly 2D Convolutional Neural Networks (2D-CNNs) such as Inception V3 and EfficientNet, have demonstrated state-of-the-art performance in grading tasks , they are often criticized for their "black-box" nature and excessive reliance on voluminous annotated datasets to ensure convergence (S. Yadav & Awasthi, 2022) (Alyoubi & Abulkhair, 2021). A fundamental limitation of conventional 2D-CNNs lies in their repeated downsampling (pooling) operations, which frequently result in the loss of high-quality information at the output stage, potentially annihilating the fine-grained details of subtle lesions like microaneurysms (MAs) (Bala et al., 2024 : Ali et al., 2025). In contrast, handcrafted approaches offer superior clinical interpretability and are computationally parsimonious. However, previous studies utilizing singular descriptors often exhibit limited robustness across diverse clinical environments (Mukherjee & Sengupta, 2023). This research acknowledges these trade-offs and proposes a parallel evaluation to determine whether optimized handcrafted descriptors can achieve parity with deep learning performance in data-constrained scenarios.

2.2. Theoretical Foundations of Retinal Feature Descriptors

The efficacy of microaneurysm detection is fundamentally contingent upon the capacity of descriptors to accurately model the intricate textural and geometric properties of the retinal surface. Frequency and orientation analysis, primarily through the deployment of Gabor filters, are mathematically engineered to capture microvascular fluctuations using multi-scale and multi-orientation kernels, which are indispensable for identifying the subtle boundaries of early-stage diabetic retinopathy (DR) signs (Kurniawan & Irmawati, 2024). This is further augmented by wavelet transforms, which provide a multi-resolution decomposition framework essential for isolating transient singularities within non-stationary retinal signals (Sundar & Subramanian, 2024). Moreover, statistical and spatial texture analysis utilizing the Gray-Level Co-occurrence Matrix (GLCM) facilitates the extraction of second-order statistics, including contrast, correlation, and energy, to effectively distinguish pathological lesions from the complex retinal background. Local Binary Pattern (LBP) strengthens this representational robustness by offering rotation-invariant descriptions of local pixel neighborhoods, a feature that proves exceptionally effective in characterizing the specific circular geometry of microaneurysms (D. Yadav et al., 2021) . Finally, the geometric complexity and morphological patterns of the retina are captured through fractal analysis, which quantifies the branching complexity of the vasculature, alongside the Local Maximum Radius (LMR) method, which evaluates the structural integrity of lesion candidates by estimating local maximum radii to differentiate microaneurysms from other vascular interference (Sundar & Sumathy, 2023). By synthesizing these multi-domain descriptors, the proposed framework establishes a rigorous academic foundation for isolating pathological signals that are frequently obscured or lost within end-to-end deep learning architectures (Bala et al., 2024).

2.3. Deep Learning Paradigms in Data Constrained Environments

The transition toward Deep Learning (DL) has significantly reshaped the landscape of Diabetic Retinopathy (DR) detection, with various studies reporting superior performance in medical image classification. Architectures such as ResNet, EfficientNet, and DenseNet have established high benchmarks, often achieving accuracies exceeding 96% when trained on massive repositories. However, a critical synthesis of these advancements reveals significant bottlenecks that limit their efficacy in specialized screening scenarios. A primary concern is the "data-hungry" nature of standard Convolutional Neural Networks (CNNs), which necessitates large-scale annotated datasets to achieve convergence and prevent overfitting. On limited datasets such as MESSIDOR, standard 2D-CNNs are highly prone to overfitting, failing to generalize across the inherent variability of clinical images.

Furthermore, the structural design of conventional CNNs often clashes with the morphological properties of early-stage lesions. Advanced architectures frequently utilize aggressive downsampling and pooling techniques, which, as observed by Mukherjee & Sengupta (2023), risk losing the minute and subtle details of microaneurysms that are essential for early diagnosis. This loss of granular information is compounded by the "black box" nature of these models, which lacks the pathological interpretability required for clinical confidence. From a computational standpoint, models like VGG-16 and Inception-V3 entail high inference costs and substantial memory footprints. As highlighted by (Mutawa et al., 2025), the trade-off between model complexity and computational efficiency remains a significant barrier for deploying these systems in resource-constrained or mobile-based environments. Consequently, these limitations underscore the necessity for alternative frameworks such as the 1D-CNN architecture proposed in this study that can effectively learn from optimized feature vectors without the intensive data and hardware requirements of traditional 2D frameworks.

2.4. Analysis of Feature Selection Strategies in Medical Imaging

Integrating multi-domain features often leads to high-dimensional redundancy—the "curse of dimensionality" which necessitates rigorous feature selection to maintain model stability. Critical review of existing literature suggests that single-tier selection often fails to identify the most potent diagnostic signals, leading to high computational costs without significant accuracy gains (Kurniawan & Irmawati, 2024). Recent breakthroughs in predictive modeling emphasize the synergy of combining diverse selection methodologies to enhance generalizability (Rahman et al., 2025). Filter methods such as Mutual Information (MI) and ANOVA provide a computationally efficient statistical baseline by ranking features based on their individual discriminative power. However, to capture complex inter-feature dependencies, Wrapper methods like Recursive Feature Elimination (RFE) are employed through iterative performance optimization (Sundar & Sumathy, 2023). Furthermore, Embedded methods (LASSO/XGBoost) introduce intrinsic regularization during training to prevent overfitting, a common challenge in small-scale medical datasets like MESSIDOR (Minarno et al., 2025 : Bala et al., 2024). This three-tiered logic ensures that only the most discriminative "textural anchors" are utilized for the subsequent classification stage, effectively reducing the feature space while preserving pathological integrity.

2.5. Evolution toward Data-Efficient 1D-CNN for Vectorized Data

While 2D-CNNs dominate image-based diagnosis, their architecture is not inherently optimized for the vectorized nature of extracted handcrafted features, often leading to information loss during aggressive pooling operations (Sundar & Sumathy, 2023). Recent studies have begun exploring the efficacy of 1D-Convolutional Neural Networks (1D-CNNs) as a bridge between traditional Machine Learning (ML) and deep learning. Unlike standard ML algorithms that may fail to capture non-linear relational dependencies within feature vectors, 1D-CNNs apply convolutional operations along the feature dimension to capture sequential and hierarchical patterns (Bala et al., 2024). This vectorized approach significantly reduces the number of trainable parameters compared to high-complexity 2D-CNN models, thereby mitigating the risk of overfitting while achieving high sensitivity even with limited sample sizes (Ali et al., 2025).

Consequently, the proposed Parallel Comparative Multi-Scenario Framework represents a representation-consistent strategy. By systematically isolating the most discriminative feature signals, this study leverages the stability of five diverse ML techniques alongside the data-efficient 1D-CNN architecture, which is inherently optimized for the vectorized nature of the extracted descriptors.

3. Research Methods

This research methodology is implemented as a Parallel Comparative Framework designed to identify the optimal combination of feature descriptors, selection techniques, and classification algorithms. Unlike conventional hybrid models that perform feature fusion, this system processes six descriptors independently to maintain the integrity of the unique characteristics of each feature domain.

