Journal of Innovative Image Processing (ISSN: 2582-4252)
www.irojournals.com/iroiip/

Deep Denoising with Dual-Branch Confidence-Aware Network for Diabetic Retinopathy Classification

Kokila A.[1*], Shankar R.[2], Duraisamy S.[3]

Department of Computer Science, Chikkanna Government Arts College, Vivekananda Nagar, Konganagiri, Tiruppur, India.

E-mail: [1*]195kokila@gmail.com, [2]shankarcgac@gmail.com, [3]sdsamy.s@gmail.com

Orcid ID: [1*]0009-0005-3676-9297, [2]0000-0001-8790-3197, [3]0000-0001-5899-4560

Abstract

Diabetic retinopathy (DR) is a condition resulting from prolonged increased blood sugar that destroys the retinal blood vessels. Early diagnosis could prevent permanent and complete blindness. Mostly, retinal fundus images are broadly utilized for the detection and localization of intra-retinal limitations. This article offers a Hybrid Deep Denoising and Dual-Branch Confidence-Aware Classification for Accurate Diabetic Retinopathy (HDD-DBCADR) from retinal images. The main purpose of the HDD-DBCADR framework is the combination of a hybrid denoising network and confidence-aware dual-branch classification, which offers a promising direction for developing reliable and clinically applicable automated DR screening systems. The proposed model involves the design of a hybrid deep denoising network, which integrates a residual convolutional neural network with an autoencoder-based refinement module to enhance image quality. For robust feature representation, a dual-branch architecture is applied, where the local branch exploits a multi-scale convolutional neural network for deriving fine-grained spatial features, and the global branch leverages a dual attention vision transformer to model long-range dependencies and contextual data. At last, a dual-branch confidence-aware classification mechanism adaptively combines predictions from both branches, depending on their confidence scores, thereby improving classification reliability. The simulation investigation of the HDD-DBCADR approach is performed using a benchmark DR detection dataset from the Kaggle repository. Extensive comparative results report the encouraging performance of the proposed HDD-DBCADR model against recent state-of-the-art models with models. achieving a maximum accuracy of 98.96% and SSIM of 0.9760. The inclusion of the denoising network with the deep learning-based DR classification model accomplishes enhanced diagnostic performance, making it useful for real-time clinical settings.

Keywords: Image Denoising; Retinal Fundus Images, Diabetic Retinopathy, Deep Learning, Transformer, Dual Attention Model.

1. Introduction

Diabetes mellitus (DM) is a long-term metabolic illness caused by inadequate insulin production in the body. Diabetes places a substantial strain on healthcare services and induces major problems if not appropriately controlled. For individuals diagnosed with diabetes, nearly

*** Corresponding Author**

DOI: https://doi.org/10.36548/jiip.2026.2.015
Received: 26.03.2026, received in revised form: 29.04.2026, accepted: 16.05.2026, published: 30.05.2026

one-third develop diabetic retinopathy (DR). Eventually, excessive glucose deteriorates the vessel walls, affecting fluid leakage, microaneurysms, and swelling of retinal tissues [1]. In the initial phase of the disease, early intervention could avoid extreme vision loss by controlling the issue before incurable injury occurs. Physicians could slow or stop the disorder's progression through interventions such as improved glycemic regulation. This lessens patients' risk of patients' developing macular edema or proliferative DR. Consequently, prioritizing timely screening and utilizing cutting-edge approaches could substantially decrease the burden of DR, improve patients' outcomes, and prevent the vision loss that impacts millions of people worldwide. Regular eye check-ups are suggested for people with diabetes, as the initial phase of DR frequently shows clear symptoms [2]. Physicians usually employ retinal scans, such as fundus images and optical coherence tomography (OCT), to recognize variations in the retina. These images enable a thorough analysis of the retinal structure and any abnormalities that could indicate the presence of DR.

In the meantime, the fields of computer science and Artificial Intelligence (AI), along with Machine Learning (ML), are developing as a sign of novelty [3]. Utilizing past performance data, ML models predict upcoming tendencies, preventing the need for human intervention or explicit programming. Essentially, ML exemplifies the substance of pattern identification, with the capability to obtain insights from vast amounts of data with unparalleled efficacy. Denoising of retinal images employs innovative image processing methods that can be a main component of the approach. This stage is vital because the quality of the image directly affects the diagnostic outcomes. Following this initial stage, complex neural networks (NNs) and ML techniques are applied for classifying the strong images [4]. These techniques must have the capability to identify small patterns and variances in retinal imagery, which can often be overlooked by conventional diagnostic methods. This can be denoted as the earlier stages of diabetic retinopathy (DR). In the ML framework, deep learning (DL) directs the central phase, boosted by an increase in computational ability and the development of "big data." Convolutional Neural Networks (CNNs) exemplify this progression, transforming DL models with their capacity to extract features, classify images, and recognize patterns at breakneck speeds [5]. In comparison to the complex mechanisms of the human brain in executing filters and intricate layers, CNNs represent an innovative stage of computational efficacy.

To improve the quality of retinal images through effective denoising techniques, this paper proposes a Hybrid Deep Denoising and Dual-Branch Confidence-Aware Classification framework for accurate Diabetic Retinopathy (HDD-DBCADR) detection from retinal images. The proposed HDD-DBCADR framework offers a unified solution for diabetic retinopathy detection by combining pre-processing, hybrid denoising, dual-branch feature extraction, and confidence-aware classification. The overall framework demonstrates superior performance and robustness when evaluated on benchmark retinal image datasets. The main contributions of this paper are:

- A two-stage denoising mechanism is employed, involving noise simulation (Gaussian and speckle) to form a mixed noise model, followed by a hybrid deep denoising network using a residual convolutional neural network (DnCNN-style) with an autoencoder (AE) refinement module.
- Feature extraction is performed using a dual-branch architecture that combines a multi-scale convolutional neural network (local features) and a dual attention vision transformer (DaViT) for global features.

- A confidence-aware adaptive classification strategy is used to dynamically weight predictions from both branches.
- Extensive evaluation on Kaggle DR Datasets, which include five class labels representing increasing levels of retinal damage caused by diabetes, demonstrates improved performance compared to existing methods.

2. Literature Review

This section reviews existing studies on DR detection from denoised retinal images, focusing on recent methods and their effectiveness. Alanazi et al. [6] presented an advanced model named Explainable MINet-ViT that integrates cutting-edge noise reduction methods with an interpretable DL method for more dependable DR identification. This method integrates a multi-level denoising network (MINet), adapted by a noise-specific pre-processing technique employing deep residual feature mapping and variance-stabilizing transform (VST). A hybrid DL method that incorporates CNNs with Vision Transformers (ViT) has been utilized for extracting global and local spatial data. The explainability approaches, namely SHAP and Grad-CAM, ensure medical interpretability by recognizing the vital retinal regions that impact the model forecasts. Singh and Singh [7] inspected the significant complexity of DR, a major complication of diabetes that arises from damaging the small blood vessels of the retina. Denoising retinal images is the initial step in the methodology that can lead to the classification of the images by employing innovative methods. The effectiveness of this technique in the accurate identification of DR highlights its ability as a useful tool for earlier identification and treatment. Chilukuri et al. [8] implemented a Quantum Denoising AE (QDAE), a hybrid quantum–classical framework that employs convolutional feature coding with parameterized quantum circuits (PQCs) in hidden space.

Rajitha et al. [9] presented an innovative hybrid DL architecture, the GAI-Net, that employs GANs, AE, and Inception-V3 to identify DR in five stages of severity. This method addresses numerous problems in DR identification, such as class imbalance and smaller datasets, by employing quality synthetic images and discriminatory hierarchical features. Li et al. [10] examined a method for accurate identification and segmentation of HRF in OCT images of DR patients. This technique is crucial for supporting ophthalmologists in the early identification and evaluation of the efficiency of diagnoses and treatments. The HRF segmentation method, which depends on KiU-Net, encompasses the Kite-Net branch by employing up-sampling coding to gather more comprehensive data and a three-layer U-Net type for extracting higher-level semantic data. To improve the capability of a single-branch network, the cross-attention block (CAB) integrates feature extraction using dual divisions. Chen [11] presented the EDDiff, a new denoising diffusion model for classifying DR. The Enhanced Dual - Granularity Prior Module (EDPM) is designed to intensify significant features and capture the intricate key - region dependency in fundus images.

