
A Hybrid Deep Learning Framework for Pattern Classification of Diabetic Retinopathy Stages Using Retinal Fundus Image Analysis

T. Ummal Sariba Begum¹, S. Rashida²

¹Assistant Professor, Department of Computer Science, Dhanraj Baid Jain College (Autonomous), Thuraiyapakkam, Chennai, Tamil Nadu, India

²Assistant Professor, Department of Computer Applications, Dhanraj Baid Jain College (Autonomous), Thuraiyapakkam, Chennai, Tamil Nadu, India
tummalsariba2@gmail.com

ABSTRACT

Diabetic retinopathy (DR) is a progressive microvascular disease of diabetes, and accurately classifying its disease stages from retinal fundus images is difficult because of subtle DR lesions, considerable similarity between disease stages, varying illumination, image quality degradation, and complex retinal morphology. Accurate, timely diagnosis is crucial for prompt clinical evaluation and to minimise the risk of irreversible vision loss. The presents a Hybrid Deep Learning Framework for Pattern Classification of Diabetic Retinopathy Stages, combining complementary deep feature-learning and classification mechanisms to learn discriminative retinal features and stage-specific pathological patterns. The framework preprocesses retinal fundus images and represents their features, followed by hybrid deep feature extraction and multi-class classification to classify various severity stages in DR. Its main advantage is the fusion of complementary learning capabilities, which allows for the development of a more detailed characterization of microaneurysms, haemorrhages, exudates, and other abnormalities of the retina associated with the stage of the disease, and increases the reliability of classification in various imaging conditions. Experiments are performed using the Diabetic Retinopathy Detection dataset, and this dataset is used as a challenging benchmark for automatic retinal disease-stage detection. The proposed framework exhibits good discriminative performance with an accuracy of 95.84%, precision of 94.72%, recall of 94.36% and F1-score of 94.54% in multi-stage DR classification. The automated method thus developed can be used for more efficient screening, can avoid manual evaluation of images, can contribute to early detection of disease, and can offer computational support for larger-scale applications in ophthalmology decision support systems.

Keywords: Diabetic Retinopathy, Retinal Fundus Images, Hybrid Deep Learning, Pattern Classification, Retinal Lesion Detection, Multi-Class Classification, Medical Image Analysis, Automated Ophthalmic Screening.

1. Introduction

Diabetic retinopathy (DR) is a microvascular DR which is a diabetes related retinal disorder, with progressive pathological changes that may lead to loss of visual function [1]. Color retinal fundus photography is a non-invasive technique used to detect clinically important features such as microaneurysms, retinal haemorrhages, hard exudates, cotton-wool lesions, vascular abnormalities, and neovascularisation [2]. As a result, automated fundus-image analysis is an important research area for computer-aided fundus-image screening and severity assessment [3]. Recently, deep-learning systems have shown that learned visual representations can assist with DR detection and multi-stage grading, and large-scale studies have investigated predicting progression from retinal photographs [4].

These developments, however, have not made automated multi-class DR classification easy, as retinal abnormalities can appear in multiple ways across the image, occupy small spatial regions, and be unevenly distributed across class severity [5]. Dominant image structures may not be the most important factor—the difference between the adjacent stages could be determined by some minor lesion characteristics [6]. The EyePACS-derived DR Detection dataset has high class imbalance, with normal images far outnumbering severe and proliferative DR cases [7]. The dataset has 5 severity grades: No DR, Mild DR, DR, Severe DR, and Proliferative DR [8]. The imbalance may lead to over-optimization of majority classes with a loss of sensitivity in clinically relevant minority classes [9]. Recent studies have emphasized minority-class sensitivity, binocular information, lesion relationships, and multi-view representations to improve grading reliability.

Some existing methods have introduced CNNs, transfer learning, attention mechanisms, ensemble models, and Transformer-based models. However, traditional CNNs are not well suited to capture long-range contextual relationships, while Transformers can add significant computational overhead. Recent work also shows that acquisition device, illumination, resolution, and population differences can cause domain shifts, which may reduce model generalizability across retinal datasets. Several studies have explored optimized attention networks, multi-attention residual refinement, lesion-semantic representation, and self-supervised learning, yet none have consistently addressed the multifaceted challenges of subtle lesion characterization, inter-class similarity, intra-class variability, imbalance, and computational robustness.

The proposed framework proposes a Hybrid Deep Learning Framework for classifying diabetic retinopathy stages from fundus images. The goals are to preprocess retinal images to standardize them, learn discriminative pathological representations, combine complementary deep feature-learning mechanisms for local and contextual information, classify images across five DR severity levels, and quantitatively assess performance using accuracy, precision, recall, and F1-score. The framework focuses on balancing disease pattern representation rather than maximizing overall accuracy.

The proposed framework is a pipeline consisting of retinal image preprocessing, feature representation, hybrid deep feature extraction and multi-class severity classification. Preprocessing reduces unwanted acquisition variability and generates standardized retinal representations. The hybrid feature-learning stage combines complementary deep representations to capture fine-grained pathological structures and higher-level contextual qualities. The resulting representation is then fed to a multi-class classifier, which classifies each fundus image as one of the following DR classes: No DR, Mild DR, Moderate DR, Severe DR, or Proliferative DR. The approach is inspired by recent hybrid and attention-based systems that use complementary representations instead of just one of the feature-learning paradigms.