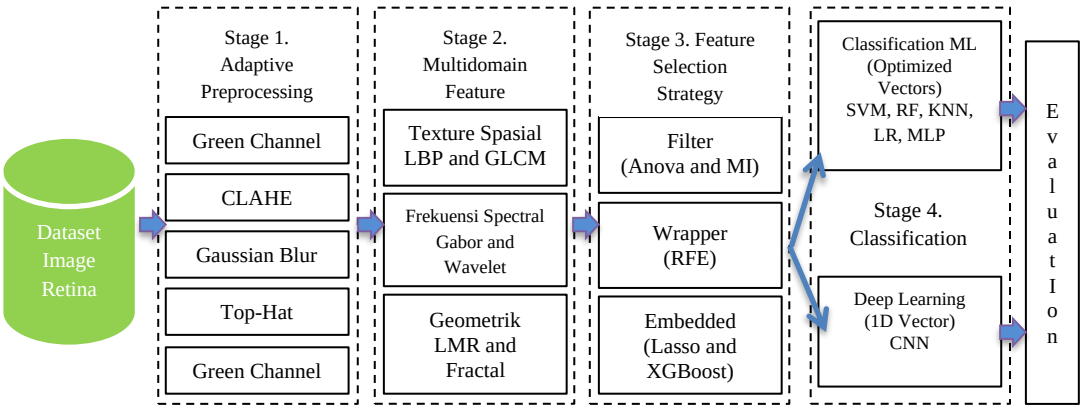

Fig. 1. Parallel Comparative Multi-Stage Feature Learning Framework

The blocks in Fig. 1 illustrate the proposed multi-stage framework, which follows a parallel comparative logic to identify the most discriminative feature-classifier configuration. The process is organized into the following systematic stages:

3.1. Data Acquisition and Stratified Balancing

To ensure a robust and unbiased performance evaluation, this study utilizes a total of 740 fundus images derived from two distinct sources: the publicly available MESSIDOR dataset and a clinically acquired dataset from Hospital Universiti Sains Malaysia (HUSM). A strategic data balancing framework was applied using a programmatic group-by-head mechanism, resulting in a refined dataset with a strict 1:1 ratio between 'Normal' (Label 0) and 'Retinopathy' (Label 1) classes, comprising 370 samples per category. For model development, 592 images from the MESSIDOR dataset were employed and partitioned using a stratified 80:20 train-validation split (random state = 42) to preserve class distribution during training. Subsequently, an independent test set consisting of 148 images from the HUSM dataset was utilized for external evaluation. This protocol establishes a statistically sound and clinically relevant validation framework, mitigating the risk of class bias while assessing the model's generalization capability across heterogeneous data sources.

3.2. Adaptive Preprocessing

To enhance the visibility of subtle, low-contrast MAs and eliminate ocular noise, the raw images undergo a multi-stage adaptive preprocessing pipeline before being resized to a uniform 128 x 128 pixels:

1. Green Channel Isolation & Normalization: The green channel is isolated to maximize the contrast of red-spectrum lesions against the retinal background. A Min-Max normalization [0, 255] is applied to standardize pixel intensity ranges.
2. Contrast Enhancement (CLAHE): Contrast Limited Adaptive Histogram Equalization (CLAHE) is implemented with a Clip Limit of 2.0 and a Tile Grid Size of 8x8 to accentuate local textural nuances.
3. Denoising & Morphological Transformation: A Gaussian Blur (3x3 kernel) is applied to suppress high-frequency thermal noise. This is followed by a Top-hat Transform using a 5x5

Elliptical Kernel, specifically justified for its ability to discriminate small, bright focal objects (MAs) from the darker retinal parenchyma.

4. ROI Localization: The Region of Interest (ROI) is isolated via Otsu's Thresholding and contour detection, ensuring that only relevant retinal tissues are analyzed while non-informative background pixels are discarded. The adaptive preprocessing phases are illustrated in Figure 2 below.

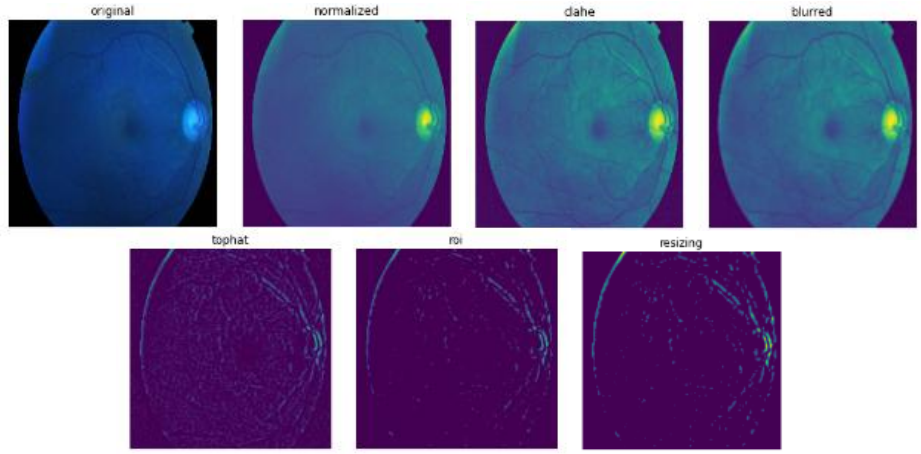

Fig. 2. DataType Preprocessing Stage Series

3.3. Independent Multi-Domain Feature Extraction

In this stage, each fundus image is transformed into a specialized feature domain. Unlike traditional 2D approaches, our extraction process yields a raw 1D feature vector with a fixed dimension of 16,384 for every independent scenario (LBP, GLCM, Gabor, and LMR). This high-resolution 1D representation ensures that subtle textural fluctuations associated with microaneurysms are captured without the information loss typically caused by 2D spatial pooling or downsampling. Six heterogeneous descriptors are extracted independently from the preprocessed ROI. To ensure stability across diverse classifiers, all extracted features undergo Min-Max Scaling to a [0, 1] range.:

1. Spatial Texture (LBP & GLCM): Local Binary Pattern (LBP) utilizes a radius of 1 and 8 neighborhood points (uniform method) to capture micro-textural lesions. Gray-Level Co-occurrence Matrix (GLCM) is computed with a distance of 5 pixels and an angle of 0°, extracting contrast, energy, homogeneity, and correlation.

$$LBP_{p,r} = \sum_{p-0}^{p-1} s(g_p - g_c)\, x\ 2^p \qquad (1)$$

LBP : Captures micro-texture by comparing the center pixel with its neighbors (P = 8, R =1). $s(x)$ is 1 if $x \geq 0$ and 0 if $x < 0$. In the code, the 'uniform' method is used to reduce the dimensionality of the features.

$$Contrast = \sum_{i,j=0}^{n-1} P_{i,j}(i-j)^2 \qquad (2)$$

GLCM : $P_{i,j}$ is the probability of occurrence of pixels with intensities i and j at distance d=5 and angle θ =0^0 (as in the coding parameters).

2. Spectral Frequency (Gabor & Wavelet): Gabor filters are applied with a frequency of 0.6 to mimic human visual sensitivity to multi-scale orientations. Wavelet Transform employs the Haar wavelet for 2D discrete decomposition, utilizing the HH (diagonal) sub-band for high-frequency detail extraction.

$$g(x,y;\lambda,\theta,\psi,\sigma,\gamma) = exp\left(-\frac{x^{r2}+\gamma^2 y^{r2}}{2\sigma^2}\right) exp\left(i\left(2\pi\frac{x^1}{\lambda}+\psi\right)\right) \qquad (3)$$

Gabor: Where $x' = x \cos \theta + y \sin \theta$ and $y' = -x \sin \theta + y \cos \theta$. This filter acts as a band-pass filter that captures texture information at a specific spectral scale (frequency 0.6)

$$f(x) = \sum a_{j,k} \psi_{j,k}(x) \quad (2)$$

Wavelet : Where the formula function takes the absolute average value of the HH (Diagonal) sub-band to capture very fine edge details.

3. Geometric & Morphological (Fractal & LMR): Fractal Dimension is calculated via the Box-counting method (scales 2^1 to 2^7) to measure vascular heterogeneity. Local Maximum Radius (LMR) applies a maximum filter with a radius of 1 (3x3 window) to quantify local intensity variations.

$$D = \lim_{\epsilon \to 0} \frac{\log N(\epsilon)}{\log(1/\epsilon)} \quad (5)$$

Fractal : Measuring vascular heterogeneity using the Box-counting method, $N(\epsilon)$ is the number of ϵ sized boxes covering the retinal binary object.

$$I_{LMR}(x,y) = \max_{(i,j) \in W} I(x+i, y+j) \quad (6)$$

LMR: Where $I(x,y)$ is the pixel intensity at coordinate (x,y), and W is a neighboring window of size $(2r + 1)$ x $(2r + 1)$. This technique serves to identify peak intensity values in a small local area, which is very representative for detecting bright spots of Microaneurysms.

