

Research Article

Optimizing Feature Selection in Retinal Disease Diagnosis Using RViT-AdaMobileNetV3 and WGOA Metaheuristics

Sashi Kanth Betha^{1,*} , Joseph Beatrice Seventline¹ 

¹ Department of Electrical & Electronics and Communication Engineering, Gandhi Institute of Technology and Management (GITAM), Visakhapatnam, 530045, India

* Corresponding Author: Sashi Kanth Betha, E-mail: sbetha@gitam.in

Article Info	Abstract
Article History	Diabetic retinopathy (DR) is a chronic microvascular complication of diabetes and remains one of the leading causes of preventable blindness worldwide. For DR, early detection and accurate stage classification through fundus image analysis are critical for timely intervention and effective treatment planning. In this work, we present a novel hybrid classification pipeline for DR screening that integrates transformer-based vision models with optimization-guided feature selection. The architecture leverages a dual-stream design, combining the global feature extraction capability of a refined vision transformer (RViT) with the local feature capturing the strength of an adaptive MobileNetV3 (AdaMobileNetV3). This design enhances the modelling of retinal lesions, such as microaneurysms, hemorrhages, and exudates, at a local scale while concurrently capturing global contextual information. A two-stage feature selection strategy is employed, initially removing highly correlated features, followed by selecting the most discriminative subset using the weighted goose optimization algorithm (WGOA). The proposed method is evaluated on two DR datasets featuring varying class distributions, including imbalanced and multi-label scenarios. The proposed architecture outperforms baseline methods (RViT-AdaMobileNetV3 and RViT-AdaMobileNetV3+GA), achieving macro F1-scores of 0.9800 on Dataset A and 0.9764 on Dataset B, thereby demonstrating its robustness, adaptability, and translational potential for cross-dataset DR classification tasks.
Received Jul 01, 2025	
Revised Sep 06, 2025	
Accepted Sep 10, 2025	
Keywords	
Diabetic Retinopathy	
Fundus Image Classification	
Vision Transformer (RViT)	
MobileNetV3	
Hybrid Deep Learning	
Feature Selection	
Weighted Goose Optimization Algorithm (WGOA)	
Recursive Feature Elimination	
Support Vector Machine	
Medical Image Analysis	



Copyright: © 2025 Sashi Kanth Betha, and Joseph Beatrice Seventline. This article is an open-access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY 4.0) license.

1. Introduction

Diabetes mellitus is a prevalent and chronic systemic metabolic disorder that results from inadequate insulin secretion or insulin resistance. This imbalance causes abnormally high blood glucose concentration, which is called hyperglycemia [1-3]. Bonaventure's analysis [4] of prevalence and incidence data from Western countries shows a sharp rise in the population-level burden of diabetes over the last several decades. The national prevalence of diabetes in China is about 12%, and the country has the largest number of diabetics in the world, due in part to its population base.

Diabetic retinopathy (DR), one of the many long-term complications of diabetes, is one of the most debilitating and is the leading cause of vision loss in working-age adults (20–64 years) [4]. DR develops as a result of prolonged hyperglycemia, which causes abnormal retinal vessel changes, such as neovascularization. These pathological changes include capillary wall weakening, hemorrhages, exudates, macular edema, and neovascularization. In the absence of treatment, DR can lead to partial or total vision loss [5]. Conventional DR signs, including microaneurysms, hard and soft exudates, and hemorrhages, are important indicators found in fundus photography [6].

To standardize diagnosis and therapy, DR has been classified into five stages based on the international clinical diabetic retinopathy disease severity scale (ICDRS): no DR, mild NPDR, moderate NPDR, severe NPDR, and PDR. One of the major challenges in DR treatment is its asymptomatic onset, whereby patients often remain unaware of the condition until significant visual impairment develops. Therefore, early detection and routine screening are necessary [7, 8]. However, the rapidly growing prevalence of DR is not matched by an adequate number of trained ophthalmologists [9]. Moreover, even experienced clinicians may find it difficult to accurately determine disease severity, thus contributing to higher misdiagnosis rates.

To overcome these difficulties, computer-aided diagnosis (CAD) systems are becoming increasingly important for clinical practice. Such systems can alleviate healthcare workload, promote diagnosis consistency, and accelerate the detection of diabetic-related complications. Convolutional neural networks (CNNs), recognized as one of the most successful techniques in machine learning, have demonstrated significant capability in analyzing medical images, particularly in DR classification because of their capacity to extract complicated visual information hierarchically [10–12].

Over the past decade, numerous deep learning-based methods for diagnosing diabetic retinopathy have emerged. The DcardNet model [13], with adaptive rate dropout and three independent stage predictions, is an example model designed based on clinical diagnostic requirements. This structure increases reliability and allows more effective training of patients for more in-depth ophthalmic examination. Another study [14] used a modified ResNet-18 with a Swish activation, and the results were state of the art compared with regular CNN models. Likewise, Kao and Lin [15] suggested incorporating EfficientNetV2 and dark channel prior through fusion to further improve lesion visibility and DR detection.

Motivated by the limited number of annotated datasets, [16] proposed a work that combines the eye-tracking of ophthalmologists with the fundus images, applying the OTSU thresholding and a class activation map (CAM)-based regularization. This resulted in a more accurate and interpretable approach that obtained 94% and 94.83% classification accuracies on the DIARETDB0 and DIARETDB1 datasets, respectively.

However, a number of concerns must be addressed to put CNN-based DR classifiers into clinical

practice. Inter-class discrimination is challenging due to concurrent lesions and pathological patterns. In addition, for DR datasets, data imbalance may cause model performance bias and weak generalization.

To combat these challenges, attention mechanisms have been widely adopted. Such mechanisms have been shown to help the models focus on the most informative parts of an image, achieving better recognition performance in fine-grained classification. For instance, CABNet [17] adopts the global attention block (GAB), incorporating the class-specific attention block (CAB) to learn global and class-related features. In [18], a texture-spatial attention mechanism is proposed to improve the saliency-based learning via an encoder-fusion structure.

To address the shortcomings of single-view fundus images, Luo, et al. [19] introduced MVDRNet, a multi-view deep learning model that integrates features extracted from different views by CNNs and attention modules. Such enhanced feature fusion results in more comprehensive lesion characterization and improved diagnostic performance. Another prominent work is the CNN-Trans hybrid model [20], which combines CNN-LSTM with Xception and Vision Transformer (ViT). This concatenation improved performance, and when combined with coordinate and self-attention, it reached 80.68% accuracy, thus presenting comparable state-of-the-art efficiency.

These studies also highlight the notable advancement of automatic DR detection. There are significant limitations, most importantly regarding overlapping features, class imbalance, and low interpretability. These difficulties call for new solutions based on previous knowledge, attention mechanisms, and hybrid models to improve the diagnostic reliability and clinical uptake.

The following research gaps (RGs)—RG-1 on global-local feature fusion and RG-2 on advanced feature selection—are systematically addressed in this study through a hybrid RViT-AdaMobileNetV3 architecture and a two-stage feature refinement strategy:

RG-1: Existing CNN architectures lack robust global-local feature fusion, limiting their ability to capture both fine-grained retinal textures and broader contextual patterns.

RG-2: Existing pipelines perform little or no effective feature selection, leading to redundancy, overfitting, and poor generalization in high-dimensional deep features.

The novelty and contributions of this study include

- i. A hybrid architecture
- ii. Robust two-stage feature selection
- iii. Improved generalization and interpretability

2. Literature Review

In recent years, the need for accurate and early diagnosis of diabetic retinopathy (DR) has driven significant advances in computer-aided diagnosis (CAD) systems. These systems aim to support

ophthalmologists in analyzing retinal fundus images, particularly by leveraging deep learning techniques, which have demonstrated superior performance in feature extraction and classification across a variety of image-based medical domains.

Several CNN-based architectures have shown promising results in DR classification tasks. In [21], the authors used MobileNetV2 with transfer learning and achieved 0.91 accuracy in DR classification and 0.92 in recognizing DR, with an AUC of 0.9. Their model also obtained macro accuracy, recall, and F1-scores of 0.78, 0.83, and 0.80, respectively, showcasing its potential for clinical adoption. In a related study, the authors in [22] formulated the DR classification problem, reporting accuracies of 0.91 and 0.90 on two different real-world datasets.

Addressing the scarcity of labeled data, Yang, et al. [23] proposed a channelized generative adversarial network (GAN) with semi-supervised learning, incorporating three modules: a generative model to produce sub-field images, a GAN-based semi-supervised classifier to utilize both labeled and unlabeled samples, and a DR-specific feature extractor for high-resolution image representation. These innovations enabled robust performance under limited supervision.

The integration of image processing with transfer learning is explored in [24], where the authors utilized the Xception model along with advanced preprocessing techniques to extract richer features. Their approach achieved 92% training accuracy and 88% test accuracy, demonstrating the effectiveness of hybrid strategies. Similarly, the authors in [25] fused features derived from ResNet-50 and applied the random forest classifier. The use of average pooling layer features from ResNet enabled efficient training of deeper models, overcoming vanishing gradient issues and enhancing predictive accuracy.