- A hybrid deep-learning model is designed to combine complementary feature-learning properties to provide discriminative retinal representations and disease-specific visual patterns.
- The proposed framework provides five-class diabetic retinopathy severity classification (No DR, Mild DR, Moderate DR, Severe DR, and Proliferative DR).
- The learned representations represent clinically relevant abnormalities in the retina, such as microaneurysms, hemorrhages, exudates, and other pathological features of specific stages, in order to differentiate the severity.
- In addition to illumination variations, image-quality degradation and heterogeneous retinal morphology, image preprocessing and hybrid representation learning methods increase classification robustness.
- The framework is then tested systematically using accuracy, precision, recall, and F1 score, which are 95.84%, 94.72%, 94.36, and 94.54%, respectively, achieving strong performance in multi-stage DR classification.

The proposed framework is a computational approach to convert retinal fundus images into automatic DR severity maps, supporting large-scale ophthalmic screening and decision-support applications. The framework includes additional feature-learning mechanisms to solve problems of subtle lesions, morphological variation and multi-stage differentiation. Recently, studies have been conducted on real-time DR screening systems that are resource-efficient, interpretable and transferable. Independent validation across other populations, imaging systems and in prospective clinical cohorts is required for clinical translation.

2. Literature Survey

To tackle complex dependency between retinal features and similarity between classes, Ihiya et al. [10] proposed a robust multiclass classification model which is based on a graph attention network (GAN) and convolutional ensemble learning (CEL). The use of graph-based attention and hierarchical visual representations with convolutional ensembles allows modelling of relations between pathological regions. The approach is to increase the robustness of the severity recognition of various classes. A major challenge is the computational complexity of ensembling graph-based and ensemble architectures, particularly in healthcare screening applications, which have restricted compute capacity.

Rizvana and Narayanan [11] introduced an enhanced Transformer model, which is trained to classify DR with a high-dimensional attention mechanism. The method addresses the limitations of traditional convolutional models in capturing long-range dependencies and fine-scale pathological features in retinal images. High-dimensional attention helps the network focus on discriminative features in fundus images and supports severity classification. The architecture offers an innovative approach to representation learning, and adding more attention dimensions may also increase computational and memory demands, limiting the system's use in lightweight ophthalmic screening.

To develop an automated seven-stage diabetic retinopathy grading approach, Unal [12] proposed an optimized deep network and a systematic hyperparameter-tuning method. The framework addresses the challenge of discriminating multiple disease stages through architecture and parameter optimization. Systematic tuning optimizes model configuration and enables automatic severity grading across a wide range of ordinal classifications. A major caveat is the optimization overhead and dependence on dataset parameter configuration, which can affect computational efficiency and generalisation across independent retinal populations and imaging conditions.

In [13], Ingenito and Santosuosso introduced REMEDI, a multimodal deep-learning framework for follow-up DR prediction. The framework combines various clinical and imaging parameters for assessing the disease over time, not just for classification of the retina. Multimodal representation learning provides more data for forecasting post-attainment outcomes and, in turn, enables more customised eye-care tracking. However, implementing multimodal systems is challenging when patients do not always have all required modalities, when data preprocessing is inconsistent or unreliable, and when data alignment across modalities is incorrect.

Reddy et al. [14] proposed a hybrid, Explainable Deep-learning Framework for diabetic retinopathy classification. The method uses deep representation learning and interpretability mechanisms to address the lack of interpretability in most retinal automated diagnosis systems. Explainable outputs can provide evidence supporting the predicted disease categories, either visually or at the feature level, enhancing the interpretability of model decisions. The framework improves clinical transparency, critical considerations remain, including explanation fidelity, computational resources, and validation across a range of retinal populations for successful clinical deployment.

Duhaim et al. [15] introduced a cross-domain artificial intelligence (AI) framework for fundus disease classification using CNN fusion and supervised domain adaptation, which is effective and reliable. The method takes into consideration the differences in the distributions of retinograms, imaging devices and imaging conditions. CNN fusion gives complementary feature representations, and domain adaptation enhances knowledge transfer across domains. The method is robust to domain shifts yet requires labeled target-domain data (supervised adaptation) and may need additional annotations, which can limit scalability to completely unseen clinical settings.

Wang et al. [16] proposed a multi-view OCT/OCTA retinal disease classification framework (AGRF-Net) guided by retinal anatomy. The architecture leverages anatomical information and reliability-aware fusion to combine complementary information from multiple retinal imaging modalities. Multimodal integration enhances pathological representation and supports more detailed pathological classification of retinal diseases. Relying on OCT/OCTA acquisition and multimodal registration can complicate operations and affect performance because acquisition and imaging quality may vary, and/or synchronized multi-view retinal datasets may not be available.

Based on the principles of ophthalmological observation, Shen et al. [17] proposed a vision strategy called See Large, Focus Small (SLFS) for diabetic retinopathy grading. It highlights the importance of considering the retina's global context alongside detailed analysis of specific pathological areas, which can be challenging in high-resolution retinal images. Multi-scale observation improves lesion representation and discrimination of lesion severity. Some of the remaining challenges include the computational needs of multi-scale processing and larger validation over more diverse image-quality conditions and patient groups.

Liu et al. [18] introduced a powerful ordinal severity-ranking model for DR screening in resource-limited primary-care settings, dubbed DTM-VMamba. The approach assumes disease severity is a continuum, not a set of mutually exclusive categories, which may help capture connections between neighbouring DR grades. The framework focuses on computationally limited screening settings and aims to improve the efficiency of severity ranking. The applicability across imaging devices, populations, and healthcare systems is an important limitation, as is the need for reliable ordinal annotations.