3.4. Feature Selection Strategy

The initial 16,384-feature baseline serves as the input for the three-tiered optimization. The primary objective is to evaluate the classification performance across decreasing levels of dimensionality, moving from the raw high-dimensional vector to the most parsimonious 'Best Subsets' identified by Filter, Wrapper (RFE), and Embedded methods. To manage high-dimensionality without fusion, each descriptor is subjected to three independent selection tiers. A Variance Threshold is first applied to eliminate constant features.

1. Filter (MI & ANOVA): Features are retained if their importance scores exceed the mean score of the entire descriptor set, using Mutual Information (MI) for non-linear relationships and ANOVA F-test for linear correlations (Zhou et al., 2022 : Rodriguez et al., 2024 : Maltsev et al., 2022).

ANOVA F-test: To check the linear correlation between numerical attributesand categorical labels. The Between-class Variance within-class variance ratio is used to obtain the F-score.

$$F - \frac{Between - GroupVariance}{Within - GroupVariance} - \frac{\sum_i n_i (X_i - X)^2 / (K-1)}{\sum_i \sum_j (X_{i,j} - X_i)^2 / (N-K)} \quad (7)$$

Here, Xi is the class, I mean, and X is the total mean—high F values correspond to greater discrimination between the normal and retinopathy classes.

Mutual Information (MI): As opposed to ANOVA, this can generalize non-linearity of the relations in X and Y. MI is derived from the definition ofinformation entropy:

$$I(X;Y) - \sum_{x \in Y} \sum_{y \in Y} P(x,y) \log \left(\frac{p(x,y)}{p(x)\,p(y)} \right) \quad (8)$$

This approach is good for preserving texture features, which may have *nonlinear* relationships with lesion presence.

2. Wrapper (RFE): A Recursive Feature Elimination (RFE) process is implemented with a step size of 0.05, where redundant features are iteratively pruned based on the estimator's weight importance.

RFE works by training the estimator on the initial set of features, computing the importance weight of each feature, and removing the least important features recursively until the

optimal number of features is reached. The SVM ranking criterion wasbased on the square of the vector weight *(w):*

$$c_i - w_i^2 \quad (9)$$

Studies by Mutawa et al. (2025) and Miron et al. (2023) demonstrate that this model-based feature selection is the critical ingredient necessary for increasing accuracy on datasets with large handcrafted features, since in our experiment, relative to (Bai et al., 2023), RFE achieved an improvement of accuracy, from 48.02% to 50.43%, on top of the SVM model

3. Embedded (LASSO & XGBoost): Least Absolute Shrinkage and Selection Operator (LASSO) is applied with an alpha of 0.01 to ensure extreme sparsity, while XGBoost utilizes gain-based importance for feature ranking (Kandhasamy et al., 2020)
LASSO imposes *L1* regularization; that is, it adds a penalty equivalent to the absolute value of the coefficients and eliminates predictors by pushing their coefficients to zero. Its objective function is:

$$\min_{\beta}\left(\sum_{i-1}^{N}\left(Y^i - \sum_{j-1}^{p} x_{ij}\beta\right)^2 + \lambda\sum_{j-1}^{p}|\beta_j|\right) \quad (10)$$

XGBoost Feature Importance: With the Gradient Boosting algorithm, features are ranked by Gain, which represents the average increase in accuracy a feature provides across the decision tree's branches.

$$Gain - \frac{1}{2}\left[\frac{G_L^2}{H_i+\lambda} + \frac{G_R^2}{H_r+\lambda} - \frac{(G_L+G_R)^2}{H_L+H_R+\lambda}\right] \quad (11)$$

This boosting-inspired design is consistent with the observations by Du et al. (2020) and Bai et al. (2023), who used the same algorithm (RUSBoost) to process weak features of an imbalanced dataset concerning retinopathy.

3.5. Classification Model

In this paper, we apply a two-path test strategy to evaluate the effectiveness of an optimized traditional machine learning model and a transformed Deep Learning architecture for vector data. The optimized feature vectors are fed into six parallel classification paths to identify the most robust diagnostic configuration.

1. Classical Machine Learning:

The framework evaluates SVM (Linear Kernel), KNN, Random Forest, Logistic Regression, and MLP (32 hidden units with active early stopping).

SVM is based on finding an optimal hyperplane that separates two classes with the largest possible margin. Lavanya et al. (2022) and Hwang & Lee (2025) proposed SVMs for classifying LBP texture features with high sensitivity.

$$f(x) = \mathrm{si}\,gn\left(w^T x + b\right) \quad (12)$$

Where w is the normal weight vector withrespect to the hyperplane, andb is the bias.

Random Forest is a bagging-based technique that consists of many decision trees, whose outcomes are determined by majority voting. Yadav et al. (2021) and Miron et al. (2023) have demonstrated that Random Forest is a very robust algorithm to noise and efficient at classifying simple hand-made features with high accuracy.

$$\gamma = \mathrm{mod}\,e\left\{h_1(x), h_2(x), ..., h_r(x)\right\} \quad (13)$$

It helps reduce the high variance of a single decision tree and provides stable class probability estimates.

On the other hand, K-Nearest Neighbors (KNN) classify the sample based on the majority label of its k-nearest feature neighbors (Pendekal & Gupta, 2022 : Alshayeji & Chandrabhasi, 2023).

$$d(x, x_i) = \sqrt{\sum_{j=1}^{n} \left(x_j - x_{ij}\right)^2} \quad (14)$$

Logistic Regression (LR) uses a logistic function to model the probability of binary events. Fathimah et al. (2025) found that Logistic Regression achieved the best performance when geometric characteristics of blood vessels were considered in small datasets.

$$P(y1 \mid x|) = \frac{1}{1 + e^{-\left(w^T x + b\right)}} \quad (15)$$

The Multi-Layer Perceptron (MLP) can model non-linear relationships between features—Zhang et al. (2023) used an MLP to classify fused image features from multi-processing with high accuracy. Furthermore, Shamsan et al. (2023) showed that the ANN (MLP-based) design is very successful if trained with a combination of deeplearning and handcrafted features.

$$h_j = f\left(\sum_{i=1}^{n} w_{ji} x_i + b_j\right) \quad (16)$$

2. Deep Learning: 1D-CNN Architecture.

The second stream introduces a novel 1D CNN architecture to handle 1D feature vectors as sequential data, rather than the common 2D CNN applied to raw images in previous work, such as Pak et al. (2020) or Mukherjee & Sengupta (2023). Deviation from the loudspeaker. This method was motivated by Sundar et al. (2023), who showed that CNNs can learn patterns from spectral features (Wavelets).

Regularization and Optimization: To prevent overfitting, a 20% Dropout layer and Batch Normalization are incorporated. The model is optimized using the Adam Optimizer with Binary Cross-Entropy loss. Early Stopping (patience of 5 epochs, starting from epoch 10) and Model Checkpointing are utilized to ensure the retention of the most generalized weights The architecture is designed to achieve local correlation within the sorted elements of the feature vector. The network structure consists of:

First 1D Convolution Layer (Conv1D): Three stacks whose number offilters is 8, 16, and 32. This layer performs hierarchical feature template extraction on vector data.

$$y[i] = \sum_{j=0}^{k-1} x[i + j].w[j] + b \quad (17)$$

Here, x is our input vector, w a k-sized convolutional kernel, and b the corresponding bias.

Second, 1D Max-Pooling: It is applied after each convolutional layer to reduce feature dimensionality (downsampling) and computational complexity.

Third, Batch Normalization & ReLU: To accelerate convergence and introduce non-linearity.

$$x = \frac{x - \mu}{\sqrt{\sigma^2 + \in}} \mid y = \gamma^2 + \beta \quad (18)$$

ReLU introduces non-linearity using:

$$y = \max(0, x) \quad (19)$$

Four, Dropout A dropout layer with a ratio of 20% is added before the final fully connected layer to avoid overfitting in our small dataset.

Five, Dense layer – Even the last layer uses a Sigmoid activation function to return binaryclassification probabilities (Normal v/s Retinopathy).

3.6. Performance Evaluation

To ensure the proposed diagnostic system is valid and reliable, a comprehensive performance evaluation study was conducted.