In [12], the main objective is to develop an automatic method for DR classification into its different phases by employing CNN methods. The effectiveness of fifteen pre-trained models is evaluated using the DRNet13 architecture. The most effective model for precise DR stage identification relies on fundus images categorized into 5 DR categories. Kirubakaran and Vijayarajan [13] presented the WaveMem-SHAPNet, an explainable DL method that integrates innovative explainable AI elements, memory-aware optimization, and preprocessing. This dual

method enhances fundus images by employing CLAHE, MF, and Adaptive Wavelet Denoising Transform (AWDT). Table 1 presents prior research on DR denoising and detection systems.

Table 1. Existing Methods for DR Denoising and Detection

References	Objectives	Models	Datasets	Results	Limitations
Alanazi et al. [6]	Develop a model capable of detecting diabetic retinopathy at early stages to prevent vision loss.	CNN and ViTs models.	APTOS-2019 and EyePACS datasets.	Accuracy of 97.6%.	May underperform on images from devices or populations not represented in the training data.
Singh and Singh [7]	Reduce noise in retinal images to enhance lesion visibility.	ML and neural networks.	APTOS 2019 dataset.	Accuracy of 94.3%.	Neural network outputs may be difficult for clinicians to interpret without additional explainability tools.
Chilukuri et al. [8]	Use convolutional feature coding combined with quantum circuits for enhanced hidden space representations.	Quantum computing and DL methods.	Diabetic Retinopathy 224x224 (2019) dataset.	SSIM of 0.96.	Integrating quantum circuits with classical CNNs requires advanced technical expertise.
Rajitha et al. [9]	Provide a fully automated, scalable solution suitable for large-scale clinical deployment.	GAN method	Diabetic Retinopathy 224x224 (2019) dataset.	Accuracy of 0.99.	Deployment may require high-end GPUs, limiting use in resource-constrained clinics.
Li et al. [10]	Assist ophthalmologists in early detection of DR through improved HRF visualization.	U-net algorithm	OCT images dataset.	DSC of 72.90%.	The model focuses only on OCT images and does not incorporate fundus images or other clinical data.
Chen [11]	Model spatial dependencies between important regions of the retina for better predictions.	CNN and diffusion approaches	APTOS 2019 dataset.	Achieves better DR detection.	High performance on benchmark datasets may not fully translate to unseen clinical data.
Shamrat et al. [12]	Compare performance with fifteen pre-trained CNN models to identify the most efficient.	CNN technique	Diabetic Retinopathy 224x224 (2019) dataset.	Accuracy of 97 %.	While efficient, real-time deployment on low-resource systems may still be challenging.
Kirubakaran and Vijayarajan [13]	Accurately detect and classify diabetic retinopathy severity from fundus images.	CLAHE and AWDT algorithms.	APTOS, EyePACS and IDRiD Models.	Accuracy of 98.74%.	Even with multiple datasets, demographic or device biases may affect generalizability.

3. Proposed Methodology

In this work, an HDD-DBCADR framework has been developed for accurate DR detection, integrating denoising and classification into a unified framework to enhance reliability in clinical screening. Fig. 1 represents the entire workflow of the HDD-DBCADR system. As depicted in figure, the process initiates with resizing the images to 224*224 pixels,

normalizing them to a range of 0-1, and black borders. Data augmentation (rotation, flipping, and scaling) is executed to standardize the images. Afterwards, a two-stage denoising approach is introduced, combining noise simulation (Gaussian and speckle noise) with a hybrid deep denoising network that integrates a DnCNN-style and AE refinement module. After denoising, feature extraction is performed using a dual-branch architecture combined with local features (multi-scale CNN) and global features (DaViT) to capture complementary information. Finally, a confidence-aware classification mechanism adaptively fuses the outputs of both branches based on its prediction confidence to improve overall classification reliability.

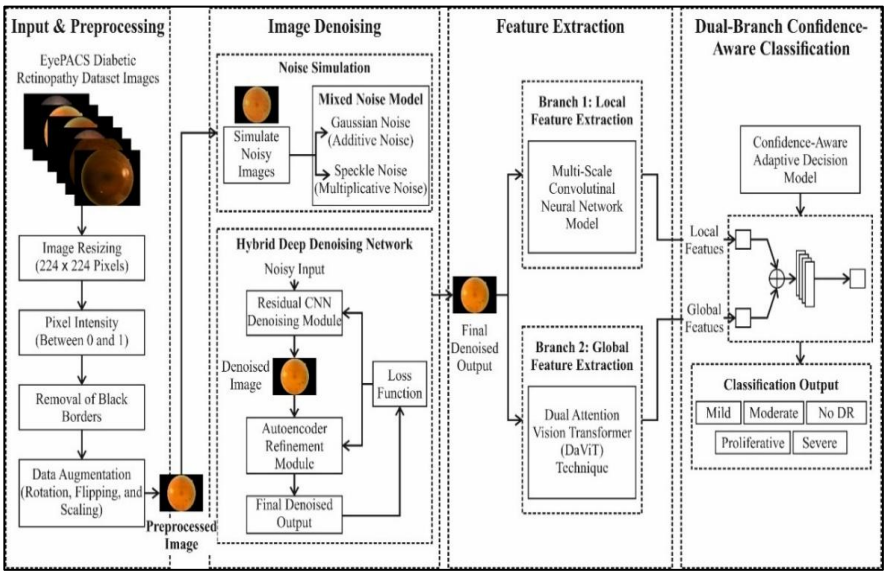

Figure 1. Overall Flow of HDD-DBCADR Framework

3.1 Preprocessing Pipelines

Before training the proposed model, numerous preprocessing steps are used for standardizing the dataset and enhancing model generality. The preprocessing steps comprise the following: Image resizing, pixel intensity normalization, removal of black borders, and data augmentation.

Resize: This technique resizes the image to align with the requirements of the DL approach. This procedure guarantees a stable image size, as DL techniques need a uniform input size. In this paper, the images are resized to 224x224 pixels. Moreover, resizing reduces high memory consumption and accelerates model convergence. This resolution provides a balanced trade-off between computational efficiency and preservation of significant retinal lesion information, allowing efficient feature learning while decreasing memory consumption and training complexity.

Pixel intensity normalization: This procedure regulates image pixels within the interval of zero and one by transforming the original pixels to be compatible with DL methods. Additionally, it enhances model training, as values in [0,1] enable gradient descent to work more efficiently. Eq. (1) demonstrates the normalization equation.

$$X_{normalized} = \frac{X - X_{min}}{X_{max} - X_{min}} \quad (1)$$

Black border removal: Black borders normally exist in fundus images owing to circular cropping during imaging. These borders do not comprise beneficial information and may bias

the model if they remain unprocessed. In this research, black borders are eliminated by identifying the non-zero-pixel regions and cropping the images accordingly. This process guarantees that only important retinal regions are utilized for training, which improves model precision. Moreover, removing black borders minimizes unwanted computation and speeds up the training procedure.

Data Augmentation: This step enhances data diversity by modifying the image using simple transformations to allow the model to learn from extensive instances, minimize overfitting, and increase the generalizability of models. In this paper, rotation, flipping, and scaling are used as augmentation techniques. Fig. 2 shows the sample images from data augmentation.

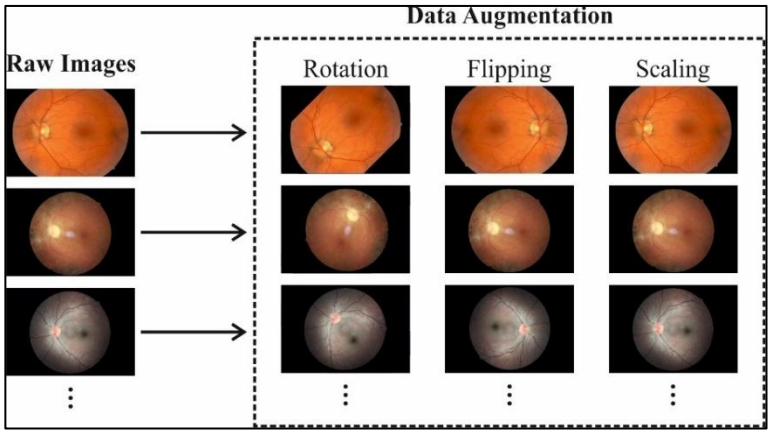

Figure 2. Sample of Augmented Images

3.2 Image Denoising

Retinal fundus images frequently contain noise introduced during image acquisition. Such noise can obscure small pathological attributes that are critical for identifying the early phases of DR. Therefore, a hybrid deep denoising framework was proposed to enhance image quality before feature extraction. The denoising process consists of two stages:

1. Noise simulation
    - Gaussian noise
    - Speckle noise
2. Hybrid deep denoising network
    - Residual CNN with an Autoencoder refinement module

3.2.1 Noise Simulation

The synthetic noise is mainly introduced during training to simulate real-world degradations usually observed in clinical fundus imaging, such as sensor noise, illumination differences, and acquisition artifacts. Though the original dataset images are relatively clean, a real clinical environment frequently comprises these uncontrolled distortions. By training the denoising model on Gaussian and speckle noise perturbations, the network learns robust feature restoration and generalizable noise-removal abilities without overfitting to the same situations. Since the dataset involves no clean images with their respective noisy counterparts, synthetic

noise is presented in the training to simulate real image conditions. Two varieties of noise are used:

1. Gaussian noise
2. Speckle noise

These noise models exemplify common distortions in retinal image systems.