A classic machine learning pipeline is reported in [26], where feature extraction was used to detect DR severity from 0 to 4. Among various classifiers tested, naïve Bayes achieved 83% accuracy, emphasizing that even traditional models remain relevant when feature engineering is effective.

Gao, et al. [27] introduced DLGP-DR, a deep learning model that integrates Gaussian processes to improve classification and probabilistic ranking. On the EyePACS dataset, the model reached a sensitivity of 93%, specificity of 92%, and AUC of 0.97. The performance remained strong on the Messidor dataset with sensitivity, specificity, and AUC of 72%, 86%, and 0.88, respectively, validating its cross-dataset applicability.

Two novel architectures, CKML-Net and VNXX, are proposed by Wang, et al. [28] for fine-grained DR classification. Inspired by GoogLeNet and VGGNet, respectively, these models incorporate multi-kernel learning, multiple loss functions, and a hybrid LGI color space, while leveraging transfer learning to address class imbalance. The best results are reported on both datasets used in the study.

Yaqoob, et al. [29] further enhanced the ResNet50 architecture by integrating attention mechanisms, achieving a training accuracy of 0.98 and an F1-score of 0.95, indicating substantial improvements over

existing techniques. Similarly, the TaNet model [30] demonstrated superior classification performance using transfer learning on Messidor-2, EYEPACS-1, and APTOS 2019 datasets, with precision, recall, and F1-score of 98.75%, 97.89%, and 98.89%, respectively.

3. Proposed Methodology

3.1. Overview of the Proposed Method

This section presents the adopted methodological framework as shown in Figure 1, used to design, implement, and evaluate the proposed RViT-AdaMobileNetV3 + WGOA classification pipeline for retinal disease detection. The procedure was structured to ensure methodological rigor and reproducibility, comprising three core components: (1) the study design, which outlines the experimental protocols, datasets, and reference benchmarks; (2) the implementation plan, detailing a step-by-step process from data preprocessing through feature extraction to optimization and final classification; and (3) the analytical strategies, encompassing performance metrics, evidence-based statistical analysis, and visualization techniques to interpret the results. Each component is interconnected, collectively demonstrating how the study was conceived, executed, and interpreted to substantiate the effectiveness of the proposed method. The method is generalizable over datasets with varying disease types and class distributions, ensuring robust clinical interpretability. The key steps of the method are summarized in Table 1:

Table 1. Proposed RViT-AdaMobileNetV3 model.

Step 1	Dataset preparation and preprocessing
Step 2	Hybrid deep feature extraction via RVit-AdaMobileNetV3
Step 3	Feature vector construction
Step 4	Feature redundancy elimination
Step 5	Feature optimization using WGOA
Step 6	Final classification with support vector machine
Step 7	Performance reporting and visualization

The weighted goose optimizer (WGOA)-based feature selection phase uses a stratified train-test split to ensure proportional representation of each class during fitness evaluation, which helps maintain class balance in the optimization loop. Additionally, the hybrid RVit-AdaMobileNet model extracts rich feature representations that improve separability across all classes, including minority ones. These features are further refined by eliminating highly correlated attributes and optionally applying recursive feature elimination (RFE) to reduce dimensionality. This sequence not only enhances model performance but also mitigates the bias typically introduced by imbalanced data, as evidenced by class-level insights drawn from the final confusion matrix.

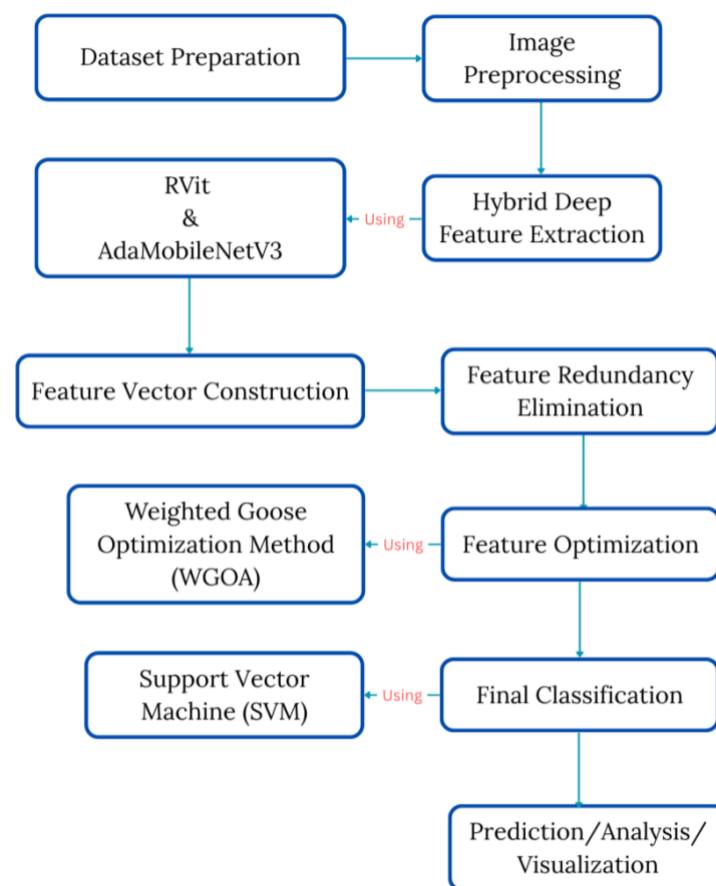


Figure 1. Implementation procedure of the proposed framework.

3.2. Detailed Implementation Steps

3.2.1. Dataset Preparation and Preprocessing

The preparation of the image dataset is carried out in structured folders, with each RGB-transformed image tensor batched with a loader that supports data shuffling and on-the-fly transformations. The $\mathcal{D} = \{(x_i, y_i)\}_{i=1}^N$ is batched to ensure consistent input dimension with 224X224 pixels. In addition, it enables handling data efficiently during feature extraction and training. We used two datasets, namely A with four classes and B with seven classes, for this work. Figure 2 and Table 2 illustrate the procedure:

Table 2. Dataset preparation and preprocessing.

Input -	
Dataset with class-wise ground truth labels	
Step 1	All images resized to 224 X 224 pixel resolution
Step 2	PyTorch transformations used to tensorize and convert to RGB format
Step 3	Implementation of a PyTorch-compliant custom dataset class with DataLoader-based mini-batch shuffling and parallelized image-label retrieval
Output -	
i.	Tensor-formatted images with class labels as (image_tensor, label) batches
ii.	DataLoader object

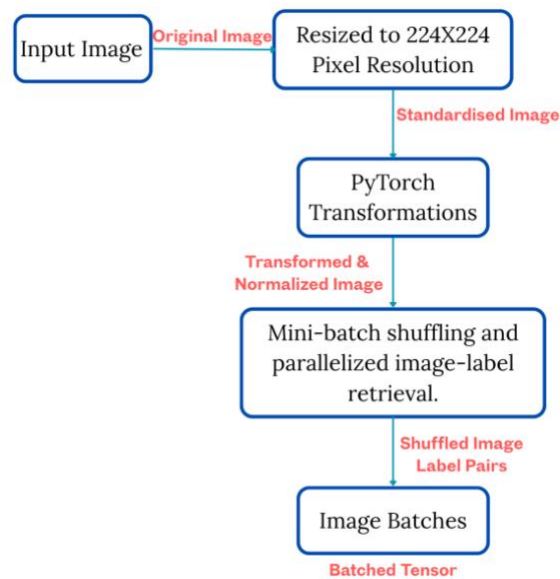


Figure 2. Implementation procedure for data preprocessing.

3.2.2. Hybrid Deep Feature Extraction via RViT-AdaMobileNetV3

A novel hybrid architecture, RViT-AdaMobileNet, is developed by integrating two powerful backbones. Table 3 demonstrates the procedure:

- i. RefinedVision transformer (RViT)
- ii. AdaMobileNetV3

Table 3. Hybrid deep feature extraction via RViT-AdaMobileNetV3 model.

Input -	
Image batches from DataLoader	
Step 1	Load pretrained refined vision transformer (RViT) and AdaMobileNetV3 without classification heads.
Step 2	Forward input through both networks and concatenate their feature vectors.
Step 3	Apply dense classification head and train using Adam optimizer and categorical cross-entropy loss.
Output -	
i.	Trained RVitAdaMobileNet hybrid feature extractor
ii.	Optimal weights learned from the dataset

The RViT module is used to encode the images through self-attention after splitting into patches. It gives an output $f_{\text{RViT}} \in \mathbb{R}^{768}$. AdaMobileNetV3 processes the image in a lightweight manner to generate feature representations, $f_{\text{Mobile}} \in \mathbb{R}^{960}$. These two are concatenated to form f_{fused} , as shown in Equation 1:

$$f_{\text{fused}} = [f_{\text{RViT}}, f_{\text{Mobile}}] \in \mathbb{R}^{1728} \quad (1)$$

A multi-layer perceptron projects the f_{fused} vector and is trained using a cross-entropy loss function, as shown in Equation 2:

$$\mathcal{L}_{CE} = -\sum_{c=1}^C y_c \log(\hat{y}_c) \quad (2)$$

This hybrid integration captures both global attention and local texture features in a most effective way. This also improves the robustness of the model more than using a single model, as shown in Figure 3:

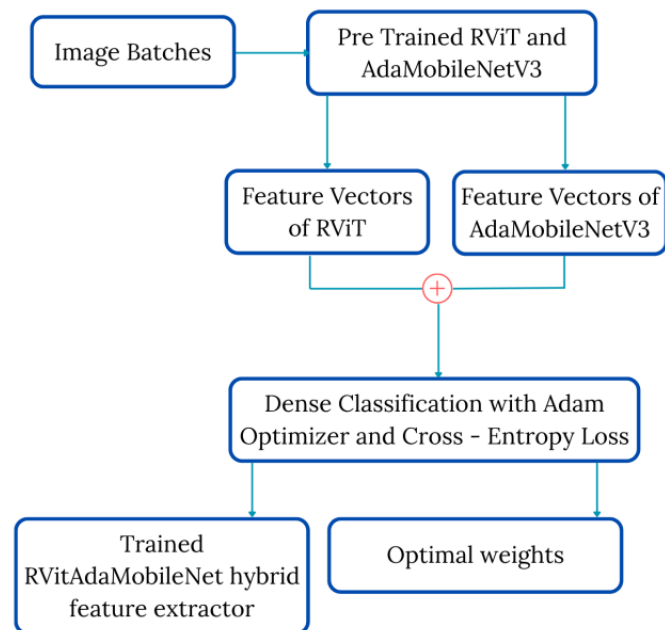


Figure 3. Implementation procedure for Hybrid RViT-AdaMobileNetV3.

3.2.3. Feature Vector Construction:

The trained hybrid model is then turned into a feature extractor by discarding the classification head. The step-wise procedure is described in Table 4. For each input image, we forward it via the model in evaluation mode to obtain high-dimensional deep features, which are then saved along with their labels for downstream tasks. Each image $x \in \mathbb{R}^{3 \times 224 \times 224}$ is mapped to a high dimensional feature vector f_{fused} . A feature vector is produced across the dataset in the form represented in Equation 3:

$$X = [f_{\text{fused}}^{(1)}, f_{\text{fused}}^{(2)}, \dots, f_{\text{fused}}^{(N)}] \in \mathbb{R}^{N \times 1728} \quad (3)$$

Table 4. Feature vector extraction using RViT-AdaMobileNetV3 model.

Input -	
i.	Trained RViTAdaMobileNet
ii.	Complete dataset via DataLoader.
Step 1	Disable gradient computation (evaluation mode).
Step 2	Pass each image through the model.
Step 3	Store resulting feature vectors and their corresponding labels.
Output -	
i.	Features: NumPy array of shape (N, D)
ii.	Labels: Ground truth vector

$y \in \mathbb{R}^N$ represents the corresponding labels. These are the deep features that are then fed as input for further selection and classification of diabetic retinopathy, as shown in Figure 4:

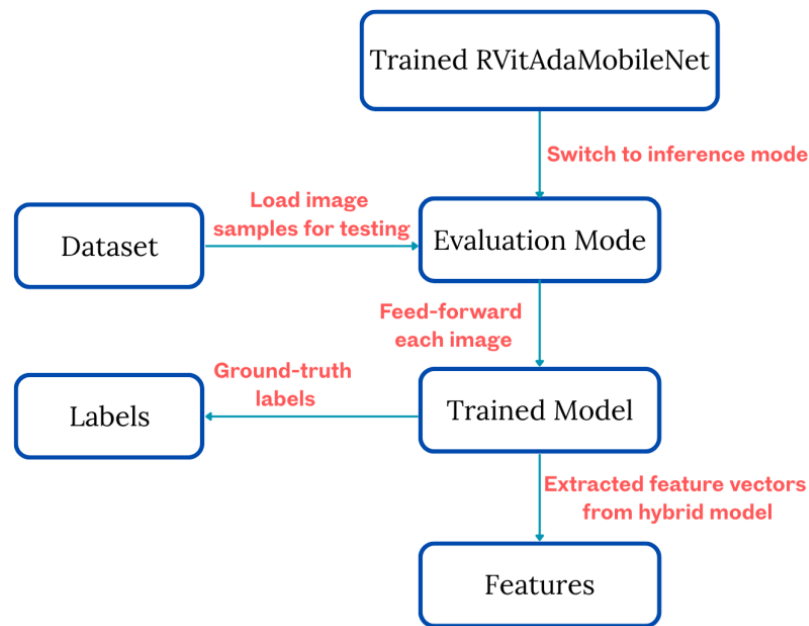


Figure 4. Implementation procedure for feature vector construction using RViT-AdaMobileNetV3.

3.2.4. Feature Redundancy Elimination

We calculated a Pearson correlation matrix to identify and eliminate highly correlated features (threshold > 0.9). Optionally, if the number of features is less than 50, then RFE with a linear SVM model is used to further reduce the dimension. Table 5 summarizes these steps. A Pearson correlation coefficient ρ_{ij} is calculated for each feature pair using Equation 4:

$$\rho_{ij} = \frac{\text{Cov}(X_i, X_j)}{\sigma_{X_i} \sigma_{X_j}} \quad (4)$$

Table 5. Feature redundancy elimination.

Input -	
i.	Extracted features
ii.	Corresponding labels
Step 1	Compute correlation matrix.
Step 2	Drop features with Pearson correlation coefficient > 0.9 .
Step 3	If remaining features exceed 50, apply recursive feature elimination (RFE) with a linear SVM.
Output -	
i.	Filtered feature matrix

In another perspective, the features that have $\rho_{ij} > 0.9$ are removed. This reduces the number of features for efficiency and to decrease overfitting. The output is called a filtered feature matrix. We employed recursive feature elimination using a linear support vector machine. This produces a reduced feature space while increasing the generalization and predictability of the machine learning model, as shown in Figure 5:

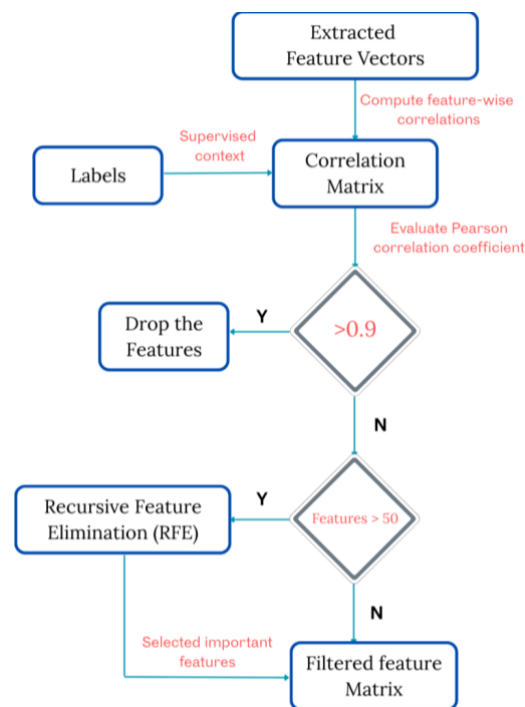


Figure 5. Implementation Procedure for Filtered Feature Matrix

3.2.5. Feature Optimization Using WGOA:

The WGOA brings in migration, mutation, and probabilistic weighting schemes to help maintain population diversity and prevent the algorithm from sinking prematurely into local optima. WGOA is an effective feature subset selection algorithm for dimensionality reduction, particularly aimed at classification, as it encourages the selection of compact feature subsets that exhibit lower redundancy and higher predictive content. This step involves WGOA to get the best feature subset. Each goose represents a candidate feature subset encoded as binary vector $g \in \{0,1\}^d$. The steps are outlined in Table 6. The fitness function in Equation 5, which is the accuracy of an SVM trained with selected features, defines each goose:

$$\text{Fitness}(g) = \text{Accuracy}\left(\text{SVM}(X_g, y)\right) \quad (5)$$

Table 6. Feature Optimization Using WGOA

Input -	
i.	Features_selected, labels
Step 1	Initialize binary population of candidate feature masks.
Step 2	Iteratively evaluate fitness (classification accuracy via SVM).
Step 3	Guide population toward best-performing solution with probabilistic migration and mutation steps (WGOA).
Output -	
i.	Optimized feature vectors
ii.	Binary mask representing chosen features

The weights in this instance are chosen to favor candidate subsets without redundant features. Geese migrate to the best-performing individual at the end of each generation, with stochastic mutation introduced at regular intervals to maintain population diversity, as in Eq(6):

$$g_{\text{new}} = \text{Mutate}(\text{Crossover}(g_{\text{top}}, g_{\text{current}})) \quad (6)$$

The subset that performed best, $\mathcal{S} \subseteq \{1, 2, \dots, d\}$, is chosen and used for classification, as shown in Figure 6:

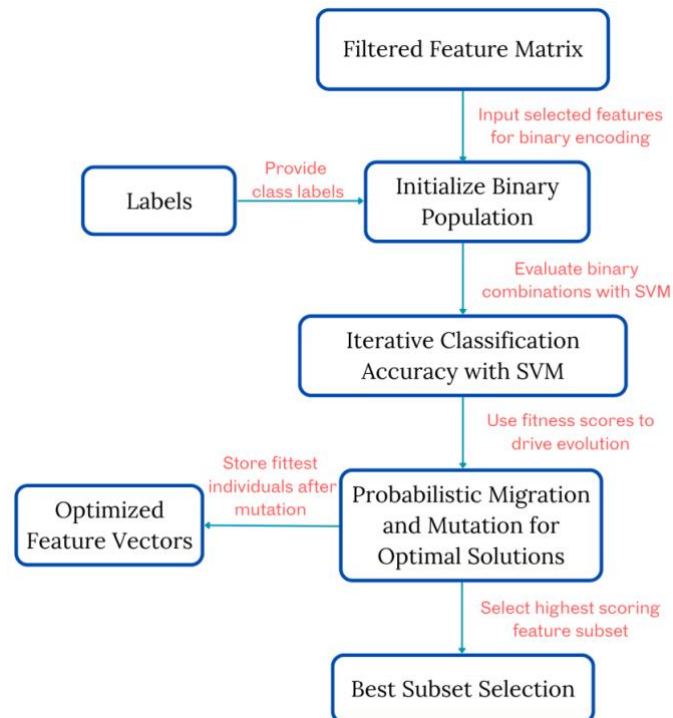


Figure 6. Implementation procedure for feature optimization using WGOA.