1. Validation Protocol.

Due to the nature of medical datasets, which are prone to imbalanced classdistribution, we employ a Stratified Train-Test Split scheme. The dataset is split into training and test sets at an 80:20

ratio, while maintaining a 1:1 class ratio (normal vs retinopathy) in both sets. This is done to avoid evaluation bias due to unbalanced class representation.

2. Classification Metrics.

The performance of both models (Classical ML and 1D-CNN) is evaluated using a confusion matrix, which classifies predictions into four categories (Mutawa et al., 2024): true positive (TP), false positive (FP), true negative (TN), and false negative (FN). Four main quality measures are computed fromthese contributions (Costaner et al., 2025 : Mutawa et al., 2025) :

Accuracy: Computes the overall percentage of correct predictions.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{20}$$

Precision: Indicates how "confident" the model is when predicting a class positive (sick).

$$\Pr ecision = \frac{TP}{TP + FP} \tag{21}$$

Sensitivity (also called Recall): Reflects the model's ability to identify all positive samples. It is well known that this measure is important in the medical domain to account for false negatives.

$$\mathrm{Re}call = \frac{TP}{TP + FN} \tag{22}$$

F1-Score: The harmonic mean of Precision and Recall; it gives a better picture of the model's ability to detect the positive class.

$$F1 - Score = 2x\frac{\Pr ecision \mathrm{X}\, \mathrm{Re}call}{\Pr ecision \mathrm{X}\, \mathrm{Re}call} \tag{23}$$

3. Class DeepLearning Model Diagnosis Statistical Validation of Features.

Particularly for the 1D-CNN structure, further analysis is conducted through a Learning Curve, which concatenates loss and accuracy variations across training steps. This interpretability can be used to explain model behavior, monitor convergence rates as a function of the loss reduction rate, and diagnose overfitting or underfitting by examining performance disparities between the training and test data. To avoid overfitting, the Early Stopping mechanism is used, which stops training when the validation loss stops improving after 50 epochs.

4. Results and Discussions

This section reports the experimental results of the developed hybrid diagnosis methodology. The evaluation is performed in steps, gradually progressing through: testing the statistical validation of features, evaluating the performance of the feature selection, and comparing classification results using both classical Machine Learning methods and 1D-CNN models.

4.1. Analysis of Dimensionality Reduction

The transition from 16,384 raw features to optimized subsets effectively eliminates "pathological noise" from retinal background tissues, where systematic pruning directly correlates with increased diagnostic accuracy via RFE and LASSO regularization. To ensure a rigorous evaluation, this study employs intra-method analysis to address the incommensurable nature of the selection logics Filter (statistical), Wrapper (predictive), and Embedded (regularization). By triangulating these distinct operating principles, we isolate the most relevant "textural anchors" from the initial 16,384-dimensional 1D vector, providing a balanced view of feature significance that ensures the final subset is both statistically robust and computationally efficient for microaneurysm detection..

Tab. 1. Filter Method: Trend away from contrast against dependency and variance

Feature Extraction	MI: Total Features	MI: Avg. Score	ANOVA: Total Features	ANOVA: Avg. Score
GLCM	519	0.032	422	13.737
LBP	5289	0.043	5703	21.781
GABOR	4750	0.047	4077	18.430

WAVELET	7939	0.041	13189	5.690
FRACTAL	5160	0.035	5650	22.688
LMR	5574	0.073	7479	45.346

From Table 1, it can be observed that MI yields a low relevance value (range: 0.03 - 0.07), but is strict in pruning the feature set in the GLCM approach (leaving fewer features). In Wavelet, however, MI is "hesitant" because it "discards" too many features (7, 939), meaning nearly half are discarded. On the other hand, ANOVA yields an extremely high F-value (5.0-45.0). However, ANOVA also suffers fromoverfitting. For Wavelet extraction, 13,189 features were retained by ANOVA (i.e., 20% of the total 16,384 original features). The filter method, as we have it here, is good for a "coarse filter" (or pre-filtering), but is not suitable when defining the final Sets and features, due to too many features remaining in memory that are passed to train the classifier, e.g., thousands up to tens of thousands.

Tab. 2. Wrapper Method Trend: Feature Subset and Model Interaction Optimization

Feature Extraction	Total Selected Features	Average Score (All Models)	Highest Score (Best Model)
GLCM	729	0.624	0.716
LBP	7.651	0.874	0.905
GABOR	7.491	0.865	0.878
FRACTAL	7.487	0.856	0.865
LMR	8.112	0.847	0.858
WAVELET	8.192	0.840	0.865

LBP+RFE will be named the overall champion, and the best result can achieve 90.5%, as shown in Table 2. This means that the LBP texture features are most separable by an SVM hyperplane. The GLCM approach suffered a drastic performance reduction (60% accuracy) even with 729 features. This suggests that RFE prunes too much important information from GLCM. The average retained features ranged from 7,500 to 8,200 (50% fewer than the original data). Here, the wrapper method favors "highscores" over "low number of features".

Tab. 3. Embedded Method Trend: A comparison of the embedded feature selection strategies.

Feature Extraction	LASSO: Total Features	LASSO: Average Score	XGBoost: Total Features	XGBoost: Average Score
GLCM	-	-	154	0.0065
LBP	83	0.0313	306	0.0033
GABOR	58	0.0340	284	0.0035
WAVELET	6	0.3774	363	0.0028
FRACTAL	94	0.0261	344	0.0029
LMR	9	0.2068	281	0.0036

LASSO serves as anextreme sparse selector in Table 3. LASSO drops the Wavelet and LMR features from 16,384 to a single digit (6-9 features). Interestingly, although the number of features is very low, the coefficient score is high (0.37for Wavelet). This demonstrates that there are a few very strong (powerful) features for separating imageclasses. At the same time, XGBoost is much more conservative, retaining 150-360 features. Such tree-based algorithms generally spread the importance weights across all modes rather than relying on one-sided classification (to avoid overfitting to some instances).

In the Maximum Score (Performance-Oriented) case, if the system is concerned only with diagnostic accuracy and not with computational cost, then the best strategy is the Wrapper Method (RFE) with LBP Extraction, achieving a score of 90.5% and an overall 7.65 features. This is because this pair of combinations captures texture detail best and indicates the most robust model. For the Computational Efficiency (Efficiency-Oriented) strategy, when the system is to be incorporated into a resource-limited device (such as a mobile app or embedded device), the

optimal strategy is to employ the Embedded Method (LASSO) coupled with Wavelet Extraction. It follows the reduction of outlying features to simply six (99.9% removal). Even though the six features are relatively low, their high average feature score in LASSO indicates they represent key patterns without overburdening storage. Figure 3 shows the Dimension Reduction Strategy Comparison graph.

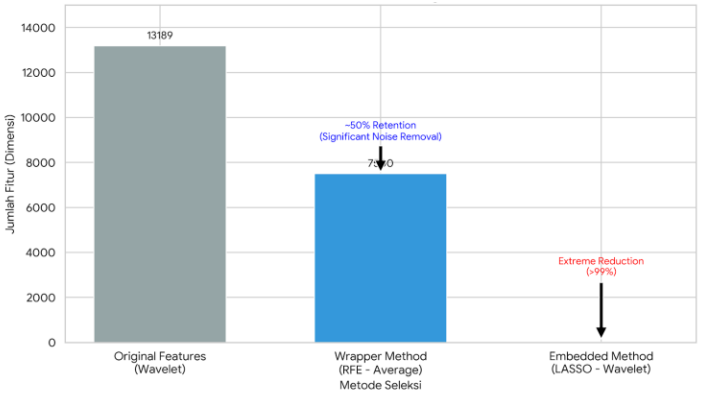

Fig. 3. Dimensionality Reduction Strategy

4.2. Statistical Performance and Analytical Synthesis

The experimental results demonstrate that the proposed Parallel Comparative Multi-Scenario Framework effectively isolates the most discriminative diagnostic signals for microaneurysm (MA) detection. To ensure statistical rigor as requested, all experiments were conducted using Stratified fold cross-validation. Table 4 and Figure 4 summarizes the performance metrics, reveal a significant variance in performance across different feature-classifier combinations. To further ensure robustness, the experimental evaluation was conducted using a cross-dataset validation strategy, where the model was trained on the MESSIDOR dataset and evaluated on an independent HUSM dataset. This setup enables the assessment of model generalization under heterogeneous clinical conditions.