Gaussian Noise: Gaussian noise simulates sensor noise created by the imaging device. It can be modeled as an additive noise distribution [14].

$$\mathrm{I_g} = \mathrm{I} + \mathrm{N}(0, \sigma^2) \tag{2}$$

Whereas I refers to the original retinal imaging. N denotes the Gaussian noise distribution. σ demonstrates the noise variance. Gaussian noise affects pixel intensity values and is usually observed in digital image systems.

Speckle Noise: The retinal images are produced by the OCT model. Because of interference, signals are influenced by the backscattered light of biotic tissues, and the actual OCT imagery generally represents the important speckle noise [15]. In order to estimate the strength of the speckle noise, a model was primarily created that precisely imitates the speckle noise in retinal imagery. The mathematical form of the multiplicative speckle noise method is denoted by

$$F(x_t y) = g(x, y) + g(\mathrm{x}, y) \times u(\mathrm{x}, y) \tag{3}$$

Here, the mathematical term $g(x, y)$ denotes the input imagery, $u(x, y)$ refers to Gaussian noise with a variance s and mean of zero that cannot be interrelated to the grayscale values of the actual images. The greater the variance of the noise, the greater the grayscale value. The computation form $F(x, y)$ signifies the noisy image.

Mixed Noise Model: To simulate realistic retinal imaging degradation, both Gaussian and speckle noise can be integrated into a mixed noise model. The Gaussian noise reduction helps to suppress random intensity differences introduced during image acquisition and sensor processing, while speckle noise management enhances the removal of multiplicative granular noise that might affect fine retinal structures. The integration of both methods improves image quality and preserves significant lesion-related characteristics. This design is derived from the separate features of additive Gaussian and multiplicative speckle noises. Gaussian noise methods address sensor-based disturbances and are formulated as:

$$I_g = I + N_g \tag{4}$$

whereas I stands for the original image and $N_g \sim N(0, \sigma^2)$ demonstrates the additive Gaussian noise.

Speckle noise has generally observed in OCT retinal imagery because of backscattered light interference, and is demonstrated as multiplicative noise:

$$I_s = g(x, y) + g(x, y) \times u(x, y) \tag{5}$$

whereas $g(x, y)$ defines the input imaging and $u(x, y))$ is zero-mean noise with variance σ^2.

By integrating these 2 noise methods, the mixed noise formulation is attained as:

$$I_{mixed} = \left(I + N_g\right) + \left(I + N_g\right) \times N_s \tag{6}$$

In which, N_s defines the speckle noise. During this formulation, Gaussian noise is 1st added to the clean imaging, and the outcome image is more degraded by multiplicative speckle noise.

3.2.2 Hybrid Deep Denoising Network

To efficiently eliminate mixed noise patterns while maintaining retinal structures, a hybrid deep denoising network has been presented. The model integrates a DnCNN-style architecture with an AE refinement module. The network learns a mapping function:

$$F\left(I_{noisy}\right) = I_{denoised} \tag{7}$$

Whereas, I_{noisy} stands for the noisy retinal image and $I_{denoised}$ refers to the reconstructed clean imaging.

As convolutional (Conv) layers are added to a CNN, the network correspondingly multiplies by more neurons [16]. The deep residual CNN approach supposes that network with insufficient layers can achieve the desired state and then increases the outcome of the mapping layer. The residual CNN denoising module operates by learning the residual noise distribution present in the retinal fundus images, rather than directly reconstructing the denoised image. Traditional layers are utilized to remove hierarchical noise-related features, while residual skip connections preserve significant retinal structures and fine lesion details during the denoising stage. The module removes unwanted noise components while maintaining clinically relevant features, thus enhancing image quality. However, x implies the input of the existing unit, and $F(x)$ depicts the mapping outcome of the existing unit guided by a non-linear transformation purpose.

$$H(x) = F(x) + x \tag{8}$$

Compared to conventional CNNs, residual CNNs generate shortcut connections that instantly connect earlier layers to later ones. These residual blocks help to evade the vanishing gradient concern in very deep networks. Unlike shallow conventional CNNs, residual CNNs can have hundreds of layers for enhanced representation. It concentrates on learning residual information, which decreases training time and enhances convergence.

Denoising CNN **(**DnCNN) signifies a deep CNN particularly intended for image denoising tasks [17]. In Gaussian noise situations, DnCNN exhibits higher denoising performance.

An autoencoder (AE) is a feedforward neural network (FFNN) in which the output layer has the same dimensionality as the input layer. This variety of neural networks is intended to work unsupervised, as they are provided with training input vectors to reconstruct the resultant vectors. The AE operates on input vectors specified by an encoder to create a latent representation H.

$$\mathrm{H} = \partial(\mathrm{W_{xh}X} + \beta_{\mathrm{xh}}) \tag{9}$$

Now, W implies a weight matrix, β denotes biased vector and ∂ refers to any activation function, specifically a sigmoid function or ReLU. An input space can be used to transform the hidden depiction vector H utilizing the decoder.

$$\widehat{\mathrm{X}} = \partial(\mathrm{W}_{\hat{\mathrm{x}}\mathrm{h}}\mathrm{h} + \beta_{\hat{\mathrm{x}}\mathrm{h}}) \tag{10}$$

The reconstruction error ε is evaluated to discover the variance amongst the reconstructed vectors $\widehat{\mathrm{X}}$ and actual input vector.

$$\varepsilon = \|\mathrm{X} - \widehat{\mathrm{X}}\| \tag{11}$$

The AE undergoes training in an unsupervised manner to diminish RE and acquire relationships among input features. Inputs with high RE are specified as consistent data, as the trained AE could precisely reconstruct normal data but not inconsistent data.

The final denoised image is:

$$I_{final} = I_{refined} \tag{12}$$

This refinement stage improves vessel visibility and lesion contrast.

3.3 Feature Extraction

After denoising, a dual-branch feature extraction framework is developed, where local features are captured using a multi-scale CNN and global dependencies are modeled using DaViT.

3.3.1 Local Feature Extraction Using Multi-Scale CNN

Feature extraction that depends on CNNs only employs the features of the last layer [18]. Accordingly, numerous comprehensive parts of data in the interlayers are reduced over these features. The MS Conv layer is employed to maintain global and local data to improve the network's ability. The MS layer integrates numerous Conv layers of various sizes in a similar Conv layer. The MS Conv mapping feature is integrated and then entered into the subsequent pooling layer for subsampling, executed to the FC layer; lastly, the outcomes are acquired by the resultant layer using softmax.

Later, in the Conv of kernels and input images, the local data is developed in a Conv layer, and after a non-linear activation function has been applied, a 3D tensor, comprising a stack of matrices named the mapping feature, is produced as the result of the Conv layer. The mathematical form of the resultant mapping feature from the Conv layer can be represented below,

$$Y_j^T = \varphi\left(\sum_{i=0}^{n} \omega_{ij}^T * Y_j^{T-1} + b_j^T\right), \tag{13}$$

Here, the symbol $*$ operator is employed for the $2-D$ Conv of channels.

During the Conv process, Y_j^{T-1} refers to j^{th} input tensor of layer T-l, and Y_j^T represents the j^{th} resultant layer T, b_j^T indicates the weighted of Conv bias, whereas the parameter ω_{ij}^T refer to the weighted of Conv kernel, the symbol φ remains a non-linear activation function, and the last outcome must be accomplished. In some instances, any comprehensive lower-level

data could be unnoticed. The mixed layer is supported for learning the lower-level features and higher-level features, as well as the aim of signifying the image features with some neurons. The outcome of the combined layer must be represented by,

$$Y = \varphi\left(\sum_{i=0}^{n_1} \omega_{ij}^1 * x_i^1 + \sum_{i=0}^{n_2} \omega_{ij}^2 * x_i^2 + \sum_{i=0}^{n_3} \omega_{ij}^3 * x_i^3 + \sum_{i=0}^{n_4} \omega_{ij}^4 * x_i^4 + b_j\right), \quad (14)$$

Here, the mathematical parameter $x_i^1, \omega_{ij}^1, x_i^2, \omega_{ij}^2, x_i^3, \omega_{ij}^3$, and x_i^4, ω_{ij}^4 signify the neurons and weighted in kernels of multi-scale Conv layer, while the mathematical term n_i indicates the n filters in $Conv2\_i$. Y, to be the outcome of mixed layer.