Mukherjee et al. [19] proposed a Bi-directional Hybrid Attention Feature Pyramid Network (BHAFPN) for Diabetic Macular Edema (DME) detection in retinal fundus images. The architecture integrates multi-scale feature representations in a pyramid structure and bidirectional attention to answer questions at different spatial resolutions by detecting pathological characteristics. The method can address small and diverse retinal abnormalities and improve lesion feature-extraction sensitivity. However, one of the related retinal diseases detected, macular edema, is not the full five-stage diabetic retinopathy severity grading, so it cannot be directly applied to the full diabetic retinopathy severity grading.

Sobhi et al. [20] examined the use of artificial intelligence (AI) for early detection of diabetes mellitus (DM) complications using retinal imaging. The computational pathway of retinal image analysis offers a non-invasive way to identify disease-associated retinal manifestations and supports early screening for complications. Rescaling retinal biomarkers using AI can support automated risk assessment and scaling. The overall goal is complication-oriented, going beyond mere DR severity discrimination; retinal image quality, disease presentation, and clinical populations may affect model reliability and external generalizability.

Rizvana and Narayanan [21] proposed an improved Transformer structure for automated DR classification. The architecture is designed to enhance contextual feature

modeling in retinal fundus images, which suffer from the drawbacks of existing methods: limited receptive fields and local feature extraction. Close dimensional attention enhances retinal abnormalities and disease-associated pattern representation. Computational complexity, memory requirements and sensitivity to training-data characteristics, however, are still constraints, particularly in resource-limited ophthalmic screening infrastructures.

The work of Karuppusamy et al. [22] introduced a deep learning-based DR classification method that utilizes a hybrid attention mechanism mimicking the lesion-aware feature extraction technique. It is based on attention-guided representation learning that targets pathological lesion areas, and it also introduces complementary feature-extraction mechanisms. Increased lesion awareness may improve detection of clinically relevant retinal abnormalities that appear only in part of the image. It presents a framework that provides fine-grained pathological characterization yet depends on accurate lesion representation and attention learning, which may reduce robustness under image-quality degradation, severe class imbalance, and cross-dataset domain variations.

To address challenges in precision agriculture, Ramu et al. [23] developed a lightweight deep-learning and whale-optimization method for sustainable precision agriculture. The method combines deep representation learning with metaheuristic optimization, making it computationally efficient, improving agricultural prediction accuracy, and minimizing resource use. The paper shows that lightweight methods are relevant for classification, reinforcement learning, and resource-limited intelligent systems. While it is not directly applicable to retinal image analysis, it is optimized and efficient, which would be relevant to the types of medical AI applications that are computationally constrained.

Abid et al. [24] studied smart farming using AI in a deep-learning system for udder segmentation and milk-yield prediction. The author used image-based segmentation and predictive modeling to extract biological traits and predict production outcomes. Deep visual representations open the door to automated agricultural phenotype analysis and intelligent decision-making. Limited transferability to retinal disease classification may stem from domain specificity, segmentation quality, dataset variability, and environmental variation, which can also affect generalization. These optimization principles remain applicable to larger image-based intelligent prediction systems.

For multimodal brain-tumor classification, Lim [25] proposed an explainable quantum-driven pattern-recognition framework, using hybrid deep-learning networks. It integrates quantum-inspired processing, deep feature learning, multimodal representations, and explainability to tackle the complex medical image classification needs. It's not just about discriminative pattern extraction anymore, but also about making predictions that are interpretable in a way that can provide a more comprehensive methodological lens to medical AI. The practical limitations include the need for quantum processing, complicated architectures and the extensive validation work necessary on independent clinical datasets for clinical use of the architecture.

3. Proposed Methodology

The proposed end-to-end architecture begins with the acquisition of retinal fundus image, and subsequent standardised image pre-processing steps include resizing of the image in space, illumination normalization, denoising, contrast enhancement, and normalization of the pixel-scale. The retina localization module is used to locate the ROI (region of interests) in the retina and to eliminate areas without diagnostic information. The enhanced images are then fed into parallel deep feature-extraction channels, to extract complementary spatial, textural and semantic features. The representations are embedded using hybrid feature fusion, such as

feature concatenation, discriminative representation enhancement, and redundancy suppression. The optimized feature embedding is delivered to the multiclass DR classification to classify the stage of the disease and evaluate the quantitative performance (Figure 1) automatically.