3.2.6. Final Classification with Support Vector Machine

The SVM is a powerful classifier that can efficiently deal with high-dimensional data with wide margins. Stratified splitting helps present a balanced class representation, which is an essential feature in medical sets with imbalanced classes. The selected features $X_{\mathcal{S}} \in \mathbb{R}^{N \times |\mathcal{S}|}$ are then split into training and test subsets. Table 7 presents the steps in the process. A linear SVM classifier with hinge loss is trained as defined in Equation 7. Predictions are then made on the test using accuracy and confusion matrix analysis, as shown in Figure 7:

$$\mathcal{L}_{\text{SVM}} = \frac{1}{2} \|w\|^2 + C \sum_{i=1}^N \max(0, 1 - y_i(w^T x_i + b)) \quad (7)$$

Table 7. Final classification with Support Vector Machine.

Input -	
i.	Selected_features, labels
Step 1	Perform an 80:20 stratified train-test split.
Step 2	Train a linear support vector classifier.
Step 3	Evaluate predictions on the test set.
Output -	
i.	Predicted labels
ii.	Confusion matrix
iii.	Accuracy score

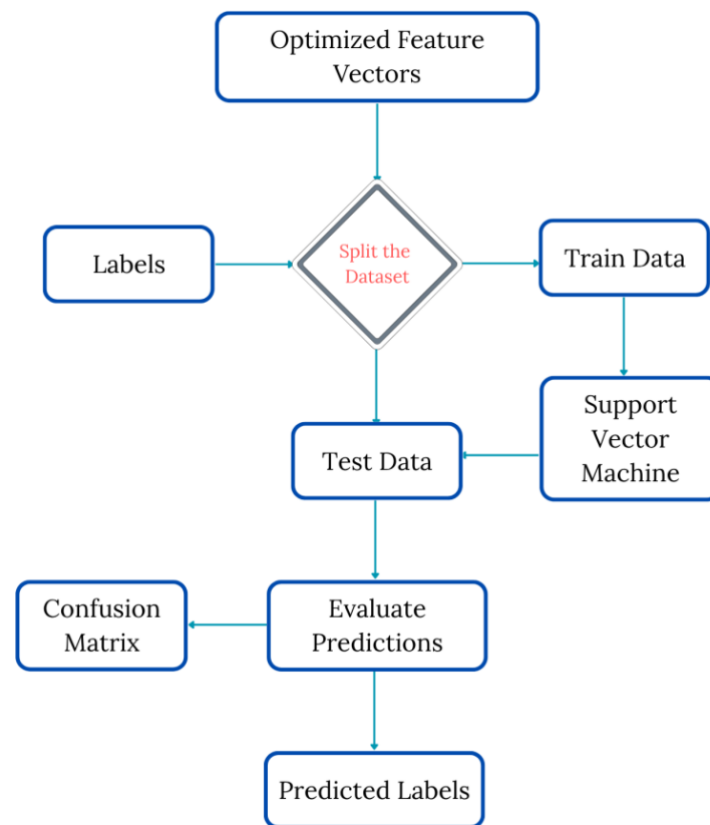


Figure 7. Implementation procedure for final classification using SVM.

3.2.7. Performance Reporting and Visualization:

Predictions are compared against the ground truth labels. Evaluation measures, such as recall, precision, and F1-score, are calculated. The confusion matrix is then drawn by heatmap to provide detailed information on how the classifier is performing in all the classes. These measures not only achieve a holistic perspective of model performance but also provide detailed per-class performance, contributing to interpretations in medical diagnosis. Visualization is also a valuable tool to realize model biases or model systematic errors in the form of class-specific predictions. Table 8 and Figure 8 outline these steps:

Table 8. Performance reporting and visualization.

Input -	
i.	Ground truth y_{test} and predicted y_{pred}
Step 1	Compute precision, recall, F1-score, and accuracy.
Step 2	Visualize confusion matrix via heatmap.
Step 3	Evaluate predictions on the test set.
Output -	
i.	Evaluation metric values
ii.	Annotated confusion matrix plot

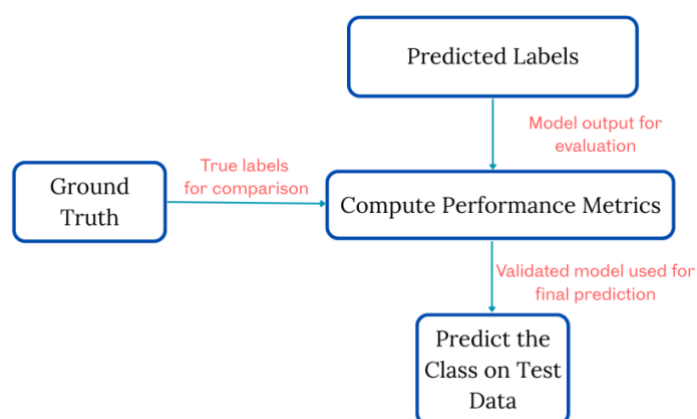


Figure 8. Implementation procedure for prediction.

4. Method Validation

In this study, the macro-averaged scores macro precision, macro recall, macro F1-score, macro false positive rate (macro FPR), and macro false negative rate (macro FNR) are employed to ensure a fair and class-balanced evaluation of the performance in medical diagnosis tasks such as diabetic retinopathy classification. The choice is particularly important due to the natural class imbalance in real-world retinal disease datasets, as some of its classes (e.g., no DR and PDR) are over-represented.

Macro averaging treats all classes equally (unlike weighted metrics), disregarding the class imbalance. This allows us to test whether the model also generalizes well on minority or rare classes, an important aspect for clinical applicability. For example, a model that does well on majority classes but fails on severe yet rare conditions would have high weighted scores but low macro scores, signifying potential diagnostic blind spots.

In addition, in healthcare, false positives and false negatives have asymmetric costs. A macro-averaged approach gives a balanced overview of error rates on all topics that prevents the model from overfitting toward the majority classes. Therefore, macro metrics are very well suited to balanced prediction, fairness across disease categories, and robustness of classification, and they are meaningful for validation of the proposed model in a realistic diagnosis scenario.

4.1. Comparative Models:

We constructed two comparative models to evaluate the performance of the proposed RViT-AdaMobileNetV3 + WGOA framework. These models enabled us to compare the influence of architectural fusion and metaheuristic feature selection on classification accuracy and on dimension reduction.

4.1.1. Method 1 - Baseline RViT-AdaMobileNetV3

We first used a baseline model to observe the training effect of our proposed RViT under the audio-visual enhancement strategy.

This setting uses RViT-AdaMobileNetV3 architecture and removes all the feature extractors except the transformer-based RViT module. The output of the last attention block is used to directly feed into one linear SVM for disease diagnosis. No further cross-convolutional fusion or feature selection was employed; this produced a baseline for comparison.

4.1.2. Method 2 - RViT-AdaMobileNetV3 + Genetic Algorithm (GA)

This approach further improves the baseline by fusing the RViT and AdaMobileNetV3 feature representations. Feature selection is carried out after feature concatenation using GA. GA emulates the evolution process—selection, crossover, and mutation—evolving binary feature masks with the highest SVM classification accuracy over generations.

4.2. Dataset Description

4.2.1. Dataset A: Diabetic Retinopathy Fundus Images (Four-Class)

Dataset A [31] comprises a curated collection of high-resolution retinal fundus images categorized into four clinical classes indicative of diabetic eye health conditions. These include

- Class 0 - Normal: Retinal images showing no signs of diabetic retinopathy (*300 images*)
- Class 1 - Cataract: Images with lens opacification, which impairs visibility of retinal features (*100 images*)
- Class 2 - Glaucoma: Images exhibiting optic nerve head cupping or related structural abnormalities (*100 images*)
- Class 3 - Retina disease: Miscellaneous retinal pathologies associated with diabetic degeneration (*100 images*)

All images were resized to 224×224 pixels during preprocessing and normalized to tensor format. Images were randomly split using an 80:20 ratio for training and testing, respectively.

The dataset serves to evaluate the capacity of the proposed RViT-AdaMobileNet + WGOA model in distinguishing between structurally subtle yet clinically significant retinal conditions in diabetic patients.