3.3.2 Global Feature Extraction using DaViT

The DaViT is a clear, effective backbone comprising the local fine-grained features and global representations [19]. Primarily, it presents the hierarchical layout of the architecture. The model is separated into four phases, where a patch embedding layer can be embedded in the primary phases. There is a two-attention module in each individual phase with the resolution and feature size maintained as similar. Consider the mathematical term of $\mathbb{R}^{P\times C}$ dimension visual feature, here P refers to the count of total patches, and C represents the count of overall channels.

$$\mathcal{A}(Q,K,V) = Concat(head_1, \ldots, head)$$

$$whereas\ head_i = Attn(Q_i, K_i, V_i)$$

$$= softmax\left[\frac{Q_i(K_i)^T}{\sqrt{C_h}}\right]V_i \quad (15)$$

Here, the mathematical form of $Q_i = X_i W_i^Q$, $K_i = X_i W_i^K$, and $V_i = X_i W_i^y$ are $\mathbb{R}^{P\times C_h}$ dimensional features with N_h heads, the term X_i signifies the i_{th} head of input data and the mathematical parameter W_i specifies the projected weighted of i^{th} head for Q, K, V, and $C = C_h * N_h$. Especially the output projection W^O can be rejected. Assumed that the P is the biggest one, for example, 128×128, the computational expenses are excessive.

In DaViT, the arrangement of channel group attention and spatial window attention can be used to acquire both global and local features, followed by the linear complexities in the spatial dimension. Fig. 3 illustrates the architecture of DaViT.

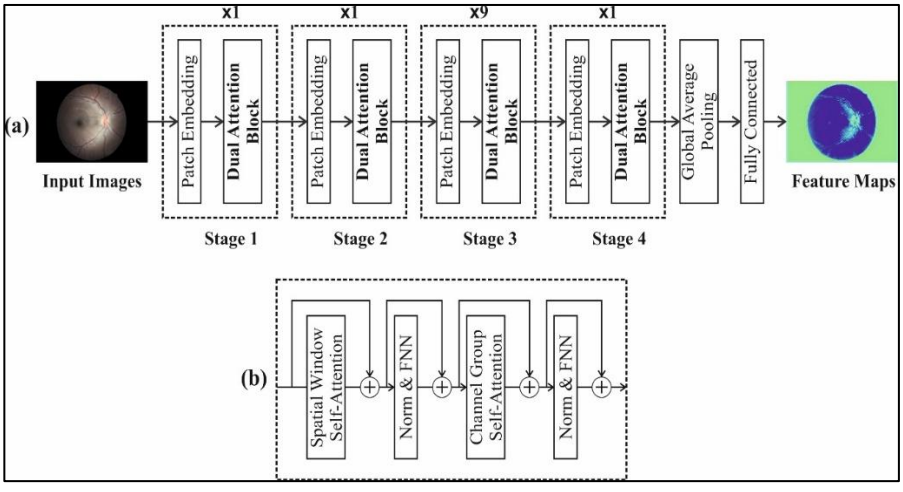

Figure 3. Architecture of DaViT

Spatial Window Attention: The window attention is calculated using self-attention in the local windows. The windows have been organized to separate the images into non-overlapping groups. Consider the mathematical term N_w for various windows, each window comprising P_w patches, here the computational term $P = P_w * N_w$. Subsequently, the window attention is denoted by,

$$\mathcal{A}_{window}(Q, K, V) = \{\mathcal{A}(Q_i, K_i, V_i)\}_{i=0}^{N_w} \tag{16}$$

Here, the mathematical form $Q_i, K_i, V_i \in \mathbb{R}^{P_w \times C_h}$ describes the local window values, keys, and queries. By using this computation, the window attention drops the capability to model global data. The channel attention certainly addresses this complexity and commonly advantages window attention.

Channel Group Attention: The self-attention originates from an alternative viewpoint and develops the channel-wise attention. Prior self-attention in vision defines tokens with patches or pixels and collects the data with spatial dimensions. Rather than executing attention on the patch- or pixel-level, the attention mechanisms (AM) operate on transfer of patch-level tokens. In order to acquire the global data from the spatial size, the count of heads is equivalent to 1. During this method, the channel tokens interact with global data under the channel size from the linear spatial-wise intricacy. It can be represented by,

$$\mathcal{A}_{channel}(Q, K, V) = \{\mathcal{A}_{group}(Q_i, K_i, V_i)^T\}_{i=0}^{N_g}$$

$$\mathcal{A}_{group}(Q_i, K_i, V_i) = soft\ \max\left[\frac{Q_i^T K_i}{\sqrt{C_g}}\right] V_i^T \tag{17}$$

Here the parameters $K_i, Q_i, V_i \in \mathbb{R}^{P \times C_g}$ are combined channel wise image level keys, query, and value. However, the transfer of token in channel attention, projection layer W, and scaling parameter $\frac{1}{\sqrt{C_g}}$ is accomplished and calculated with the channel size, instead of spatial one.

3.4 Dual-Branch Confidence-Aware Classification

In traditional DL methods, feature extraction from numerous branches is generally integrated by employing operations like averaging or concatenation. Although these direct fusion approaches handle entire feature sources equivalently, they disregard the dependability and discriminatory improvement of every individual branch for a specified input. The dual-branch confidence-aware classification is mainly used to manage low-confidence predictions produced by both the CNN-based local and DaViT-based global feature extraction branches. The confidence scores attained from both branches are evaluated during feature fusion, and adaptive weighting is used to emphasize the more reliable estimation. In cases where both branches deliver low-confidence results, the fusion module integrates the complementary local and global feature representations and chooses the final class label driven the highest aggregated probability score. In medical image evaluation, particularly for DR detection, the significance of local lesion features and global contextual data can vary across images. Thus, a more adaptable and consistent decision-making method is essential. To overcome this limitation, a dual-branch confidence-aware classification approach has been developed. Rather than combining the features at an earlier phase, each individual branch independently executes the classification and generates its own probability distribution using the DR classes. A

confidence-based weighting approach has been employed to dynamically integrate the predictions, thus permitting the model to depend on the branch that is most confident for a specific input image.

Independent Branch Predictions: Later, after the feature extraction, the branches are transformed into relevant feature representations by employing the fully connected (FC) layer, and the softmax activation is applied to produce the class probability distributions.

Local Branch Output: $P_{local} = [p_0, p_1, p_2, p_3, p_4]$

Global Branch Output: $P_{global} = [q_0, q_1, q_2, q_3, q_4]$

Every vector signifies the predictable probabilities for the 5 DR severity classes (0 to 4). The mathematical form of probabilities is given by,

$$\sum_{i=0}^{4} p_i = 1, \sum_{i=0}^{4} q_i = 1 \tag{18}$$

The local branch concentrates on fine-grained lesion features, whereas the global branch captures the whole retinal structure and contextual correlation.

Confidence Estimation: In order to calculate the reliability of each predictive branch, a confidence score can be computed that depends on the maximum predicted probability. The insight is that the maximal probability denotes robust confidence in a specific class prediction.

$$C_{local} = max(P_{local}) \tag{19}$$

$$C_{global} = max\left(P_{global}\right) \tag{20}$$

Here the parameter C_{global} refers to the confidence of the global branch, and the term C_{local} indicates the confidence of the local branch.

Adaptive Weight Computation: The confidence scores have been standardized to calculate adaptive weights for each individual branch. Such weights are determined by the involvement of each branch in the ultimate decision.

$$w_{global} = \frac{C_{local}}{C_{local} + C_{global}} \tag{21}$$

The weights satisfy the restriction,

$$w_{local} + w_{global} = 1 \tag{22}$$

This normalization assures a stable and explainable integration. When one branch has a significantly greater confidence, it can obtain a proportionally greater weight.