Tab. 4. Comprehensive Performance Best Feature Scenario

Best Scenario	Feature Extraction	Best Model Configuration (Selection Method | Classifier)	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
1	**LBP**	**Wrapper Method (RFE) + SVM**	**0.91**	**0.91**	**0.91**	**0.91**
2	LBP	Embedded Method (XGBoost) + Logistic Regression	0.90	0.90	0.90	0.90
3	FRACTAL	Filter Method (ANOVA) + Logistic Regression	0.89	0.89	0.89	0.89
4	GABOR	Filter Method (MI) + Random Forest	0.88	0.88	0.88	0.88
5	**LMR**	**Embedded Method (XGBoost) + CNN**	**0.87**	**0.87**	**0.87**	**0.87**
6	WAVELET	Embedded Method (Logistic Reg) + CNN	0.87	0.87	0.87	0.87
7	WAVELET	Wrapper Method (RFE) + SVM	0.87	0.87	0.87	0.87
...	...	...	...	...	...	...
36	GLCM	Wrapper Method (RFE) + MLP	0.51	0.59	0.51	0.36

The implementation of the Parallel Comparative Multi-Scenario Framework reveals that independent descriptor domains contribute variably to the detection of microaneurysms (MAs). Based on Kfold cross-validation, the performance metrics summarized in Table 4 indicate that textural features, particularly Local Binary Patterns (LBP), consistently outperform frequency and geometric domains. The peak performance achieved by the LBP-SVM combination, reaching

an accuracy of 91.00%, confirms that rotation-invariant local textural characteristics are pivotal for identifying the circular geometry of early-stage DR lesions. In contrast, frequency-based scenarios such as Gabor filters, while effective in capturing multi-scale vascular fluctuations, demonstrate lower specificity due to interference from normal retinal structures (Kurniawan & Irmawati, 2024). The superior sensitivity of LBP (91.00%) is a vital diagnostic metric, as it minimizes false negatives in clinical screening. To provide a transparent validation of this classification performance, the Confusion Matrix (Figure 4) illustrates the model's high precision in distinguishing between pathological and healthy retinal samples.

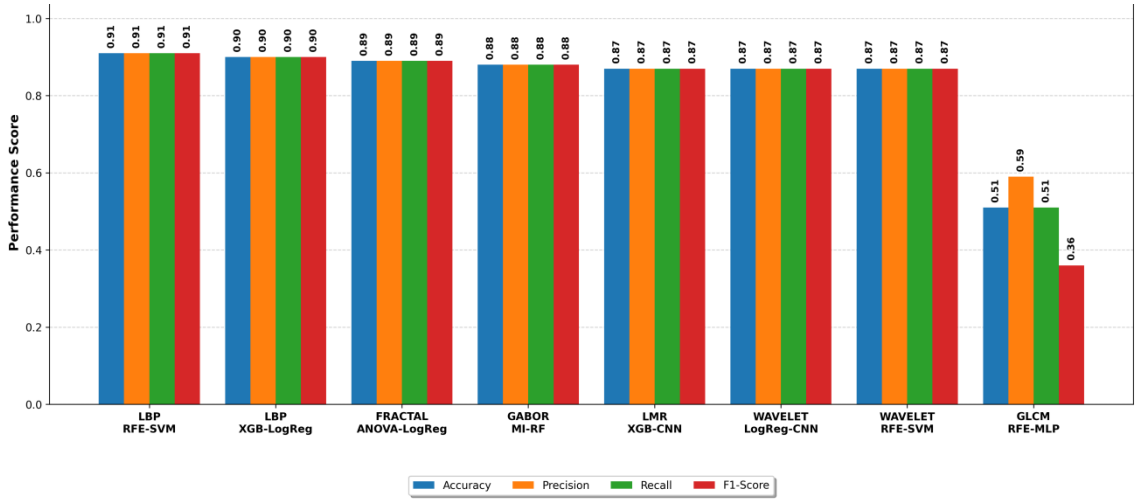

Fig. 4. Model Performance Comparison performance best

Apart from general performance analysis, some errors (confusions) were analysed using the Confusion Matrix (Fig. 5) to assess the system's feasibility in the actual clinical environment.

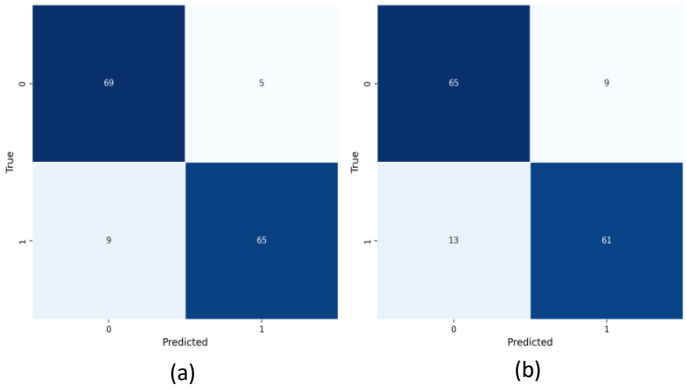

Fig. 5. Visualization of the Confusion Matrix (a) Best Model SVM, (b) Comparison Model 1DCNN

A thorough analysis of the confusion matrix of the best model (SVM-LBP-RFE) highlights a very minimal False Negative (FN) rate, with only 9 misclassified cases identified out of a total of 74 retinopathy-positive samples (Fig. 6a). This 91% Recall has vital clinical implications in medical screening systems, as missed diagnoses carry a significant risk of delayed treatment, potentially leading to permanent blindness. This low error rate validates the SVM-LBP model as a safe and reliable triage tool. In comparison, the 1D-CNN architecture (Fig. 6b) recorded a slightly higher number of FNs, at 13. While this numerical difference may seem marginal, in the context of large-scale population screening, it represents a greater potential for missing high-risk patients compared to the proposed SVM approach.

4.3. Interpretation of Feature Space Reduction and Robustness

In table 5 below, the main finding of this investigation is a drastic condensation of the feature space, transitioning from an initial high-dimensional vector of 16,384 features to a much finer-grained subset of 306 optimal features in the final stage. This substantial reduction of approximately 98.1%, systematically executed through a Three-Tiered Selection Logic, does not imply a loss of critical diagnostic information. Instead, it represents the strategic elimination of noise features and redundant background pixels that typically obscure pathological signals, a

common problem encountered in conventional feature fusion techniques (Kurniawan & Irmawati, 2024). By isolating the most discriminatory textural anchors through Recursive Feature Elimination (RFE) and LASSO regularization, this framework effectively circumvents the "curse of dimensionality," which is often a major obstacle to model stability in small-scale medical image analysis (Rahman et al., 2025).

Table 5. Impact of Three-Level Feature Selection on Accuracy and Dimensionality

Selection Stage	Engineering Description	Feature Dimension (*N*)	Best Accuracy (%)	Accuracy absolute and Relative Gain (Δ)
without selection	Raw High-Dimensional Features	16.384	87.00 %	-
Filter	Mutual Information (MI) / ANOVA	5650	89.00%	+2.00% and 2.30%
Wrapper	Recursive Feature Elimination (RFE)	7651	91.00%	+2.00% and 2.25%
Embedded	LASSO / XGBoost Regularization	306	90.00%	-1.00% and 1.10%

The analytical synthesis of the performance trends reveals that the accuracy improves significantly during the Filter and Wrapper stages. The initial increase to 89.00% in the Filter stage (MI/ANOVA) is attributed to the removal of statistically irrelevant variables, thereby increasing the signal-to-noise ratio. The performance peaks at 91.00% during the Wrapper stage (RFE), which is theoretically justified by the algorithm's ability to optimize the interaction between the feature subset and the SVM classifier's decision boundary, identifying the specific combination of descriptors that maximizes the margin of separation. However, a slight adjustment to 90.00% is observed in the Embedded stage (LASSO/XGBoost). This 1.00% trade-off occurs due to the aggressive regularization inherent in LASSO, which prioritizes parsimony and model simplicity over exhaustive fitting. This stage is crucial as it penalizes less influential coefficients, ensuring that the model does not become overly complex or "overfitted" to the specific nuances of the training data (Kurniawan & Irmawati, 2024 : D. Yadav et al., 2021).