4.2.2. Dataset B: Hybrid Retinal Disease Classification Dataset (Seven-Class)

Dataset B [32] is a hybrid compilation of retinal fundus images designed to evaluate multi-class classification performance across varying diagnostic categories. It combines two sources:

1. A structured CSV-based dataset containing class labels from 0 to 4, and
2. A folder-based dataset representing additional classes labeled as 5 and 6.

All images were standardized by resizing to 224×224 pixels and normalized to ensure compatibility with the model input pipeline.

The dataset consists of the following categories:

- Class 0: No disease (*129 images*)

- Class 1: Mild non-proliferative diabetic retinopathy (22 images)
- Class 2: Moderate non-proliferative diabetic retinopathy (156 images)
- Class 3: Severe non-proliferative diabetic retinopathy (84 images)
- Class 4: Proliferative diabetic retinopathy (64 images)
- Class 5: Cataract (lens opacification) (100 images)
- Class 6: Glaucoma (optic nerve cupping) (101 images).

Dataset B serves as a comprehensive benchmark for validating the performance of hybrid deep learning models. It assesses not only the classification capability across a spectrum of retinal disease severity but also the model's adaptability in fusing multi-source data, a critical requirement for practical medical image analysis workflows.

After preparing the dataset using a custom class and applying uniform transformations (resizing to 224×224 and tensor normalization), the proposed methodology implicitly addresses class imbalance through stratified sampling and robust feature selection.

4.3. Experimental Results

4.3.1. Dataset A

Table 9 shows that RViT-AdaMobileNetV3 with WGOA-based feature selection significantly demonstrated better performance in all evaluation metrics for Dataset A than any other compared model and that the proposed method had an overall accuracy of 98.35%, while the macro-averaged precision, recall, and F1-score were more than 96%, which implies outstanding generalization performance among all the classes in the case of the infrequent inter-class overlapping. The confusion matrix as shown in Figure 9(c) indicates a small number of misclassifications and strong diagonal dominance, achieving near-perfect separation across all four retinal disease categories.

Method 1 (RViT-AdaMobileNetV3 without feature selection) was much worse in comparison, with 60.33% accuracy and extreme misclassifications where Classes 0 and 2 were completely mistaken for Class 3, and so on. The weak balance of predictions and strong class bias resulted in a low precision and a macro F1-score of approximately 34.32%. Method 2 (RViT-AdaMobileNetV3 with GA) achieved a remarkable improvement with an accuracy of 76.03% and a macro F1-score of 71.26%, which confirmed the advantage of evolutionary feature selection in the classification. However, its confusion matrix still showed significant confusion between Classes 1 and 2, illustrating the poor discriminative power of the method.

In general, our results confirm the efficacy of the proposed hybrid model structure and of the WGOA-driven selection of features. The model not only shrank the size of the fused feature set but also retained—and even enhanced—the discriminative structure required for multi-class classification of

diabetic retinopathy-related fundus images.

Table 9. Classification performance metrics of models on dataset A.

Method	Method 1 (Baseline)	Method 2 (GA)	Proposed (WGOA)
Accuracy	0.6033	0.7603	0.9835
Macro precision	0.3354	0.7139	0.9924
Macro recall	0.3832	0.7306	0.9690
Macro F1-score	0.3432	0.7126	0.9801
Weighted precision	0.4410	0.7837	0.9840
Weighted recall	0.6033	0.7603	0.9835
Weighted F1-score	0.4940	0.7642	0.9833

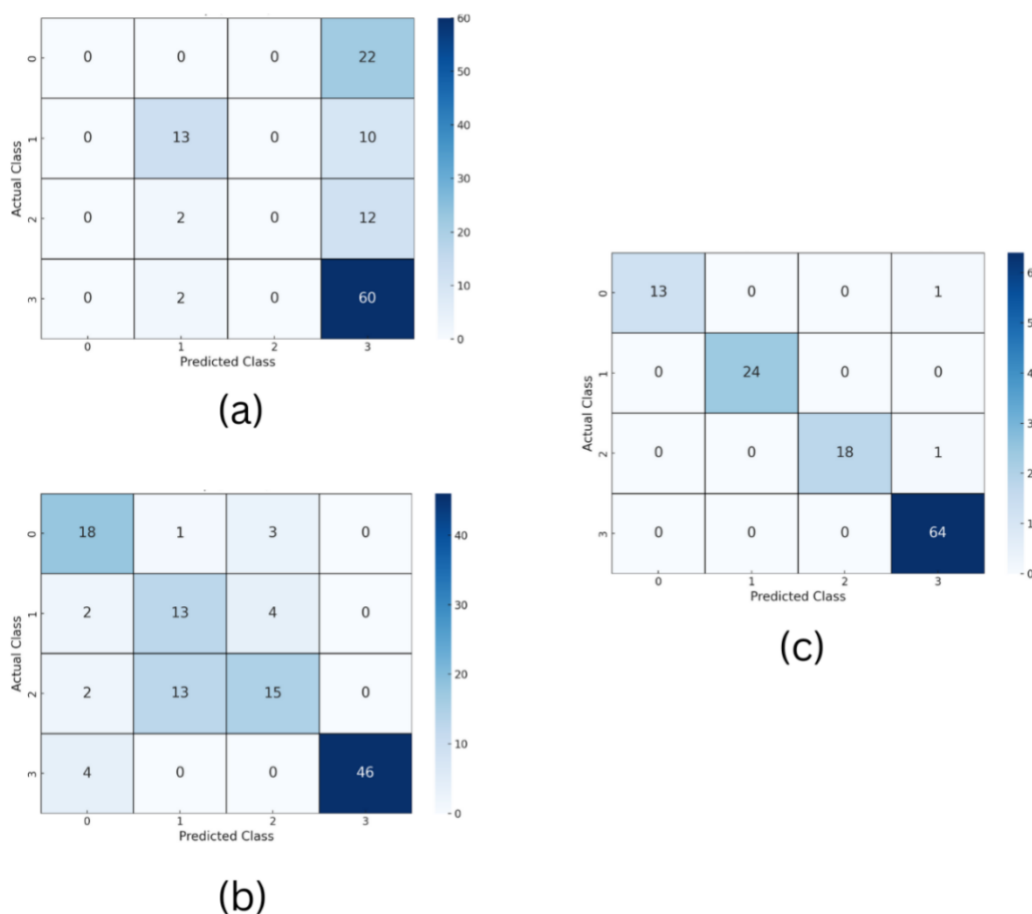


Figure 9. Confusion matrices for dataset A of (a) RViT-AdaMobileNetV3 model (b) RViT-AdaMobileNetV3+GA model (c) Proposed RViT-AdaMobileNetV3+WGOA model.

Table 10. Misclassification performance metrics of models on dataset A.

Method	Macro FPR	Macro FNR
Method 1 (baseline)	0.1966	0.6168
Method 2 (GA)	0.0737	0.2694
Proposed (WGOA)	0.0088	0.0310

4.3.2. Dataset B

The experimental results of the proposed RViT-AdaMobileNetV3 + WGOA on Dataset B (seven different retinal disease classes), as shown in Table 11, demonstrate the excellent classification performance. We also observe an excellent confusion matrix with only single misclassification across all categories. On the whole, accuracy was up to 99.24%, whereas macro-averaged precision, recall, and F1 were close to 1.0, revealing very reliable and balanced classification between the classes. This slight departure from complete diagonal dominance in the matrix proves the discriminative power of the RViT-AdaMobileNetV3 hybrid feature extractor and the effectiveness of the weighted goose optimization algorithm (WGOA) in retrieving concise and discriminative features from high-dimensional deep representations. The findings confirm the built-in robustness of the proposed model and its clinical relevance to challenging multi-class contexts.