Final Prediction Fusion: It must be acquired through the calculation of a weighted integration of the probability distributions of these two branches,

$$P_{final} = w_{local} \cdot P_{local} + w_{global} \cdot P_{global} \tag{23}$$

The projected DR category can be denoted by choosing the class with the maximum probability,

$$y = arg\ max\left(P_{final}\right) \tag{24}$$

4. Experimental Validation

In this section, we present a detailed dataset description, denoising analysis, classification results, and an ablation study. In this study, the dataset was split into 70% for training and 30% for testing. The method runs on Python 3.6.5 with an i5-8600k CPU, a 4GB GPU, 16GB RAM, a 250GB SSD, and a 1TB HDD, using a learning rate of 0.01, ReLU, 50 epochs, a dropout rate of 0.5, and a batch size of 5. 4.1. Data Used

The empirical assessment of the proposed technique was conducted by employing the DR detection dataset [20]. The dataset is provided with higher-resolution retinal images obtained using different camera models and imaging settings, which may have impacted the visual orientation of the right and left eyes. All the images have been anatomically generated, with the optic nerve on the right and the macula on the left eye, while changes appear reversed, as observed using a microscope with a condensed lens in a general live eye analysis. An image is considered inverted when the macula, the small dark central region, is somewhat above the midline of the optic nerve; if the macula is positioned below this midline, it is not inverted. Likewise, the presence of a notch on the side of the images (circle, square, or triangle) indicates that it is not inverted, while images without a notch will be reversed. For every subject, both left and right regions are provided, and all images must be labeled with a subject ID and the respective eye orientation. Table 2 shows the number of samples in the DR detection dataset. Fig. 4 depicts the sample images of the original and the denoised versions. Fig. 5 represents the pre-processed sample images of the original and denoised versions. Fig. 6 shows the feature extraction of five classes.

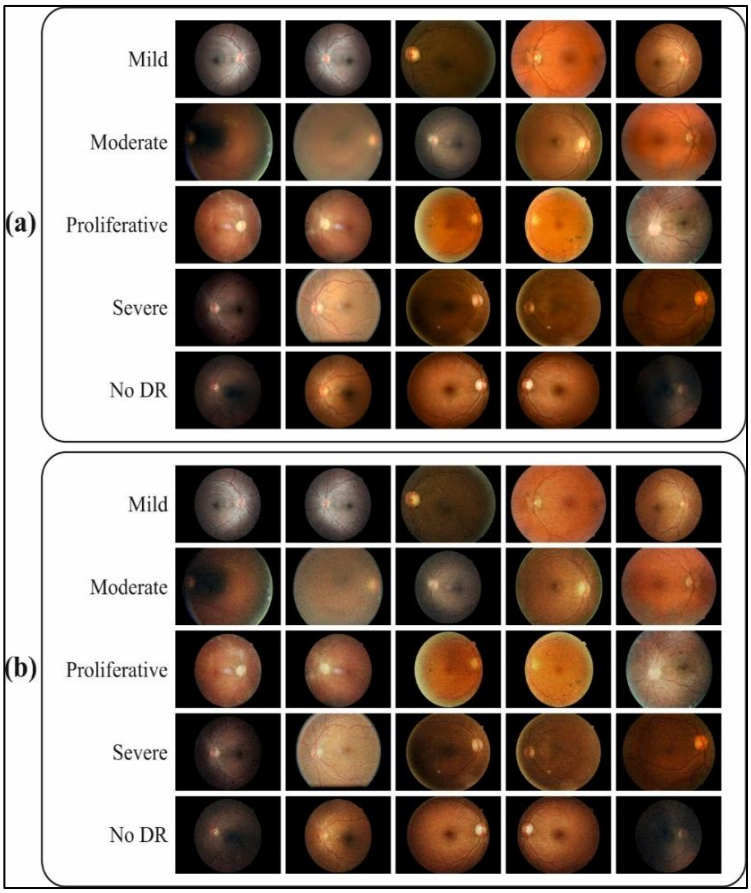

Figure 4. Sample of (a) Original and (b) Denoised Images

Table 2. Number of Samples of DR-Detection Data Set

Classes	No. of Samples
NoDR	5000
Mild	2000
Moderate	5000
Severe	800
Proliferative	700
Total Samples	13500

Fig. 7 displays intensity profiles with pixel positions for five different classes, all of which are compared with three lines: clean (original), noisy, and denoised signals. The noisy lines show asymmetrical variations influenced by noise, whereas the denoised lines closely follow the clean signals, effectively smoothing out noise and maintaining the true intensity patterns. This represents the denoising capability to decrease noise and improve the original signal shape reliably across each individual class, from mild to severe.

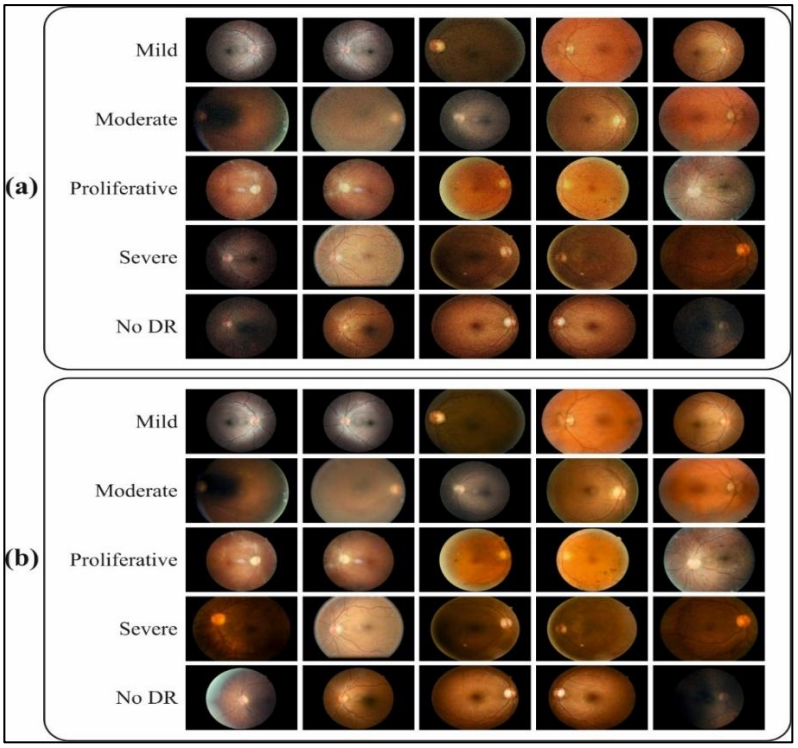

Figure 5. Preprocessing Sample of (a) Original and (b) Denoised

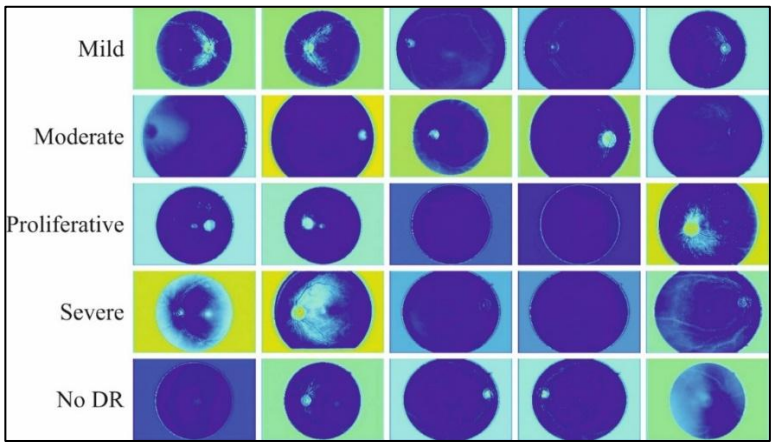

Figure 6. Feature Extraction of Five Classes

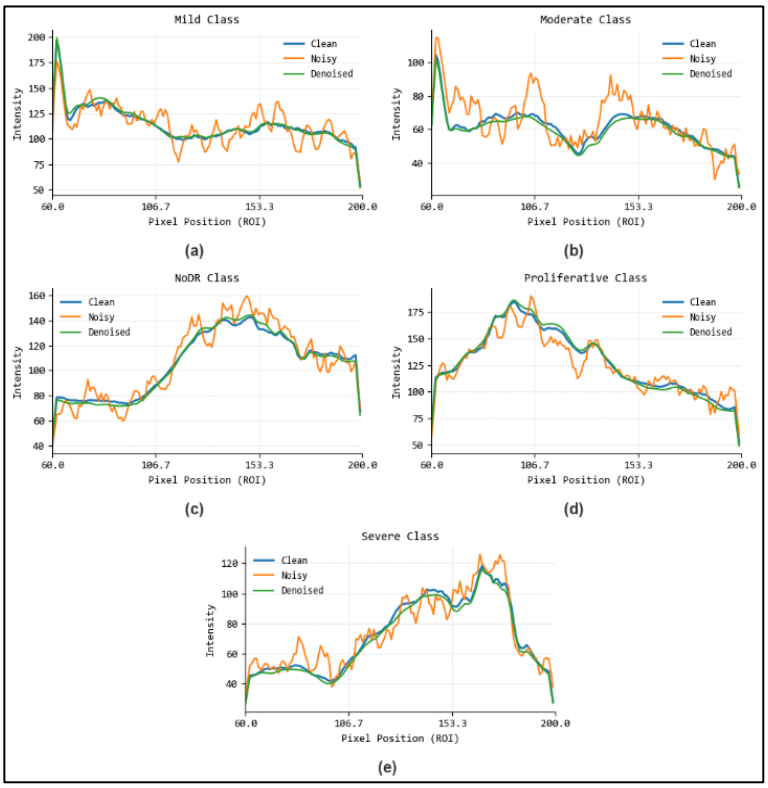

Figure 7. Different Classes with (a) Mild, (b) Moderate, (c)NoDR, (d) Proliferative, and (e) Severe

4.2 Performance Measures

Performance measures calculate the efficacy of a method in both prediction and classification studies through comparison of predicted outcomes with definite ground truth values. They support the computation of error, similarity, and accuracy to ensure robust and reliable method outcomes.