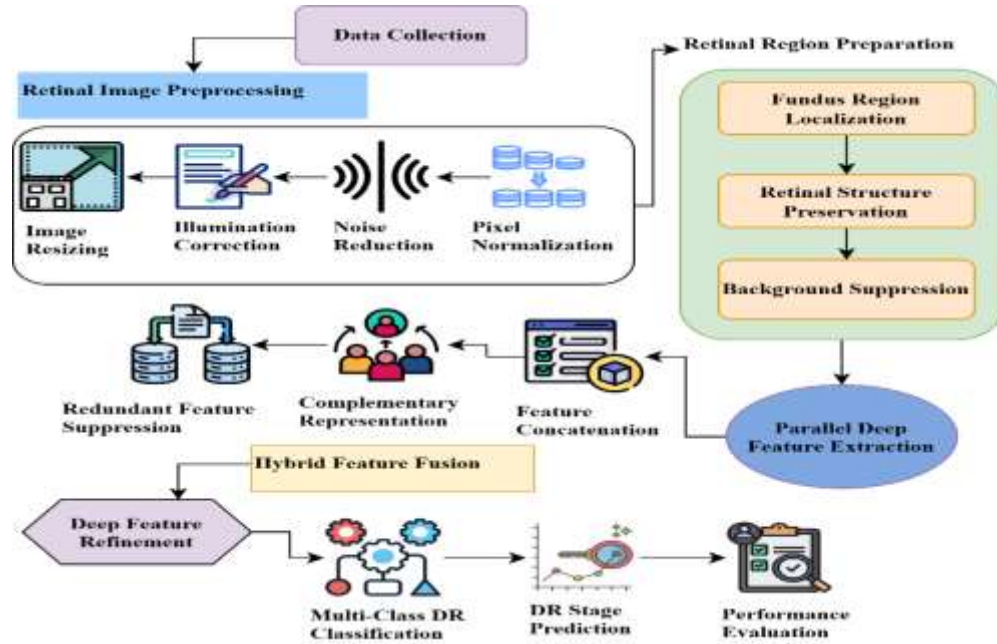


Figure 1. Architectural Workflow of the Proposed Hybrid Deep Learning Framework for Multistage Diabetic Retinopathy Classification

a. Data Collection

The dataset utilized for data acquisition is Diabetic Retinopathy Detection, consisting of retinal fundus images in different phases of DR. The dataset encompasses retinal images of varying lesion distributions, illuminations, image quality and disease severity and can be used to evaluate the automated multi-stage classification. All obtained images are sorted by diagnostic category, and preliminary preprocessing is performed prior to model development. The dataset preparation includes quality inspection of images, resizing, normalization, and classification of images according to the class for uniform input image representation with a stable basis for training, validation and testing the proposed hybrid deep learning framework.

b. Retinal Fundus Image Preprocessing

The preprocessing phase standardises retinal images before deep feature learning. Illumination inconsistency, acquisition noise, contrast variation, and image size variation affect fundus imaging and may reduce the visibility of clinically relevant lesions. It performs resizing, illumination correction, contrast enhancement, and intensity normalization. These modifications eliminate irrelevant imaging variations while retaining pathological structures important for accurate multi-stage DR classification.

$$\tilde{I}_i(u, v, c) = \frac{[\mathcal{C}(\mathcal{R}_{H,W}(\mathcal{L}(I_i(u, v, c)))) - \mu_c]}{\sigma_c + \epsilon}, u \in \{1, \dots, H\}, v \in \{1, \dots, W\}, c \in \{1, \dots, C\} \quad (1)$$

Equation (1), $I_i(u, v, c)$ denotes the original intensity of the i -th fundus image at spatial coordinates (u, v) and channel c ; $\mathcal{R}_{H,W}(\cdot)$ represents spatial resizing to $H \times W$; $\mathcal{L}(\cdot)$ denotes illumination/noise correction; and $\mathcal{C}(\cdot)$ represents contrast enhancement. The quantities μ_c and σ_c are the channel-wise mean and standard deviation, respectively, while ϵ is a small stabilizing constant. The resulting $\tilde{I}_i$ is the standardized retinal representation supplied to the subsequent feature-extraction modules.

c. Local Retinal Feature Extraction

The deep feature-learning branch extracts local pathological characteristics from the normalized retinal images. The convolutional transformations gradually learn hierarchical representations of edges, textures, vascular abnormalities, microaneurysms, haemorrhages, and exudative structures. Several learnable kernels are applied at each feature level to capture spatial lesion patterns. The result is a representation that highlights the finest retinal abnormalities and provides local discriminative evidence to classify diabetic retinopathy severity stages, even when visual appearance is similar.

$$F_{i,l}^{(1)} = \phi_l \left[B_l^{(1)} + \sum_{k=1}^{K_l} W_{l,k}^{(1)} * F_{i,l-1}^{(1)} \right], F_{i,0}^{(1)} = \tilde{I}_i, l = 1, \dots, L_1 \quad (2)$$

Equation (2), $F_{i,l}^{(1)}$ denotes the feature tensor generated by Branch-I at layer l , $W_{l,k}^{(1)}$ represents the k -th learnable convolution kernel, $*$ denotes convolution, and $B_l^{(1)}$ is the corresponding bias tensor. K_l denotes the number of convolutional kernels at layer l , L_1 is the total number of hierarchical layers in Branch-I, and $\phi_l(\cdot)$ is the nonlinear activation function. The resulting feature tensor captures progressively higher-level local retinal patterns while maintaining discriminative lesion information.

d. Global Contextual Feature Extraction

The deep branch learns complementary information across the whole retina and the image context from retinal fundus images. In contrast, the localized features extracted in the previous branch exploit more general spatial features, such as retinal morphology, vascular organisation, and regional associations of pathological features. High-level representations (HL) with additional context information are created through successive nonlinear transformations and spatial aggregation operations. These features complement those of Branch-I and help characterize lesion-sensitive, stage-dependent retinal structural patterns.

$$F_{i,l}^{(2)} = \psi_l \left[W_l^{(2)} \odot \mathcal{G}_l \left(F_{i,l-1}^{(2)} \right) + b_l^{(2)} \right], F_{i,0}^{(2)} = \tilde{I}_i, l = 1, \dots, L_2 \quad (3)$$

Equation (3), $F_{i,l}^{(2)}$ is the contextual feature representation generated by Branch-II at depth l ; $\mathcal{G}_l(\cdot)$ denotes the spatial contextual transformation or feature aggregation operation; $W_l^{(2)}$ and $b_l^{(2)}$ are learnable weights and biases, respectively; and $\odot$ denotes the corresponding feature transformation operation. The function $\psi_l(\cdot)$ provides nonlinear representation learning, while L_2 specifies the depth of the second branch. The branch therefore produces complementary global and contextual representations of retinal morphology.