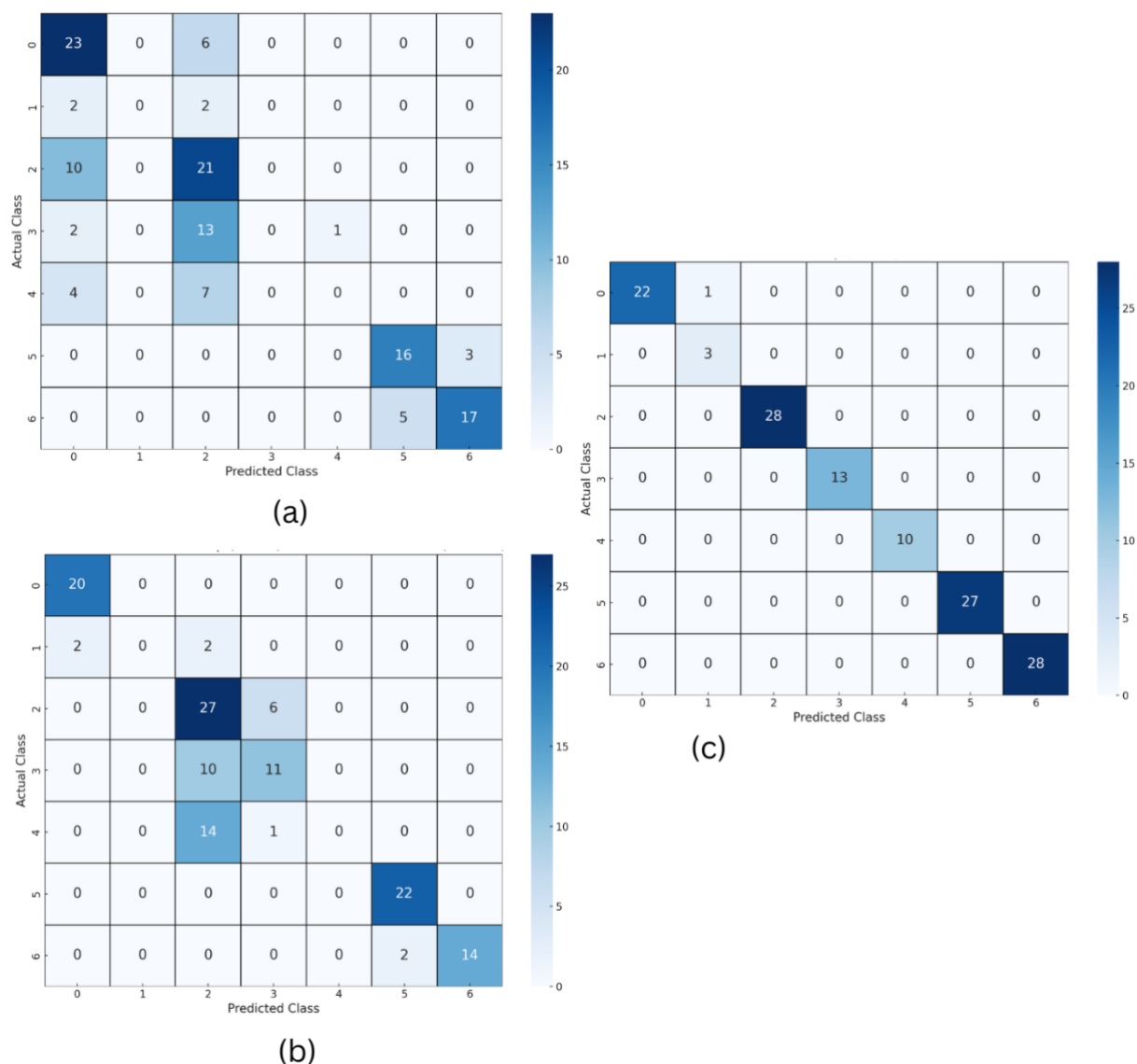


Figure 10. Confusion matrices for dataset B of (a) RViT-AdaMobileNetV3 model (b) RViT-AdaMobileNetV3+GA model (c) Proposed RViT-AdaMobileNetV3+WGOA model.

In contrast, Method 1 (RVit-AdaMobileNetV3 without feature selection) achieved an accuracy as low as 58.33% and generated a macro F1-score of 39.88%. The confusion matrix in Figure 10(c) indicates massive misclassification among the closest disease classes, particularly Classes 2, 3, and 4, implying that the model found it difficult to differentiate similar retinal diseases when no feature selection strategy was used. Method 2 (RVit-AdaMobileNetV3 + genetic algorithm) performed better, obtaining an accuracy of 71.76% and a macro F1-score of 57.63%, but still misclassified borderline cases, particularly between Classes 2 and 3 and Classes 4 and 2. The GA also gave improvements, but it did not have the same accuracy or class distinction as the WGOA.

Table 11. Classification performance metrics of models on dataset B.

Method	Method 1 (Baseline)	Method 2 (GA)	Proposed (WGOA)
Accuracy	0.4719	0.7276	0.9924
Macro precision	0.4721	0.7307	0.9643
Macro recall	0.4719	0.7276	0.9938
Macro F1-score	0.4713	0.7242	0.9764
Weighted precision	0.4576	0.7265	0.9943
Weighted recall	0.4719	0.7276	0.9924
Weighted F1-score	0.4597	0.7248	0.9929

These results highlight the ability of the proposed approach to handle that type of more complex and imbalanced multi-class datasets. The proposed WGOA feature improved classification performance and was calculated with lower computational time to extract a subset of features, demonstrating its applicability to real clinical classification tasks in various retinal diseases.

Table 12. Misclassification performance metrics of models on dataset B.

Method	Macro FPR	Macro FNR
Method 1 (baseline)	0.076	0.5592
Method 2 (GA)	0.052	0.3976
Proposed (WGOA)	0.0088	0.0310

4.4. Feature Extraction Analysis

The following provides a detailed description of the statistical measures used to evaluate the efficacy of the feature extraction pipeline across both datasets (A and B), focusing on apparent feature variance and mutual information (MI). These measures assess the diversity and discriminative power of extracted features, which are crucial for achieving strong classification performance.

The average variance for features in Dataset A was 36.4680, with a maximum of 50.6347 and a minimum of 22.7864 (nominal). This suggests a set of high-to-medium variance values that are fairly evenly distributed across feature pairs, although not all features vary at the same rate. In contrast, Dataset B had a lower average variance of 34.7332 but a wider range, with a maximum of 56.2747 and a minimum of 14.4307. This indicates that while some features in Dataset B exhibit abnormally high variance, others show very low variability, potentially introducing noise into the classification process.

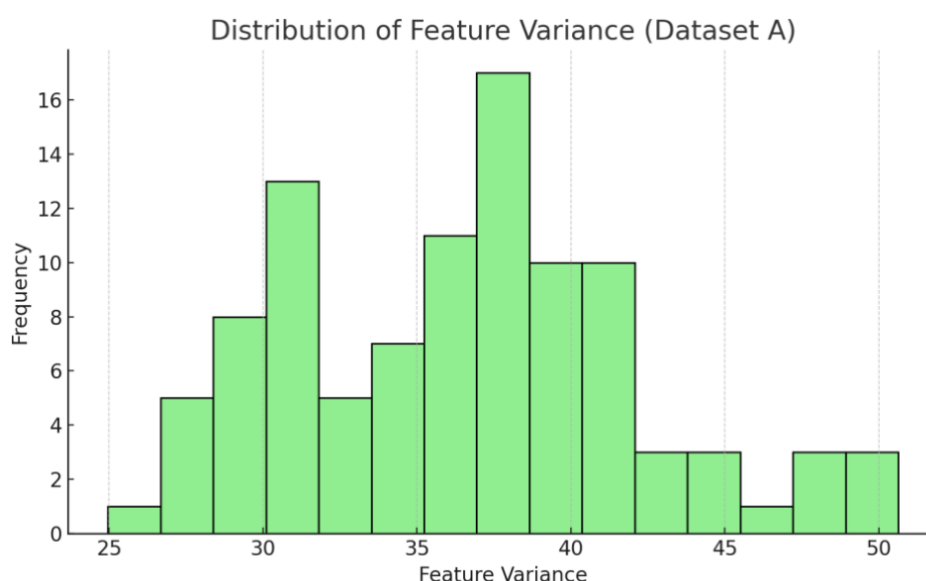


Figure 11. Feature variance distribution of dataset A.

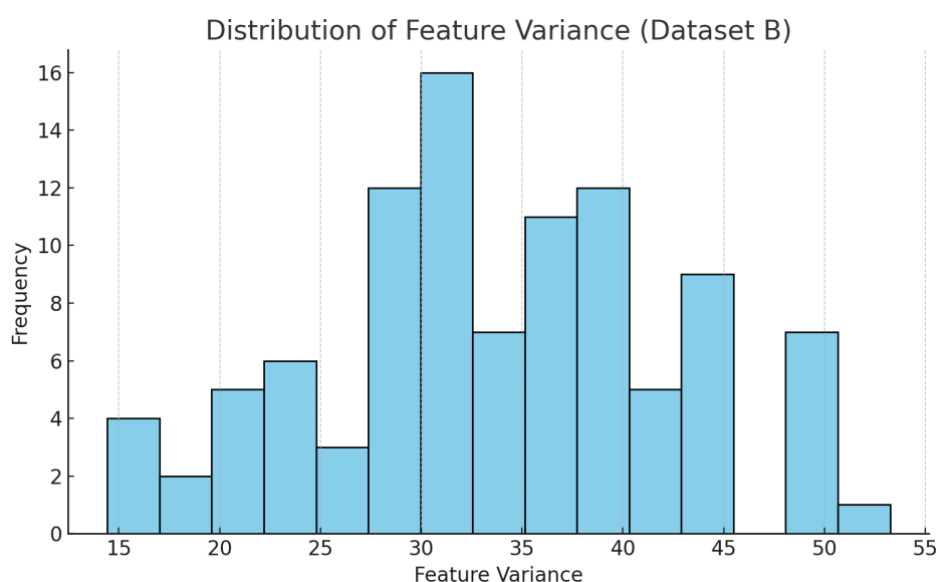


Figure 12. Feature variance distribution of dataset B.