4.2.1 Classification Metrics

Classification metrics compute how well a system allocates distinct class labels. They quantify accuracy by applying terms such as TN, TP, FN, and FP to examine the solution.

Accuracy: It calculates the ratio of properly predicted instances out of the overall count of instances.

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \tag{25}$$

Precision: It calculates the ratio of appropriately predicted positive instances out of each predicted positive instances.

$$\text{Precision} = \frac{TP}{TP+FP} \tag{26}$$

Recall: Recall is the percentage of properly recognized positive instances out of every actual positive instances.

$$\text{Recall} = \frac{TP}{TP+FN} \tag{27}$$

F1-score: F1-score is the harmonic mean of precision and recall utilized for balanced performance evaluation.

$$F1 = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \tag{28}$$

ROC-AUC: It calculates the ability model to distinguish between classes through every threshold.

$$\text{TPR} = \frac{TP}{TP+FN} \tag{29}$$

$$\text{FPR} = \frac{FP}{FP+TN} \tag{30}$$

4.2.2 Prediction Metrics

Prediction metrics calculate how close continuous or reconstructed outcomes are to the ground truth values. They measure numerical error or visual similarity among predicted and actual outputs.

MSE: MSE is the average of squared differences among predicted and actual values.

$$\text{MSE} = \frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 \tag{31}$$

PSNR: PSNR is the ratio of maximal signal power to noise power utilized to measure image quality.

$$\text{PSNR} = 10 \log_{10} \left(\frac{MAX^2}{\text{MSE}} \right) \tag{32}$$

SSIM: SSIM is the calculation of structural similarity among predicted and reference images.

$$\text{SSIM}(x, y) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)} \tag{33}$$

4.3 Result Analysis of Image Denoising Approach

Table 3 and Fig. 8 present the denoising analysis of the HDD-DBCADR approach, including training and testing phases using various measures. With the training set, the HDD-DBCADR model has attained an average MSE of 0.000130, PSNR of 38.92869, and SSIM of 0.976031. Meanwhile, with the testing set, the HDD-DBCADR method has obtained average value with MSE of 0.000132, PSNR of 38.89387, and SSIM of 0.975867, respectively.

Table 3. Denoising Analysis of the HDD-DBCADR Method with Training and Testing Data

Metrics	Training Data	Testing Data
MSE	0.000130	0.000132
PSNR	38.92869	38.89387
SSIM	0.976031	0.975867

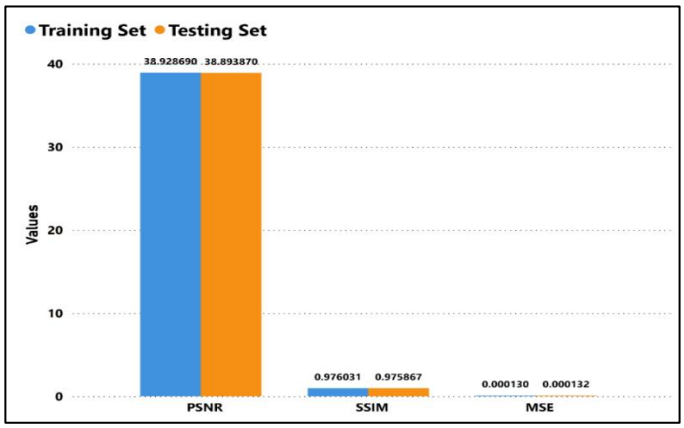

Figure 8. Denoising Analysis Using the Average Value of the HDD-DBCADR Method

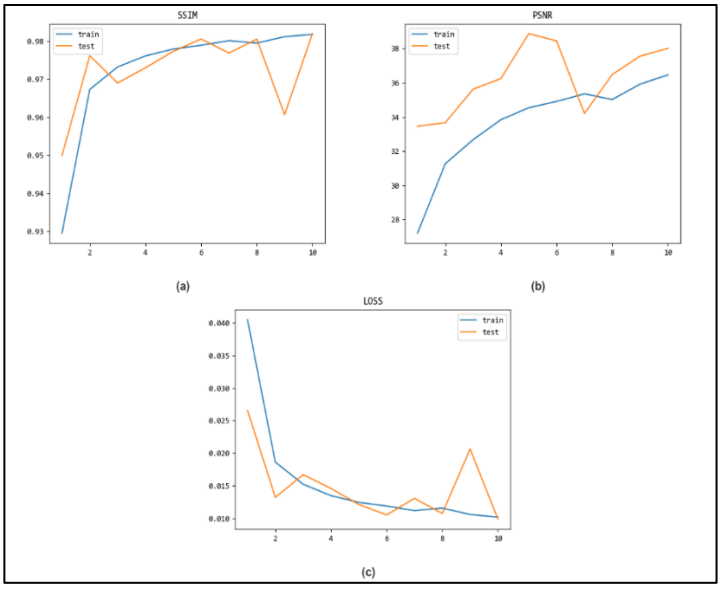

Figure 9. Training and Testing Performance of the HDD-DBCADR Model with (a) SSIM, (b) PSNR, and (c) Loss

Fig. 9 demonstrates the training and testing performance of a model at 10 epochs by employing three metrics: (a) SSIM, (b) PSNR, and (c) Loss. The SSIM and PSNR values rise with training, denoting the improved image quality, with test performance similar to and also exceeding the training with respect to time. The loss progressively decreases for both sets, thus exhibiting proficient learning and higher predictive accuracy.

Table 4. Comparison Analysis of the HDD-DBCADR with Other System for Denoising Analysis

Methods	MSE	PSNR	SSIM
OPA [21]	0.00121	20.7466	0.8401
Bi-LSTM [21]	0.00152	25.3096	0.8346
CNN [21]	0.00487	26.0574	0.7023
ISCL [21]	0.00354	29.6708	0.8495
EPLL [22]	0.00560	30.9600	0.7886
DDM [22]	0.00169	33.0200	0.8175
GCDN [22]	0.00035	28.9300	0.7951
HDD-DBCADR [Proposed]	0.00013	38.9287	0.9760

Table 4 shows a comparative examination of the HDD-DBCADR with other models for denoising analysis [21, 22]. The simulation results of the HDD-DBCADR method demonstrate superior performance. Based on MSE, the HDD-DBCADR method has a lower

value with an MSE of 0.00013, while the OPA, Bi-LSTM, CNN, ISCL, EPLL, DDM, and GCDN models have obtained higher values with MSEs of 0.00121, 0.00152, 0.00487, 0.00354, 0.00560, 0.00169, and 0.00035, respectively. Moreover, based on PSNR, the HDD-DBCADR approach has reached a greater value with a PSNR of 38.9287, while the OPA, Bi-LSTM, CNN, ISCL, EPLL, DDM, and GCDN models have achieved lesser values with PSNRs of 20.7466, 25.3096, 26.0574, 29.6708, 30.9600, 33.0200, and 28.9300, respectively. Lastly, based on SSIM, the HDD-DBCADR approach has attained the maximal value with an SSIM of 0.9760, while the OPA, Bi-LSTM, CNN, ISCL, EPLL, DDM, and GCDN models have gained minimal values with SSIMs of 0.8401, 0.8346, 0.7023, 0.8495, 0.7886, 0.8175, and 0.7951, respectively.