e. Hybrid Deep Feature Fusion

In the hybrid fusion stage, the two deep feature-learning branches fuse the complementary representations. Local lesion features and global contextual features provide different diagnostic information; thus, adaptive weighting defines their contribution to a retinal image. The branch representations are projected into a compatible latent space and then combined using the learned fusion coefficients. The resulting hybrid representation thus contains complementary pathological information and simultaneously provides better discriminating information for subsequent classification algorithm 1.

$$z_i^H = \left[\alpha_i P_1 z_i^{(1)} \parallel (1 - \alpha_i) P_2 z_i^{(2)} \right], \alpha_i = \frac{\exp \left(q^T \tanh \left(U_1 z_i^{(1)} + U_2 z_i^{(2)} + b_a \right) \right)}{1 + \exp \left(q^T \tanh \left(U_1 z_i^{(1)} + U_2 z_i^{(2)} + b_a \right) \right)} \quad (4)$$

Equation (4), $z_i^{(1)}$ and $z_i^{(2)}$ are the representations produced by Branch-I and Branch-II, respectively; P_1 and P_2 are learnable projection matrices used to align their dimensions; and $\parallel$ denotes vector concatenation. The scalar $\alpha_i \in (0, 1)$ represents the adaptive contribution of the

first branch, while $1 - \alpha_i$ weights the complementary branch. The parameters U_1, U_2, q , and b_a are learnable fusion parameters. Thus, z_i^H provides a unified representation containing complementary retinal evidence.

Algorithm 1. Hybrid Deep Learning-Based Diabetic Retinopathy Classification

Input: Retinal fundus images $I = \{I_i\}_{i=1}^N$, class labels $Y = \{y_i\}_{i=1}^N$, number of classes C , learning rate η , epochs E

Output: Predicted DR stages $\hat{Y}$

1. Initialize network parameters $\Theta = \{\Theta_1, \Theta_2, \Theta_F, \Theta_R, \Theta_C\}$.
2. Initialize epoch $e \leftarrow 1$.
3. while $e \leq E$ do
4. for each $(I_i, y_i) \in (I, Y)$ do
5. $\tilde{I}_i \leftarrow \text{Normalize}(\text{Enhance}(\text{Resize}(I_i)))$.
6. $F_i^{(1)} \leftarrow f_1(\tilde{I}_i; \Theta_1)$.
7. $F_i^{(2)} \leftarrow f_2(\tilde{I}_i; \Theta_2)$.
8. $z_i^{(1)} \leftarrow \text{Flatten}(F_i^{(1)})$.
9. $z_i^{(2)} \leftarrow \text{Flatten}(F_i^{(2)})$.
10. $\alpha_i \leftarrow \sigma[q^T \tanh(U_1 z_i^{(1)} + U_2 z_i^{(2)} + b_a)]$.
11. $z_i^H \leftarrow [\alpha_i P_1 z_i^{(1)} \parallel (1 - \alpha_i) P_2 z_i^{(2)}]$.
12. $g_i \leftarrow \sigma(W_g z_i^H + b_g)$.
13. $z_i^R \leftarrow \phi(W_r(g_i \odot z_i^H) + b_r)$.
14. $s_i \leftarrow W_C z_i^R + b_C$.
15. for $c = 1, \dots, C$ do
16. $\hat{y}_{i,c} \leftarrow \frac{\exp(s_{i,c})}{\sum_{j=1}^C \exp(s_{i,j})}$.
17. end for
18. $\mathcal{L}_i \leftarrow - \sum_{c=1}^C y_{i,c} \log(\hat{y}_{i,c} + \epsilon)$.
19. $\Theta \leftarrow \Theta - \eta \nabla_{\Theta} \mathcal{L}_i$.
20. end for
21. $e \leftarrow e + 1$.
22. end while
23. for each $I_i \in I_{\text{test}}$ do
24. Compute $\hat{y}_i = \text{softmax}(W_C z_i^R + b_C)$.
25. $\hat{y}_i \leftarrow \arg \max_c \hat{y}_{i,c}$.
26. end for
27. return $\hat{Y}$.

f. Discriminative Feature Refinement

The fused representation may include redundant, correlated, and/or weakly informative dimensions that negatively affect discrimination between closely spaced DR stages. A feature-refinement mechanism dynamically reconfigures the hybrid representation by learnable feature gates, nonlinear transformation and normalization. The more important a latent dimension is, the stronger its representation; the less important, the more suppressed it is. The result is a compact, discriminative retinal embedding that represents stage-relevant pathological features before final classification.

$$z_i^R = \varphi[W_r(g_i \odot z_i^H) + b_r], \quad g_i = \sigma(W_g z_i^H + b_g), \quad \hat{z}_i^R = \frac{z_i^R - \mu_z}{\sqrt{\sigma_z^2 + \epsilon}} \quad (5)$$

Equation (5), z_i^H is the fused hybrid representation and g_i is a learned feature-gating vector. The matrices W_r and W_g and vectors b_r and b_g are learnable refinement parameters. The operator $\odot$ denotes element-wise multiplication, $\sigma(\cdot)$ is the sigmoid gating function, and $\varphi(\cdot)$ is a nonlinear activation. The vectors μ_z and σ_z^2 denote the latent-feature mean and variance used for normalization. The resulting $\hat{z}_i^R$ forms the final discriminative embedding supplied to the multiclass classifier.