The histograms (Figures 11 and 12) illustrate these distributions, showing that Dataset A's variance is more tightly clustered, whereas Dataset B displays a broader spread. To assess feature importance, we conducted mutual information (MI) analysis. Dataset A showed its highest MI scores at 0.5038, 0.5399, 0.7450, and 0.7694, indicating a moderate degree of dependence between the top-ranked features and class

labels. While these features are informative, their discriminative power could still be improved. In comparison, Dataset B achieved higher MI scores—0.9573, 0.9797, 0.9899, 1.0266, and 1.1381—indicating that its selected features were more strongly correlated with the target classes than those in Dataset A. The bar charts (Figures 13 and 14) make these differences in feature relevance visually apparent:

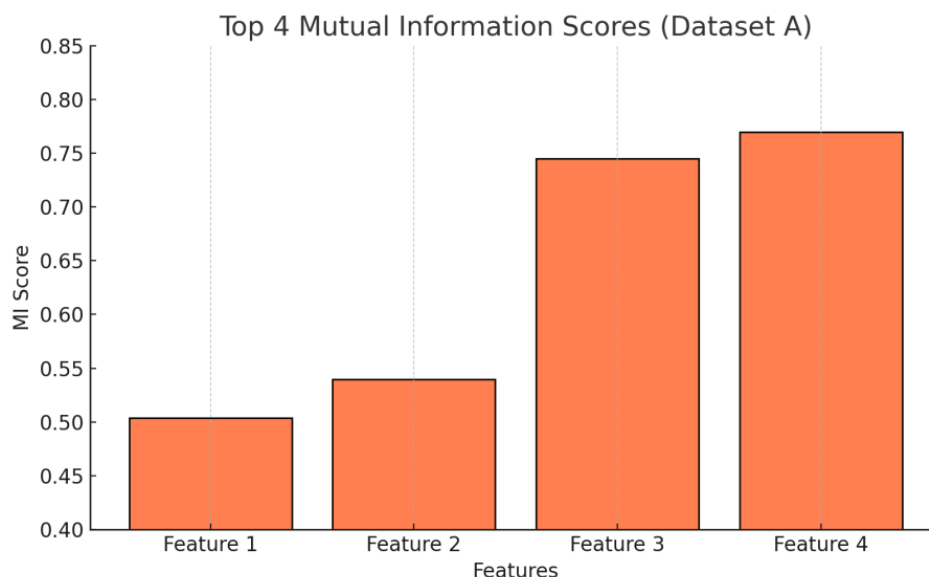


Figure 13. Top four mutual information scores of dataset A.

Comparing feature extraction results from Datasets A and B is valuable because it provides insight into both the statistical properties of the features and their practical significance in real-world classification contexts. Firstly, the MI values generally indicate that Dataset B provides more discriminatory features. MI measures how much information a feature contributes about the class labels. In real-world applications such as medical diagnosis, remote sensing, or financial fraud detection, a higher MI implies that the features are more informative for predicting outcomes.

For example, in a medical imaging application, if a feature representing lesion texture has a high MI value, it would provide strong evidence of the condition's presence. Such features are more likely to yield accurate and confident classifications, especially in tasks requiring high specificity and precision.

Dataset A, although having lower MI scores compared to Dataset B, shows less bias in the between-feature variance. Variance reflects the spread or diversity of feature values across samples. If the variance pattern is consistent across all features, it suggests that the features are evenly distributed across different strata of systemic factors, reducing the likelihood of overfitting.

In practical applications such as industrial defect detection or real-time monitoring, models trained on such features tend to generalize well to unseen data because the features remain stable and robust over time. Therefore, for applications requiring long-term stability, Dataset A, despite its slightly lower initial discriminatory capability, may ultimately be more reliable.

The two datasets exhibit complementary strengths, demonstrating that no single statistical measure

(such as MI or variance) can fully determine a feature's value. Similarly, a high MI alone does not guarantee model robustness, and uniform variance does not necessarily ensure strong class separability. Consequently, the analysis supports employing adaptive feature selection techniques: algorithms or pipelines capable of dynamically balancing discriminative power (e.g., MI, ANOVA, chi-square) with statistical consistency (e.g., variance, entropy) throughout the entire feature subset search process.

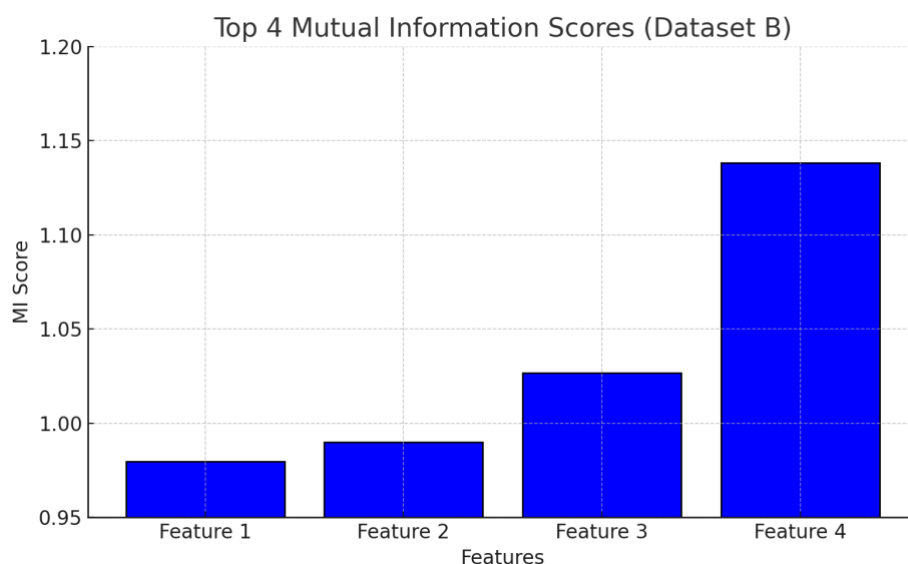


Figure 14. Top four mutual information scores of dataset B.

For instance, in a production-grade AI system, high-MI features from Dataset B could be used for primary classification, followed by the integration of stable features from Dataset A to enhance robustness against noise or data drift. Finally, combining feature sets through ensemble learning or hybrid selection methods (e.g., PSO-RFE, LASSO with MI weighting) could enable the development of models that are both highly accurate and robust, making them suitable for large-scale real-world deployment.

In conclusion, this in-depth study not only enhances understanding of the datasets' technical characteristics but also establishes a strategic foundation for developing intelligent, robust AI systems capable of adapting to changes in data behavior and application constraints.

4.5. Discussion

The experimental results from both Dataset A (four classes) and Dataset B (seven classes) demonstrate consistent and significant improvement with each refinement. The baseline model (Method 1), using the RViT-AdaMobileNetV3 module alone, without any feature-level screening, performed worst on both datasets, attaining 60.33% accuracy on Dataset A with macro-fused precision and recall of 0.3354 and 0.3832, respectively, and 47.19% accuracy on Dataset B with macro-fused precision and recall values of 0.4721 and 0.4719, respectively. These results indicate relatively weak class separation, particularly in the face of class imbalance, and lead to a higher misclassification rate, as also shown in Table 10 by the macro-values of FPR and FNR (0.1966 and 0.6168, respectively, for Dataset A).

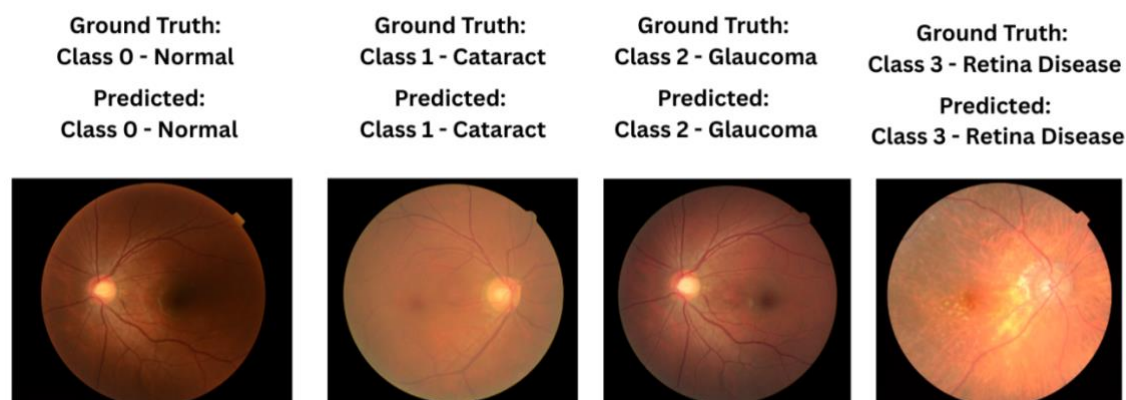


Figure 15. Prediction of images vs. ground truth values of dataset A using proposed model.

This outcome was improved in Approach 2, with the feature selection used on the extracted features based on RViT-AdaMobileNetV3 by a genetic algorithm (GA). This further resulted in substantial improvements: the accuracy on Dataset A increased to 76.03% using the macro precision and recall of 0.7139 and 0.7306, respectively, and macro FPNR and FNR were 0.0737 and 0.2694, respectively. On Dataset B, similar increases were achieved: the accuracy improved to 72.76%, and macro precision and recall reached 0.7307 and 0.7276, respectively. These findings confirm the efficiency of GA-based optimization to improve generalization and minimize the rate of misclassifications; yet, the model still suffered from moderate errors, particularly in detecting minority classes.