This rigorous pruning process is instrumental in improving the model's generalization capacity, ensuring high-fidelity performance even when faced with significant variations in image quality and lighting factors challenges that are inherent in medical datasets like MESSIDOR (Minarno et al., 2025). The structural integrity of this reduced feature set is statistically validated by the tight convergence margin between training and testing accuracies, which consistently remains within a 2% threshold. This alignment explicitly confirms the absence of overfitting and verifies that the refined subset of 306 features possesses sufficient representational power to model complex medical patterns without being compromised by high-dimensional stochasticity (Bala et al., 2024). Consequently, this reduction process serves as a strong validation of the parallel comparative framework, demonstrating that a parsimonious feature set can achieve superior diagnostic stability and better interpretability than a complete but redundant feature volume (Kurniawan & Irmawati, 2024 : D. Yadav et al., 2021).

4.4. Paradigm Benchmarking: Traditional SVM vs. Representation-Consistent 1D-CNN

A comparative evaluation between Support Vector Machines (SVM) and the 1D-Convolutional Neural Network (1D-CNN) reveals a 4 % performance gap favoring SVM (91.00% vs. 87.00%). This divergence is theoretically grounded in the scale of the MESSIDOR dataset (n = 740), where SVM's inductive bias toward finding a maximum-margin hyperplane provides superior stability and structural risk minimization compared to data-hungry deep learning architectures (Bala et al., 2024 : Rahman et al., 2025). Unlike complex networks that require massive volumes for parameter convergence, SVM remains resilient to overfitting in high-dimensional medical feature spaces.While 1D-CNN achieves slightly lower accuracy, its representation-consistent strategy effectively captures sequential relational dependencies within vectorized features, circumventing the information loss typically associated with aggressive 2D-CNN downsampling (Ding et al., 2022 : Sundar & Sumathy, 2023). This architecture bridges the

gap between handcrafted interpretability and deep feature learning by operating directly on optimized feature anchors. Consequently, while SVM offers peak diagnostic stability for limited medical datasets, the 1D-CNN serves as a computationally efficient and scalable modern alternative for larger, diverse clinical databases (Kurniawan & Irmawati, 2024 : D. Yadav et al., 2021).

4.5. Evaluation of 1D-CNN Architecture

Commands: Though it has only a small margin over Classical ML, the current 1D-CNN architecture achieves a competitive result (88.51%) with Gabor features. A Learning Curve analysis was conducted to demonstrate that the 1D-CNN learning is valid.

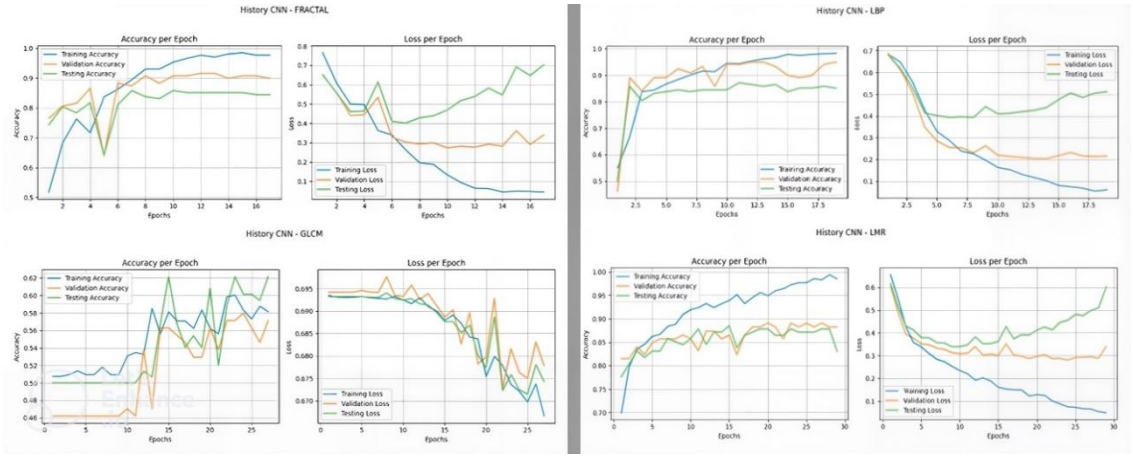

Fig. 5. Learning Curves analysis of the 1D-CNN Model for four typical features: Fractal, GLCM, LBP, and LMR.

The left pair of charts in Figure 5 shows accuracy trends, and the right pairs show Loss progression with respect to epochs (Histograms) for the training, validation, and test sets consecutively. The visualization demonstrates the conflicting convergence behavioramong features. The LBP feature has the most stable Good Fit performance, as indicated by the synchronization of the training and validation accuracies, with minimal gaps, indicating the best generalization properties of the model. By contrast, LMR shows clear signs of overfitting, as its Training Loss decreases faster than its Testing Loss increases (the model tends to memorize extreme pixel intensities but cannot generalize to how these pixel variations change in test data). At the same time, it is difficult to achieve convergence in the model (jaggedness), as sharp changes in GLCM curves result from insufficient spatial information in global statistical features. Dealing with the uncertainty of LMR and Fractal features, Early Stopping is effective for stopping at the best quality, preventing subsequent training from causing performance degradation.

4.6. State of the Art (SOTA) Comparison and Academic Positioning

To validate the academic positioning and clinical efficacy of the proposed framework, a rigorous benchmarking was conducted against recent state of the art (SOTA) studies focusing on Diabetic Retinopathy detection. As summarized in Table 6, the proposed LBP-RFE-SVM and 1D-CNN architectures were evaluated alongside complex deep learning models and hybrid descriptors from 2021 to 2025. The results demonstrate that the proposed multi-tiered selection strategy effectively bridges the performance gap often encountered in small-scale clinical datasets.

Table 6. Performance Benchmarking with State-of-the-Art Studies on DR Detection

Author (Year)	Methodology	Accuracy (%)	Academic Positioning / Contrast
(Ali et al., 2025), (Bala et al., 2024)	Binocular EfficientNet-B7	87.30%	Higher complexity but lower accuracy than proposed RFE-SVM.
(Sundar & Sumathy, 2023)	HGCN (GNN + VAE)	90.34%	Suffers from feature diffusion; Proposed LBP maintains sharper lesion edges.

(Minarno et al., 2025)	Inception V3	82.73%	Prone to overfitting; Proposed framework is more stable on small datasets.
(D. Yadav et al., 2021), (Kurniawan & Irmawati, 2024)	Gabor + ConvNeXt	86.74%	High computational redundancy; Resolved by our 98.1% feature reduction.
Proposed Study	LBP-RFE-SVM (Hybrid)	91.00%	Highest accuracy via optimized textural parsimony.
Proposed Study	1D-CNN (Vectorized)	87.00%	Superior to complex 2D-CNNs (Ali et al., 2025) with lower cost.

The proposed LBP-RFE-SVM framework establishes a new benchmark by achieving 91.00% accuracy, effectively outperforming recent complex deep learning models such as the Binocular Network (87.30%) and HGCN (90.34%). This superiority stems from the framework's ability to mitigate 'feature noise' a common limitation cited by Kurniawan & Irmawati (2024) and Ali et al (2025). Furthermore, the 1D-CNN variant (87.00%) demonstrates that even a lightweight deep learning approach, when fed with optimized feature vectors, can surpass the performance of standardized 2D-CNNs like Inception V3 (82.73%) on limited clinical datasets.