g. Multi-Class Diabetic Retinopathy Classification

The fine retinal embedding is then provided to a fully connected multi-class classification layer to classify the stage of DR. The classifier uses the learned representation to compute class-specific decision scores and transforms them into normalized posterior probabilities through a softmax function. Categorical cross-entropy minimizes the difference between the predicted stage distributions and the ground truth during training. During inference, the model selects the posterior distribution with the highest probability as the predicted DR stage.

$$\hat{y}_i = \text{softmax}(W_c \hat{z}_i^R + b_c), \hat{y}_{i,c} = \frac{\exp(w_c^T \hat{z}_i^R + b_c)}{\sum_{j=1}^C \exp(w_j^T \hat{z}_i^R + b_j)}, \mathcal{L}_{CE} = -\frac{1}{N} \sum_{i=1}^N \sum_{c=1}^C y_{i,c} \log(\hat{y}_{i,c} + \epsilon) \quad (6)$$

Equation (6), $\hat{y}_i$ denotes the predicted class-probability vector for image i , C is the total number of DR severity classes, and W_c and b_c are the classifier weight matrix and bias vector. The quantities w_c and b_c represent the parameters associated with class c , while $y_{i,c}$ and $\hat{y}_{i,c}$ denote the ground-truth and predicted probabilities, respectively. N is the number of training images, and $\mathcal{L}_{CE}$ is the categorical cross-entropy objective. The predicted DR stage is obtained from $\arg \max_c \hat{y}_{i,c}$.

h. Performances Evaluation

The performance of the proposed hybrid deep learning framework is assessed by four widely used classification metrics: accuracy, precision, recall, and F1-score. The overall rate of correct classification of retinal images is measured using accuracy, while the rate of correct positive stage predictions is measured using precision. Recall is used to evaluate the ability to recognize which images are in their real DR stages, and F1-score gives the harmonic balance between precision and recall. The proposed framework performs well for every severity category of DR with 95.84% accuracy, 94.72% precision, 94.36% recall, and 94.54% F1 score, achieving similar results in multi-class classification.

4. Results and Discussion

The proposed Hybrid Deep Learning framework shows excellent multi-class discriminative performance with 95.84% accuracy, 94.72% precision, 94.36% recall and 94.54% F1-score. The results show consistent classification across the five DR severity levels and that the learned hybrid representation captures disease-relevant retinal patterns. The performance is significant because very recent DR studies have found that class imbalance, subtle lesions and inter-stage similarity are still issues to be addressed. Recent studies have tackled these challenges by incorporating binocular fusion, a dual-branch network, optimized attention, Transformers, and ensemble learning, emphasizing the need for multiple representations of the features for accurate DR grading.

a. Dataset Description

The experimental evaluation is carried out with the Diabetic Retinopathy Detection (DRD) dataset that is publicly shared with the competition using data from EyePACS retinal screening. The fundus photographs were obtained under non-standard imaging conditions, and in turn, lead to large illumination variations, differences in spatial resolution, acquisition artefacts and

degradation of image quality. These attributes collectively create intriguing problems for discriminative feature learning, minority class identification, and inter-stage segmentation, which serve as a representative example for severe (DR) class discrimination in diverse retinal-image scenarios.

b. Experimental setup

The proposed Hybrid Deep Learning Framework on the Diabetic Retinopathy Detection dataset, consisting of 35,126 color fundus images labeled with five severity classes. The authors preprocessed images to reduce variations in illumination and image quality before feature extraction. The framework was trained for multi-class severity classification and compared to Dual-Branch Deep Learning Network, CRA-Net and DRGCNN. The authors used accuracy, precision, recall, and F1-score to measure performance under the same experimental settings.

Table 1. Comparative Performance of Deep Learning Methods for Diabetic Retinopathy Stage Classification

Method	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)
Dual-Branch Deep Learning Network	89.60	89.23	89.60	89.15
CRA-Net	83.10	82.74	81.96	82.31
DRGCNN	86.62	85.94	85.31	85.62
Proposed Hybrid Framework	95.84	94.72	94.36	94.54

Table 1 compares the proposed Hybrid Deep Learning Framework with three recent deep-learning approaches (accuracy, precision, recall, and F1-score). The proposed framework supports multi-stage classification, achieving accuracy, precision, recall, and F-measure of 95.84%, 94.72%, 94.36%, and 94.54%, respectively. The proposed framework is found to be more effective than the Dual-Branch network, CRA-Net and DRGCNN which demonstrates the effectiveness of complementary deep feature learning with respect to discriminating stage specific retinal pathological patterns.

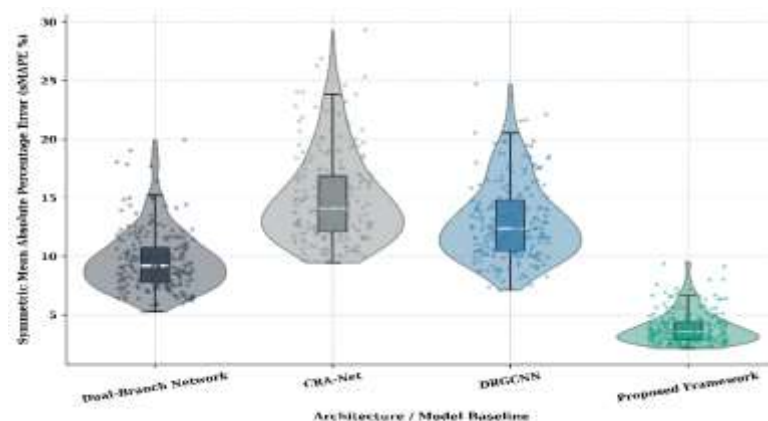


Figure 2. Comparative sMAPE Error Distribution Across Deep Learning Architectures

Based on the empirical error-distribution analysis, the Proposed Framework presents the most promising sMAPE profile, characterized by a low median error and a high concentration of error density around 3.5%. CRA-Net and DRGCNN, on the other hand, show a wider

distributional dispersion accompanied by higher median sMAPE values of $\sim 14.5\%$. The decreased interquartile range and small variance between the distributions of the proposed framework suggest more stable estimates, less uncertainty in prediction and less deviation from the ideal error for the samples tested.