Table 5 depicts the computational performance of the HDD-DBCADR model compared with other techniques [23]. Based on inference time, the proposed method provides a lesser value of 3.67 ms, with a parameter count of 1.09 M, FLOPS of 0.26 G, and GPU memory of 38 M, whereas the ResNet50, DenseNet201, EfficientNet-B3, Hybrid EfficientNet + Swin, Enhanced ResNet50 + Attention, and EDSCDR-CNN techniques attain higher times of 12.80 ms, 15.20 ms, 10.40 ms, 18.60 ms, 14.30 ms, and 6.70 ms, respectively.

Table 5. Comparative Computational Results of HDD-DBCADR Model

Models	Parameters (M)	FLOPs (G)	Inference Time (ms)	GPU (M)
ResNet50	23.50	4.10	12.80	98
DenseNet201	20.00	4.40	15.20	102
EfficientNet-B3	12.00	1.80	10.40	74
Hybrid EfficientNet + Swin	45.20	6.30	18.60	165
Enhanced ResNet50 + Attention	28.60	5.10	14.30	112
EDSCDR-CNN	3.20	0.90	6.70	42
HDD-DBCADR	1.09	0.26	3.67	38

4.4 Result Analysis of Classification Model

Fig. 10 demonstrates the classifier outcomes of the HDD-DBCADR system. Figs. 10a-10b display the confusion matrix results with the identification of 5 class labels. Fig. 10c exemplifies the PR analysis, demonstrating maximal performance across all classes. Lastly, Fig. 10d exemplifies the ROC analysis, demonstrating higher-level outcomes with improved ROC results for diverse class labels.

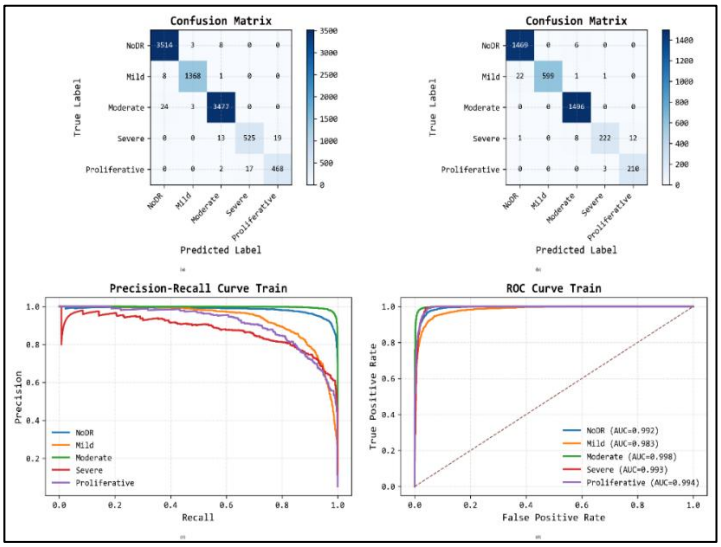

Figure 10. Classifier Result of the HDD-DBCADR Method with (a-b) Confusion Matrix and (b-c) PR/ROC Curve

The DR classification outcomes of the HDD-DBCADR method are clearly established in Table 6. The experimental outcome values demonstrate the effective ability of the HDD-DBCADR method in the recognition process. According to the training set, the HDD-DBCADR method achieves averages $accur_y$ of 98.96%, $preci_n$ of 98.96%, $recal_l$ of 98.96%, $F1_{score}$ of 98.96%, and ROC-AUC of 99.31%. Moreover, based on testing set, the HDD-DBCADR system gains average $accur_y$ of 98.67%, $preci_n$ of 98.68%, $recal_l$ of 98.67%, $F1_{score}$ of 98.66%, and ROC-AUC of 99.29%.

Table 6. DR Classifier Result of the HDD-DBCADR Approach with Training and Testing Data

Metric	Training Data	Testing Data
Accuracy	98.96	98.67
Precision	98.96	98.68
Recall	98.96	98.67
F1-Score	98.96	98.66
ROC-AUC	99.31	99.29

In Table 7, the overall comparison evaluation of the HDD-DBCADR is evidently described [24, 25]. The simulation outcomes showed that the HDD-DBCADR method achieves effective performance. Based on $accur_y$, the HDD-DBCADR system provides better $accur_y$ of 98.96% while the MAPCRCI-DMPLC, MVDRNet, MPDCNN, AlexNet, SUNet, pre-trained VGG19, and Triple-DRNet techniques realize lesser $accur_y$ of 94.28%, 88.71%, 86.81%, 96.85%, 81.55%, 95.40%, and 92.08%. Alternatively, with the $preci_n$, the HDD-DBCADR method gains the greater $preci_n$ of 98.96%, whereas the MAPCRCI-DMPLC, MVDRNet, MPDCNN, AlexNet, SUNet, Pre-trained VGG19, and Triple-DRNet techniques obtain lower $preci_n$ of 94.64%, 85.41%, 89.92%, 83.51%, 83.92%, 87.18%, and 86.79%, respectively.

Table 7. Comparison Outcome of HDD-DBCADR Approach with Other Systems

Methods	$Accur_y$	$Preci_n$	$Recal_l$	$F1_{Score}$
MAPCRCI-DMPLC [24]	94.28	94.64	91.73	80.51
MVDRNet [24]	88.71	85.41	93.77	86.19
MPDCNN [24]	86.81	89.92	81.78	80.78
AlexNet [25]	96.85	83.51	88.50	87.49
SUNet [25]	81.55	83.92	94.54	83.06
Pre-trained VGG19 [25]	95.40	87.18	83.67	92.13
Triple-DRNet [25]	92.08	86.79	92.42	86.15
HDD-DBCADR [Proposed]	98.96	98.96	98.96	98.96

4.5. Ablation Study

An ablation study is conducted for the systematic evaluation of the contribution of every individual element in the proposed architecture. Various models are assessed by selectively eliminating or modifying key modules, namely, preprocessing, feature extraction, and classification approaches. The performance comparison encompasses the implications of every component on overall accuracy and strength. The experimental outcome values demonstrate that the comprehensive model, integrating all modules, achieves greater performance compared to its simplified forms, thus ensuring the effectiveness of the proposed design. Table 8 presents the ablation study of the denoising model with various measures. The Only CNN approach has obtained lower performance with an MSE of 0.0021, PSNR of 34.1687, and SSIM of 0.6540. Next, the Residual Convolutional Neural Network (DnCNN-style) system has achieved a slightly higher result with an MSE of 0.0016, PSNR of 35.7687, and SSIM of 0.7640. In

addition, the Residual CNN + Autoencoder Neural Network refinement methodology has reached moderate outcomes with an MSE of 0.0009, PSNR of 37.2887, and SSIM of 0.8760. However, the HDD-DBCADR (hybrid model Residual CNN + Autoencoder) approach has achieved effective results with an MSE of 0.0001, PSNR of 38.9287, and SSIM of 0.9760.

Table 9 demonstrates the ablation study of the HDD-DBCADR system with various aspects. Based on $accur_y$, the HDD-DBCADR (Dual-branch + confidence-aware adaptive decision model) approach has attained higher $accur_y$ of 98.96% while Baseline (CNN), Multi-Scale CNN, Global (Multi-Scale CNN +DaViT), and Dual-Branch (Classification heads for local & global features) systems have lesser $accur_y$ of 96.19%, 96.83%, 97.60%, and 98.37%. Meanwhile, based on $recal_l$, the HDD-DBCADR (Dual-branch + confidence-aware adaptive decision model) method gains maximal $recal_l$ of 98.96%, whereas Baseline (CNN), Multi-Scale CNN, Global (Multi-Scale CNN +DaViT), and Dual-Branch (Classification heads for local & global features) techniques are lower $recal_l$ of 96.14%, 96.90%, 97.60%, and 98.30%. In conclusion, based on $F1_{score}$, the HDD-DBCADR (Dual-branch + confidence-aware adaptive decision model) system has gained higher $F1_{score}$ of 98.96%, although Baseline (CNN), Multi-Scale CNN, Global (Multi-Scale CNN +DaViT), and Dual-Branch (Classification heads for global & local features) models have smaller $F1_{score}$ of 96.06%, 96.72%, 97.48%, and, 98.21%.