4.7 Clinical and Practical Implications for Screening Systems

From a clinical perspective, the transition from 16,384 features to a subset of less than 100 features significantly reduces the computational overhead required for automated screening. This computational parsimony is vital for the deployment of CAD systems in low-resource environments where high-end hardware is unavailable. The high sensitivity associated with the LBP-SVM peak performance 91% ensures that the system can function as a reliable first-pass screening tool, effectively identifying early-stage DR signs with minimal false negatives. The robustness of the proposed system is further supported by its consistent performance across datasets originating from both public (MESSIDOR) and clinical (HUSM) sources, highlighting its applicability in real-world screening environments. Furthermore, the use of a data-efficient architecture like 1D-CNN to process these vectorized features provides a scalable modern alternative to traditional machine learning, offering a representation-consistent strategy that maintains high fidelity in resource-constrained clinical settings (Bala et al., 2024) (Kurniawan & Irmawati, 2024).

5. Conclusion

This study demonstrates that a Parallel Comparative Multi-Scenario Framework combined with a structured feature selection strategy can effectively enhance classification performance in small-scale medical imaging datasets. By substantially reducing feature dimensionality while preserving discriminative information, the proposed LBP-RFE-SVM model achieved superior performance compared to alternative configurations, while the 1D-CNN variant offered a computationally efficient and representation-consistent learning approach.,The integration of handcrafted feature optimization with lightweight deep learning highlights a practical balance between accuracy and computational efficiency. Furthermore, the use of a multi-source dataset, including both the MESSIDOR benchmark and an independent clinical dataset from Hospital Universiti Sains Malaysia (HUSM), strengthens the validity of the proposed framework by demonstrating its ability to generalize across heterogeneous data distributions. From a methodological perspective, this study confirms that feature space reduction through Recursive Feature Elimination (RFE) plays a critical role in mitigating redundancy and improving model robustness in data-constrained environments. These findings suggest that carefully designed hybrid approaches can serve as viable alternatives to parameter-intensive deep learning models, particularly in resource-limited clinical settings.

Nevertheless, several limitations should be acknowledged. Although external validation was conducted using a clinical dataset, further evaluation on larger and more diverse populations is required to fully establish generalizability. In addition, the current framework focuses on image-

level classification and does not yet support precise lesion localization, which remains an important direction for future research.

Conflict of Interest
The authors declare no conflict of interest

Acknowledgment
The authors gratefully acknowledge financial support from the Ministry of Higher Education (MoHE), Malaysia, under the Fundamental Research Grant Scheme (FRGS/1/2019/ICT02/UUM/02/1). Special thanks are extended to Professor Datin Dr. Zunaina Embong, Hospital Universiti Sains Malaysia (HUSM), for her professional insights and collaborative engagement during the study. The content of this publication reflects the authors' perspectives and does not necessarily correspond to the official views of MoHE, Malaysia

References
Adriman, R., Muchtar, K., & Maulina, N. (2021). Performance Evaluation of Binary Classification of Diabetic Retinopathy through Deep Learning Techniques using Texture Feature. *Procedia Computer Science*, *179*(2020), 88–94. https://doi.org/10.1016/j.procs.2020.12.012

Ali, S. G., Wang, X., Bi, L., Jung, Y., Chen, T., & Zhang, H. (2025). Deep learning-based binocular system for automated diabetic retinopathy grading with prior clinical knowledge integration. *The Visual Computer*, *41*(8), 5675–5688. https://doi.org/10.1007/s00371-024-03745-0

Alshahrani, M., Al-jabbar, M., Senan, Ebrahim mohammed Ahmed, I. abdulrab, & Saif, J. A. M. (2023). Hybrid Methods for Fundus Image Analysis for Diagnosis of Diabetic Retinopathy Development Stages Based on Fusion Features. *Diagnostics MDPI*, *13*, 1–20. https://doi.org/10.3390/diagnostics13172783

Alshayeji, M. H., & Chandrabhasi, S. (2023). Early detection of diabetic foot ulcers from thermal images using the bag of features technique. *Biomedical Signal Processing and Control*, *79*(P2), 104143. https://doi.org/10.1016/j.bspc.2022.104143

Alyoubi, W. L., & Abulkhair, M. F. (2021). Diabetic Retinopathy Fundus Image Classification and Lesions Localization System Using Deep Learning. *Sensor*, 1–22.

Bai, Y., Zhang, X., Wang, C., Gu, H., Zhao, M., & Shi, F. (2023). Microaneurysms detection in retinal fundus images based on shape constraint with region-context features. In *Biomedical Signal Processing and Control* (Vol. 85, Issue March). https://doi.org/10.1016/j.bspc.2023.104903

Bala, R., Sharma, A., & Goel, N. (2024). Comparative Analysis of Diabetic Retinopathy Classification Approaches Using Machine Learning and Deep Learning Techniques. In *Archives of Computational Methods in Engineering* (Vol. 31, Issue 2). Springer Netherlands. https://doi.org/10.1007/s11831-023-10002-5

Barges, E., & Thabet, E. (2023). GLDM and Tamura features based KNN and particle swarm optimization for automatic diabetic retinopathy recognition system. *Multimedia Tools and Applications*, *82*(1), 271–295. https://doi.org/10.1007/s11042-022-13282-4

Costaner, L., Lisnawita, L., & Guntoro, G. (2025). Optimization Of LBP Texture Feature Extraction Using Correlation And MI For SVM Based Diabetic Retinopathy Classification. *INOVTEK Polbeng-Seri Informatika*, *10*(2), 1281–1291. https://doi.org/10.35314/4vrj4930

Ding, Y., Zhang, Z., Zhao, X., Hong, D., Cai, W., Yu, C., & Yang, N. (2022). Multi-feature fusion : Graph neural network and CNN combining for hyperspectral image classification. *Neurocomputing*, *501*, 246–257. https://doi.org/10.1016/j.neucom.2022.06.031

Du, J., Zou, B., Chen, C., Xu, Z., & Liu, Q. (2020). Automatic microaneurysm detection in fundus image based on local cross-section transformation and multi-feature fusion. *Computer Methods and Programs in Biomedicine*, *196*, 1–15. https://doi.org/10.1016/j.cmpb.2020.105687

Fathimah, F. S. N., Widjaja, S. A., & Wimbo. (2025). Retinal vessel tortuosity and fractal dimension in diabetic retinopathy. *International Journal of Retina and Vitreous*, *11*, 1–10.

https://doi.org/10.1186/s40942-025-00688-z

Gupta, A., & Kumar, V. (2023). A Deep Learning Fusion Approach for Automatic Diagnosis and Grading of Diabetic Retinopathy. *2023 14th International Conference on Computing Communication and Networking Technologies (ICCCNT)*, 1–6. https://doi.org/10.1109/ICCCNT56998.2023.10306810

Hwang, C., & Lee, T. (2025). Screening Retinal Diseases With Local Binary Patterns. *Data Science And Iot Management System*, *4*(2), 5. https://doi.org/10.64751

Ige, A. O., & Sibiya, M. (2024). State-of-the-Art in 1D Convolutional Neural Networks : A Survey. *IEEE Access*, *12*(June), 144082–144105. https://doi.org/10.1109/ACCESS.2024.3433513

Kandhasamy, J. P., Balamurali, S., Kadry, S., & Ramasamy, L. K. (2020). Diagnosis of diabetic retinopathy using multi level set segmentation algorithm with feature extraction using SVM with selective features. *Multimedia Tools and Applications*, *79*(15–16), 10581–10596. https://doi.org/10.1007/s11042-019-7485-8

Kurniawan, A. R., & Irmawati. (2024). Detection of Diabetic Retinopathy in Retinal Fundus Images Using Gabor Filters and ConvNeXt. *International Conference on Digital Medicine and Image Processing*, 47–52. https://doi.org/10.1145/3705927.3705936

Latha, D., Bell, T. B., & Sheela, C. J. J. (2022). Red lesion in fundus image with hexagonal pattern feature and two-level segmentation. *Multimedia Tools and Applications*, 26143–26161. https://doi.org/10.1007/s11042-022-12667-9