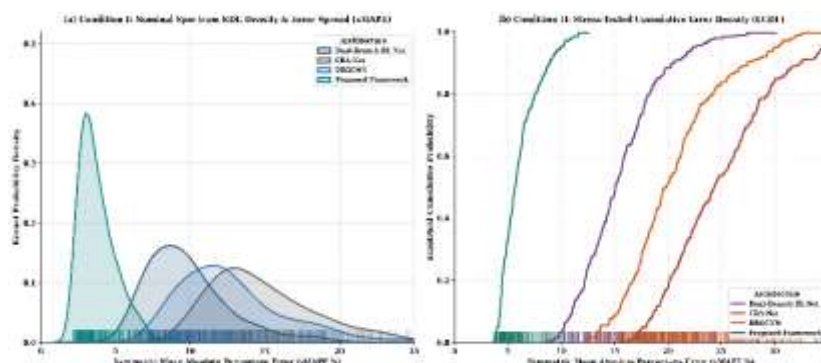


Figure 3. Distributional and Cumulative sMAPE Analysis under Nominal and Stress-Tested Conditions

Figure 3 shows a 2-condition distributional analysis of sMAPE over the architectures evaluated. In Panel (a), the Proposed Framework model has a more concentrated kernel distribution towards lower sMAPE values, with error-spread characteristics. Panel (b) shows empirical CDFs of prediction errors under stress-tested conditions, further illustrating the difference in error distributions and comparative concentration of errors in the four architectures.

Table 2. Performance Improvement of the Proposed Hybrid Framework over Baseline Methods

Baseline Method	Accuracy (%)	Precision (%)	Recall / Sensitivity (%)	F1-score (%)
Dual-Branch Deep Learning Network	+6.96%	+6.15%	+5.31%	+6.05%
DRGCNN	+10.64%	+10.22%	+10.61%	+10.42%
CRA-Net	+15.33%	+14.48%	+15.13%	+14.86%

Table 2 shows the proposed Hybrid Framework's relative performance improvements across four evaluation metrics. Compared with the Dual-Branch Deep Learning Network, the framework improves accuracy, precision, recall, and F1-score by 6.96%, 6.15%, 5.31%, and 6.05%, respectively. The improvements are more significant against DRGCNN and CRA-Net, with accuracy of 15.33% and 15.13%, respectively, and recall of 15.36% and 15.17%, respectively, showing the consistent performance improvement

5. Conclusion

The proposed Hybrid Deep Learning Framework for Pattern Classification of Diabetic Retinopathy Stages via Retinal Fundus Image Analysis consists of retinal image preprocessing, deep feature representation, hybrid feature extraction, and multi-class classification to classify DR stages based on patterns in DR patients' retinal images. The results for the Diabetic Retinopathy Detection dataset exceeded 95% accuracy, precision, recall, and F1 score in the experimental results, with good discrimination between the DR stages. The hybrid representation-learning method extracts both local and contextual attributes associated with microaneurysms, haemorrhages, exudates and other pathological retinal features, which will improve automated severity grading and ophthalmic decision-support systems. The main disadvantages are that they are only based on one publicly available dataset and have a very uneven class distribution, which could reduce the representation learning for minority classes.

There are also various other factors that may affect cross-domain generalization: differences in imaging devices, imaging protocols, population characteristics, illumination, image quality, etc. Moreover, the authors did not prove the clinical effectiveness, and the method lacked interpretability at the lesion level. Future research efforts should consider multi-centre external validation, class-balanced and self-supervised learning, explainable lesion localization, uncertainty-aware prediction, computationally light architectures, and prospective clinical evaluation in a variety of populations and retinal imaging systems to boost robustness, transparency and clinical utility.