Table 8. Ablation Study of Denoising Model with Various Measures

Methods	MSE	PSNR	SSIM
Only CNN	0.0021	34.1687	0.6540
Residual Convolutional Neural Network (DnCNN-style)	0.0016	35.7687	0.7640
Residual CNN + Autoencoder Neural Network refinement	0.0009	37.2887	0.8760
HDD-DBCADR (hybrid model Residual CNN + Autoencoder)	0.0001	38.9287	0.9760

Table 9. Ablation Study of the HDD-DBCADR System with Different Aspects

Methods	$Accur_y$	$Preci_n$	$Recal_l$	$F1_{Score}$
Baseline (CNN)	96.19	96.04	96.14	96.06
Multi-Scale CNN	96.83	96.67	96.90	96.72
Global (Multi-Scale CNN +DaViT)	97.60	97.45	97.60	97.48
Dual-Branch (Classification heads for local & global features)	98.37	98.17	98.30	98.21
HDD-DBCADR (Dual-branch + confidence-aware adaptive decision model)	98.96	98.96	98.96	98.96

5. Conclusion

In this manuscript, we have presented a novel HDD-DBCADR framework for retinal images, which integrates hybrid denoising with confidence-aware dual-branch classification, offering an encouraging approach for developing reliable and clinically applicable automated DR screening systems. The proposed system begins with effective image pre-processing, including resizing images, normalizing pixels, removing black borders, and executing data augmentation techniques to enhance generalization. The denoising process consists of two phases: noise simulation comprising Gaussian and speckle noise, followed by a hybrid deep denoising network that combines a DnCNN-style network with an AE-based refinement module to improve image quality. For feature representation modules, a dual-branch architecture has been utilized. The local branch utilizes a multi-scale CNN to capture fine-grained spatial features, and the global branch leverages a DaViT to model long-range dependencies and contextual data. Finally, a dual-branch confidence-aware classification

mechanism adaptively combines predictions from both branches, depending on their confidence scores, thereby improving classification reliability. The simulation examination of the proposed HDD-DBCADR framework is performed using a benchmark DR detection dataset from the Kaggle repository. Extensive comparative outcomes report the superior performance of the HDD-DBCADR algorithm over recent systems.

Data Availability Statement

The data that support the findings of this study are openly available in the Kaggle repository at https://www.kaggle.com/competitions/diabetic-retinopathy-detection/data, reference number [20].

Acknowledgment

Not Applicable

Declarations

Conflict of Interest

The authors declare no conflict of interest.

Funding Sources

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Ethics Approval of Research

Not Applicable.

References

[1] Liang, Weizhe, Chee-Onn Chow, Raymond Wong Jee Keen, and Jeevan Kanesan. "Diabetic Retinopathy Classification Network with Multi-Frequency Contextual Attention Module." Medical Engineering & Physics 147, no. 3 (2026): 035008.

[2] Aftab, Shabib, and Samia Akhtar. "Diabetic Retinopathy Severity Classification Using Data Fusion and Ensemble Transfer Learning." Journal of Software Engineering and Applications 18, no. 1 (2025): 1-23.

[3] Radha, K., and Yepuganti Karuna. "Retinal Vessel Segmentation to Diagnose Diabetic Retinopathy Using Fundus Images: A Survey." International Journal of Imaging Systems and Technology 34, no. 1 (2024): e22945.

[4] Krishnadhas, Anugirba, Lal Raja Singh Ravi Singh, and RS Rimal Isaac. "D-RetinoNet: Diabetic retinopathy Stage Classification via Deep Duo-Branch S2 Feature Based Neural Network." Biomedical Signal Processing and Control 119 (2026): 109901.

[5] Devi, T. M., P. Karthikeyan, B. Muthu Kumar, and M. Manikandakumar. "Diabetic Retinopathy Detection via Deep Learning Based Dual Features Integrated Classification Model." Technology and Health Care 33, no. 2 (2025): 1066-1080.

[6] Alanazi, Sultan, Sajid Ullah Khan, Faisal M. Alotaibi, and Mohammed Alonazi. "A Novel Noise Removal and Interpretable Deep Learning Model for Diabetic Retinopathy Detection." International Journal of Imaging Systems and Technology 35, no. 6 (2025): e70245.

[7] Singh, Atulesh Pratap, and Ajay Singh. "Denoising and Diabetic Retinopathy Detection using Machine Learning and Neural Network." health 2025. 1: 2.

[8] Chilukuri, Rajitha, Praveen P, Ranjith Kumar Gatla, and Reem A. Almenweer. "Quantum Denoising Autoencoder Improves Retinal Fundus Image Quality for Early Diabetic Retinopathy Screening." Scientific Reports 16, 5970 (2026).

[9] Rajitha, Chilukuri, P. Praveen, and K. Rajchandar. "Enhanced Detection of Diabetic Retinopathy Using GAI-Net: Deep Learning Model based on GAN with Autoencoder." Ain Shams Engineering Journal 17, no. 3 (2026): 104050.

[10] Li, Yixiao, Boyu Yu, Mingwei Si, Mengyao Yang, Wenxuan Cui, Yi Zhou, Shujun Fu, Hong Wang, Xuya Liu, and Han Zhang. "Enhancing Diabetic Retinopathy Diagnosis: Automatic Segmentation of Hyperreflective Foci in OCT via Deep Learning." International Ophthalmology 45, no. 1 (2025): 79.

[11] Chen, Qiyuan. "Denoising Diffusion with Enhanced Dual-granularity Prior for Diabetic Retinopathy Grading." In 2025 19th International Conference on Complex Medical Engineering (CME), IEEE, 2025, 14-18.

[12] Shamrat, FM Javed Mehedi, Rashiduzzaman Shakil, Bonna Akter, Md Zunayed Ahmed, Kawsar Ahmed, Francis M. Bui, and Mohammad Ali Moni. "An Advanced Deep Neural Network for Fundus Image Analysis and Enhancing Diabetic Retinopathy Detection." Healthcare Analytics 5 (2024): 100303.

[13] Kirubakaran, M., and V. Vijayarajan. "WaveMem-SHAPNet: A Transparent Deep Learning Approach to Early Diagnosis of Diabetic Retinopathy." SN Computer Science 7, no. 1 (2026): 73.

[14] Yu, Jingning. "Based on Gaussian Filter to Improve the Effect of the Images in Gaussian Noise and Pepper Noise." In Journal of Physics: Conference Series, vol. 2580, no. 1, IOP Publishing, 2023, 012062.

[15] Cheng, Jilan, Guoli Long, Zeyu Zhang, Zhenjia Qi, Hanyu Wang, Libin Lu, Shuihua Wang, Yudong Zhang, and Jin Hong. "WaveNet-SF: A Hybrid Network for Retinal Disease Detection Based on Wavelet Transform in Spatial-Frequency Domain." arXiv preprint arXiv:2501.11854 (2025).

[16] Yao, Qihai, Yong Wang, and Yixin Yang. "Underwater Acoustic Target Recognition Based on Data Augmentation and Residual CNN." Electronics 12, no. 5 (2023): 1206.

[17] Li, Fangfang, Qizhou Wu, Bei Jia, and Zhicheng Yang. "Speckle Noise Removal in OCT Images via Wavelet Transform and DnCNN." Applied Sciences 15, no. 12 (2025): 6557.

[18] Rassam, Murad A. "Autoencoder-Based Neural Network Model for Anomaly Detection in Wireless Body Area Networks." IoT 5, no. 4 (2024): 852-870.

[19] Huang, Tengda, Sheng Fu, Haonan Feng, and Jiafeng Kuang. "Bearing Fault Diagnosis Based on Shallow Multi-Scale Convolutional Neural Network with Attention." Energies 12, no. 20 (2019): 3937.

[20] Ding, Mingyu, Bin Xiao, Noel Codella, Ping Luo, Jingdong Wang, and Lu Yuan. "Davit: Dual Attention Vision Transformers." In European conference on computer vision, Cham: Springer Nature Switzerland, 2022, 74-92.

[21] https://www.kaggle.com/competitions/diabetic-retinopathy-detection/data

[22] Jebur, Rusul Sabah, Mohd Hazli Bin Mohamed Zabil, Dalal Abdulmohsin Hammood, Lim Kok Cheng, and Ali Al-Naji. "Image Denoising Using Hybrid Deep Learning Approach and Self-Improved Orca Predation Algorithm." Technologies 11, no. 4 (2023): 111.

[23] Tiantian, Wang, Zhihua Hu, and Yurong Guan. "An Efficient Lightweight Network for Image Denoising Using Progressive Residual and Convolutional Attention Feature Fusion." Scientific Reports 14, no. 1 (2024): 9554.

[24] Karthik, S. A., M. N. Geetha, K. Prabhavathi, Dhananjaya Shashank, K. P. Suhaas, and M. Narender. "Early Detection and Severity Classification of Diabetic Retinopathy Using Convolutional Neural Networks." SN Computer Science 6, no. 7 (2025): 819.

[25] Muthusamy, Dharmalingam, and Parimala Palani. "Deep Learning Model Using Classification for Diabetic Retinopathy Detection: An Overview." Artificial Intelligence Review 57, no. 7 (2024): 185.