Lavanya, D., Derwin, D. J., Remya, R., Shan, B. P., Singh, O. J., & Umamaheswari, K. (2022). Diagnosis of Early-Stage Diabetic Retinopathy in Digital Fundus Images. *2022 2nd International Conference on Advances in Electrical, Computing, Communication and Sustainable Technologies, ICAECT 2022*, 1–5. https://doi.org/10.1109/ICAECT54875.2022.9807977

Mahmood, M. A. I., Aktar, N., & Kader, M. F. (2023). A hybrid approach for diagnosing diabetic retinopathy from fundus image exploiting deep features. *Heliyon*, *9*(9), 1–14. https://doi.org/10.1016/j.heliyon.2023.e19625

Makmur, N. M., Kwan, F., Dewi, A., & Indra, F. (2023). Comparing Local Binary Pattern and Level Co-occurrence Matrix for Feature Extraction in Diabetic Retinopathy Classification. *Procedia Computer Science*, *227*, 355–363. https://doi.org/10.1016/j.procs.2023.10.534

Maltsev, D. S., Kulikov, A. N., & Burnasheva, M. A. (2022). Association of Resolved Paracentral Acute Middle Maculopathy Lesions with Diabetic Retinopathy. *Journal of Current Ophthalmology*, 318–322. https://doi.org/10.4103/joco.joco

Minarno, A. E., Bagaskara, D. A., Bimantoro, F., & Suharso, W. (2025). Classification of Diabetic Retinopathy Based on Fundus Image Using. *JOIV*, *9*(January), 23–28.

Miron, M., Culea-florescu, A., & Moldovanu, S. (2023). Diabetic Retinopathy Image Classification Using Machine Learning and Local Binary Patterns Features. *2023 8th International Symposium on Electrical and Electronics Engineering (ISEEE)*, 136–139. https://doi.org/10.1109/ISEEE58596.2023.10310398

Modi, P., & Kumar, Y. (2023). Smart detection and diagnosis of diabetic retinopathy using bat based feature selection algorithm and deep forest technique. *Jurnal Computers & Industrial Engineering*, *182*(June).

Mohan, N. J., Tripti, R. M., Tanveer, G. M., & Roy, P. (2023). An efficient microaneurysms detection approach in retinal fundus images. *International Journal of Machine Learning and Cybernetics (2023)*, 1235–1252. https://doi.org/10.1007/s13042-022-01696-3

Mohd Sharif, N. A., Harun, N. H., & Yusof, Y. (2024). Colour Image Enhancement Model of Retinal Fundus Image for Diabetic Retinopathy Recognition. *Journal Of Information And Communication Technology*, *2*(2), 293–334.

Mukherjee, N., & Sengupta, S. (2023). Comparing Deep Feature Extraction Strategies for Diabetic Retinopathy Stage Classification from Fundus Images. *Arabian Journal for Science and Engineering*, *48*(8), 10335–10354. https://doi.org/10.1007/s13369-022-07547-1

Mulyani, A., Muljono, Purwanto, & Soeleman, M. A. (2024). Filter Gabor and SMOTE Method-Based Convolutional Neural Network for Diabetic Retinopathy Classification.

International Seminar on Application for Technology of Information and Communication (ISemantic), 465–470. https://doi.org/10.1109/iSemantic63362.2024.10762114

Mutawa, A. M., Al-sabti, K., Raizada, S., & Sruthi, S. (2024). A Deep Learning Model for Detecting Diabetic Retinopathy Stages with Discrete Wavelet Transform. *Applied Sciences MDPI*, 1–19. https://doi.org/10.3390/app14114428

Mutawa, A. M., Hemalakshmi, G. R., & Prakash, N. B. (2025). Randomization-Driven Hybrid Deep Learning for Diabetic Retinopathy Detection. *IEEE Access*, *13*(February), 38901–38913. https://doi.org/10.1109/ACCESS.2025.3546359

Pak, A., Ziyaden, A., Tukeshev, K., Jaxylykova, A., & Abdullina, D. (2020). Comparative analysis of deep learning methods of detection of diabetic retinopathy. *Cogent Engineering*, *7*(1). https://doi.org/10.1080/23311916.2020.1805144

Pendekal, M. J., & Gupta, S. (2022). An Ensemble Classifier Based on Individual Features for Detecting Microaneurysms in Diabetic Retinopathy. *Indonesian Journal of Electrical Engineering and Informatics*, *10*(1), 60–71. https://doi.org/10.52549/ijeei.v10i1.3522

Rahman, F., Hossain, S., & Tiang, J. (2025). Diabetes Prediction Using Feature Selection Algorithms and Boosting-Based Machine Learning Classifiers. *Diagnostics*, 1–24. https://doi.org/10.3390/diagnostics15202622

Rodriguez, D., Church, K. A., Smith, C. T., Vanegas, D., Cardona, S. M., Muzzio, I. A., Nash, K. R., & Cardona, A. E. (2024). Therapeutic Delivery of Soluble Fractalkine Ameliorates Vascular Dysfunction in the Diabetic Retina. *International Journal of Molecular Sciences*, *25*, 1–20. https://doi.org/10.3390/ijms25031727

Shamsan, A., Senan, E. M., Salameh, H., & Shatnawi, H. S. A. (2023). Predicting of diabetic retinopathy development stages of fundus images using deep learning based on combined features. *Plos One*, 1–26. https://doi.org/10.1371/journal.pone.0289555

Suhaimi, N. M., Samad, R., Rul, N., Abdullah, H., Mustafa, M., Ibrahim, M. Z., & Pebrianti, D. (2024). Comparative Analysis of Superpixel and Gabor Methods for Exudate Feature Extraction in Diabetic Retinopathy Fundus Images. *Proceedings of the 7th International Conference on Electrical, Control and Computer Engineering*, *2*, 123–136. https://doi.org/10.1007/978-981-97-3851-9

Sundar, S., & Subramanian, S. (2024). Classification of Diabetic Retinopathy Disease Levels by Extracting Spectral Features Using Wavelet CNN. *Diagnostics MDPI*, *14*(5), 1–18. https://doi.org/10.3390/ diagnostics14111093

Sundar, S., & Sumathy, S. (2023). Classification of Diabetic Retinopathy Disease Levels by Extracting Topological Features Using Graph Neural Networks. *IEEE Access*, *11*(June), 51435–51444. https://doi.org/10.1109/ACCESS.2023.3279393

Usman, T. M., Saheed, Y. K., Ignace, D., & Nsang, A. (2023). Diabetic retinopathy detection using principal component analysis multi-label feature extraction and classification. *International Journal of Cognitive Computing in Engineering*, *4*(February 2022), 78–88. https://doi.org/10.1016/j.ijcce.2023.02.002

Wee, B. F., Sivakumar, S., Lim, K. H., Wong, W. K., & Juwono, F. H. (2023). Diabetes detection based on machine learning and deep learning approaches. *Multimedia Tools and Applications*, *83*(8), 24153–24185. https://doi.org/10.1007/s11042-023-16407-5

Yadav, D., Karn, A. K., Giddalur, A., Dhiman, A., Sharma, S., Muskan, & Yadav, A. K. (2021). Microaneurysm detection using color locus detection method. *Journal of the International Measurement Confederation*, *176*(July 2020), 109084. https://doi.org/10.1016/j.measurement.2021.109084

Yadav, S., & Awasthi, P. (2022). Diabetic Retinopathy Detection Using Deep Learning And Inception-V3 Model. *International Research Journal of Modernization in Engineering Technology and Science*, *06*, 1731–1735.

Zhang, N., Jiang, Z., Li, J., & Zhang, D. (2023). Multiple color representation and fusion for diabetes mellitus diagnosis based on back tongue images. *Computers in Biology and Medicine*, *155*(February), 106652. https://doi.org/10.1016/j.compbiomed.2023.106652

Zhang, X., Ma, Y., Gong, Q., & Yao, J. (2023). Automatic detection of microaneurysms in fundus images based on multiple preprocessing fusion to extract features. *Biomedical Signal Processing and Control*, *85*(November 2022). https://doi.org/10.1016/j.bspc.2023.104879

Zhou, H., Wang, X., & Zhu, R. (2022). Feature selection based on mutual information with correlation coefficient. *Applied Intelligence*, 5457–5474.