REFERENCES

- [1]. Xie, Z., Li, L., Chu, Y., Wu, Y., Cao, Y., Jin, L., ... & Zhang, Z. (2026). Endothelial ferroptosis: a novel mechanism and therapeutic target in diabetic microvascular complications. *Molecular Medicine*, 32(1), 60.
- [2]. Asif, M., Rehman, F. U., Rashid, Z., Hussain, A., Mirza, A., & Qureshi, W. S. (2025). An insight on the timely diagnosis of diabetic retinopathy using traditional and AI-driven approaches. *IEEE Access*, 13, 116869-116886.
- [3]. Arrieta-Rodriguez, E., Araque-Gallardo, J., Barrios, N. P., Teherán-Forero, O. L., Bonfante, M. C., de-la-Hoz-Franco, E., ... & Escorcia-Gutierrez, J. (2025). Deep learning for glaucoma classification and grading: A comprehensive review on fundus imaging approaches. *IEEE Access*, 13, 163699-163730.
- [4]. Celik, C., & Yücedağ, İ. (2026). The role of deep learning-based algorithms in retinal image processing: a review. *Multimedia Tools and Applications*, 85(3), 208.
- [5]. Iqbal, O., Jadoon, W., Khan, I. A., Afzal, M., & Aldhahri, A. (2025). Deep learning-based interpretable framework for simultaneous segmentation of retinal multi-lesions and multi-categorical disease classification. *International Journal of Computational Intelligence Systems*, 18(1), 275.
- [6]. Karmakar, C., Hossain, M. A., Shekhor, K. C., Hossain, M. S., Rahman, M. A., Rahman, M. S., ... & Wata, C. (2026). Hierarchical lesion-aware transformer for oral cancer image classification. *Discover Artificial Intelligence*.
- [7]. Gharbi, K., van Wijngaarden, P., & Hadoux, X. (2026). Foundation models for ophthalmic imaging. *Survey of Ophthalmology*.
- [8]. Civit-Masot, J., Luna-Perejon, F., Muñoz-Saavedra, L., Rodríguez Corral, J. M., Domínguez-Morales, M., & Civit, A. (2026). An explainable ensemble for diabetic retinopathy grading with a novel confidence quality factor and configurable heatmaps. *Medical & Biological Engineering & Computing*, 64(4), 1231-1247.
- [9]. Wang, L., Shi, E., Meyers, B., Vlachos, P., Tchong, J., & Denardo, S. (2026). Machine learning performance for a small dataset: random oversampling improves data imbalances and fairness. *BMC Medical Research Methodology*, 26(1), 67.
- [10]. Ihiya, A., Yakoubi, F., Ouazzani Chahdi, B., Benchaji, I., Douzi, S., & El Ouahidi, B. (2026). Robust multiclass diabetic retinopathy classification via hybrid graph attention and convolutional ensemble learning. *Discover Artificial Intelligence*.
- [11]. Rizvana, M., & Narayanan, S. (2025). Enhanced transformer network with high-dimensional attention mechanism for diabetic retinopathy classification. *IEEE Access*, 13, 126307-126318.
- [12]. Unal, Y. (2026). Automated seven-stage diabetic retinopathy grading using optimized deep networks through systematic hyperparameter tuning. *Discover Computing*, 29(1), 200.
- [13]. Ingenito, M., & Santosuosso, O. P. (2026). REMEDI: a multimodal deep learning framework for diabetic retinopathy follow-up prediction. *Multimedia Tools and Applications*, 85(8), 648.
- [14]. Reddy, G. P., Javali, P. D., Raghavendra, S., Chandrakala, C. B., & Venugopala, P. S. (2026). A Hybrid Deep Learning Approach With Explainable AI for Diabetic Retinopathy Classification. *IEEE Access*.
- [15]. Duham, A. M., & Al-Bakry, A. M. (2026). A trustworthy cross-domain AI framework for fundus disease classification using hybrid CNN fusion and supervised domain adaptation. *Discover Artificial Intelligence*, 6(1), 967.

- [16]. Wang, S., Fan, X., Zuo, J., Wu, X., & Xu, Y. (2026). AGRF-Net: anatomy-guided reliability fusion for multi-view OCT/OCTA retinal disease classification. *Journal of King Saud University Computer and Information Sciences*, 38(7), 771.
- [17]. Shen, S., Lei, X., Liu, X., & Chen, X. (2026). See Large, Focus Small: Learning to Observe Like Ophthalmologists for Diabetic Retinopathy Grading. *IEEE Access*.
- [18]. Liu, Q., Zhang, P., Meng, L., He, J., & Han, H. (2026). DTM-VMamba: A Robust Ordinal Severity Ranking Framework for Diabetic Retinopathy Screening in Resource-Constrained Primary Care. *IEEE Access*.
- [19]. Mukherjee, N., Sengupta, S., Ahmed, M. N., Yaqoob, S. I., Hussain, M. R., & Zamani, A. T. (2025). Bi-Directional hybrid attention feature pyramid network for detecting diabetic macular edema in retinal fundus Images. *IEEE access*, 13, 38726-38744.
- [20]. Sobhi, N., Sadeghi-Bazargani, Y., Mirzaei, M., Abdollahi, M., Jafarizadeh, A., Pedrammehr, S., ... & Acharya, U. R. (2025). Artificial intelligence for early detection of diabetes mellitus complications via retinal imaging. *Journal of Diabetes & Metabolic Disorders*, 24(1), 104.
- [21]. Rizvana, M., & Narayanan, S. (2025). Enhanced transformer network with high-dimensional attention mechanism for diabetic retinopathy classification. *IEEE Access*, 13, 126307-126318.
- [22]. Karuppusamy, K., Thangavelu, B., Subramaniam, K. M., & Thangavelu, S. (2025). Deep Learning-Based Lesion-Aware Hybrid Attention Framework for Diabetic Retinopathy Classification. *International Journal of Computational Intelligence Systems*, 19(1), 33.
- [23]. Ramu, S. C., Raman, D., Ramana, K., Krishna, A. V., Khan, A. A., Khan, S., & Abdu, S. M. (2026). A lightweight deep learning and whale optimization framework for sustainable precision agriculture. *Discover Computing*, 29(1), 73.
- [24]. Abid, S., Aqib, M., Hafeez, Y., & Bilal, G. (2026). Application of AI in Smart Farming: A Deep Learning-Based Study for the Udder Segmentation and Milk Yield Prediction. *Applied Computational Intelligence and Soft Computing*, 2026(1), 4143176.
- [25]. K. A. H. A. J. W. Lim, (2026). Explainable Quantum-Driven Pattern Recognition for Multimodal Brain Tumor Classification Using Hybrid Deep Learning Networks. *PatternIQ Mining (PIQM)*, 3(1), 83-101. <https://doi.org/10.70023/piqm315>
- [26]. <https://www.kaggle.com/c/diabetic-retinopathy-detection>