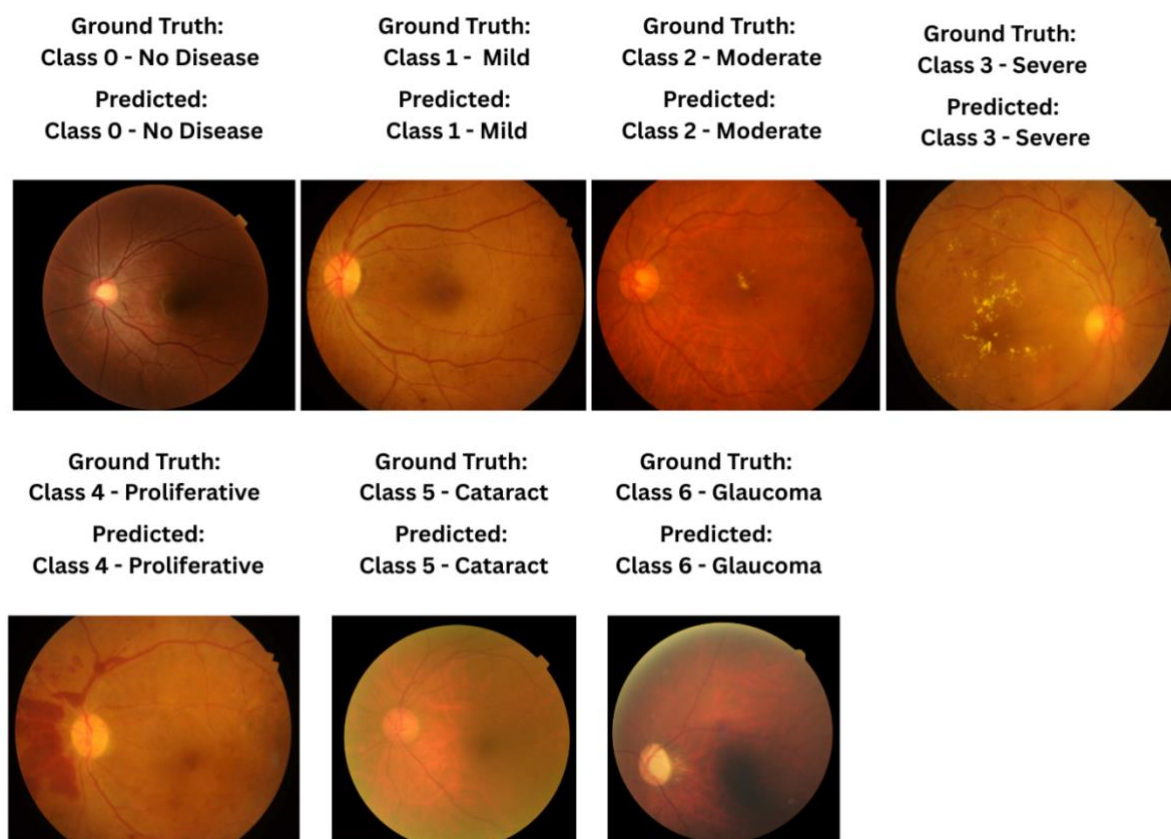


Figure 16. Prediction of images vs. ground truth values of dataset B using proposed model.

In contrast, the proposed approach, RViT-AdaMobileNetV3 with weighted goose optimizer (WGOA), outperformed the state of the art for both datasets. The F1-score of the model on Dataset A is 98.35%, with macro-precision, macro-recall, and macro-F1-score all larger than 0.96; macro-FPR and macro-FNR are kept reduced to 0.0088 and 0.0310, respectively, as shown in Table 12. Additionally, on Dataset B, the accuracy reached a maximum value of 99.24%, and macro metrics overcame 0.96. Weighted scores were also high in both cases, across all detection metrics, thus underlining robustness against class imbalance. These results indicate that hybrid feature fusion demonstrates better discriminative performance, while WGOA can select a subset of features from high-dimensional data, making the proposed approach robust and clinically applicable for the automatic classification of retinal diseases.

Figure 15 shows correctly classified examples from Dataset A containing four categories: normal, cataract, glaucoma, and retinal disease. In each case, the model accurately recognized the class, with ground truth and predicted labels displayed above each image. Figure 16 presents correctly classified examples from the multi-class retinal disease classification model for seven distinct categories: no disease, mild, moderate, severe, proliferative, cataract, and glaucoma. Ground truth labels are shown for each row, with the corresponding predicted labels placed alongside them, demonstrating the model's ability to accurately identify both disease type and severity across diverse retinal pathologies. These examples provide strong evidence of the model's discriminative power, maintaining high accuracy across different disease groups and adapting to varying visual patterns in retinal fundus images.

4.6. Limitations

Although this study highlights several strengths, certain limitations must also be acknowledged. The datasets employed possess properties well-suited for the intended use cases; however, their relatively small size may constrain the model's ability to generalize to broader populations. Moreover, the high-MI feature scenarios observed—particularly in Dataset B—suggest a potential risk of overfitting if not mitigated through robust regularization or adaptive feature selection methods. Additionally, the current validation is limited to the datasets used in this study, meaning that performance under varying imaging conditions and acquisition protocols may not fully generalize to multi-center clinical environments.

5. Conclusion

In this article, we propose a hybrid deep learning and optimization-driven classification pipeline for automated detection of diabetic retinopathy and retinal diseases using fundus images. The integration of RViT and AdaMobileNetV3 helps gather detailed information about the overall context and specific textures in the images. The pipeline fine-tunes the feature set using correlation-based filtering and RFE and then uses a nature-mimicking optimization strategy, the weighted goose optimization algorithm (WGOA), to choose a pool of strongly discriminative features. This methodological fusion not only overcomes the challenges of dimensionality and redundancy but also preserves pertinent task information for high-risk

classification.

The results from testing on two different datasets—one for four types of diabetic eye disease and another for seven types of disease—show that the new method works better than the previous methods, RViT-AdaMobileNetV3 and GA-enhanced, across different measures of success. The results show that the new combined method works better for different types of datasets that have various class distributions and low similarities between classes. Analysis of confusion matrices indicates the minimization of misclassification and enhancement of sensitivity in minority classes.

The higher mutual information in Dataset B enhanced feature discriminability, resulting in a macro F1-score of 0.9764, while the more uniform variance in Dataset A promoted stability and excellent generalization, achieving a macro F1-score of 0.9800. These results further indicate that combining Dataset A with the high-MI features of the underperforming classes from Dataset B yields the most reliable and accurate predictions, consistent with expectations from the feature extraction analysis.

Moreover, cross-dataset experiments demonstrate the robustness and generalization performance of the model, indicating its potential for practical clinical applications. The architecture is modular and scalable, with the possibility to incorporate attention refinement layers, cross-modality inputs, and ensemble decision strategies. In short, a simple and flexible test is suggested that uses transformer vision, lightweight CNNs, and smart methods to choose important features, offering an effective and precise way to check for and assess retinal diseases.

Based on these results, future work will investigate the adaptive weighting of high-MI and stable features during training, enabling the model to dynamically balance discriminability and robustness through a feature fusion strategy. Further research should also explore alternative methods for domain adaptation and continual learning to improve performance on evolving datasets. Extending the approach to multimodal retinal data and other medical imaging domains could further enhance its generalizability.

Data Availability Statement: The datasets used in this study are publicly available on Kaggle (doi: 10.34740/KAGGLE/DSV/12208994; doi: 10.34740/KAGGLE/DSV/12204660). No new datasets were generated.

Funding: The authors declare that no funding was received for this research.

Declaration of Competing Interest: The authors declare they have no known competing interests.

Research Involving Human/Animal Participants: This article does not involve any studies conducted by authors on animals or human participants.

References

- [1] A. T. Kharroubi and H. M. Darwish, "Diabetes mellitus: The epidemic of the century," *World Journal of Diabetes*, vol. 6, no. 6, p. 850, 2015, doi: <https://doi.org/10.4239/wjd.v6.i6.850>.

- [2] S. Yan, L. Ding, and Y. Cai, "Proportional-integral observer-based H_∞ fuzzy controller design of nonlinear time-varying delayed glucose-insulin system with input saturation," *Nonlinear Dynamics*, vol. 113, no. 15, pp. 19887-19899, 2025, doi: <https://doi.org/10.1007/s11071-025-11192-2>.
- [3] S. Yan, L.-l. Chu, and Y. Cai, "Robust H_∞ control of T-S fuzzy blood glucose regulation system via adaptive event-triggered scheme," *Biomedical Signal Processing and Control*, vol. 83, p. 104643, 2023, doi: <https://doi.org/10.1016/j.bspc.2023.104643>.
- [4] M. M. Engelgau *et al.*, "The Evolving Diabetes Burden in the United States," *Annals of Internal Medicine*, vol. 140, no. 11, pp. 945-950, 2004, doi: <https://doi.org/10.7326/0003-4819-140-11-200406010-00035>.
- [5] A. R. Wahab Sait, "A Lightweight Diabetic Retinopathy Detection Model Using a Deep-Learning Technique," *Diagnostics*, vol. 13, no. 19, p. 3120, 2023, doi: <https://doi.org/10.3390/diagnostics13193120>.
- [6] C. P. Wilkinson *et al.*, "Proposed international clinical diabetic retinopathy and diabetic macular edema disease severity scales," *Ophthalmology*, vol. 110, no. 9, pp. 1677-1682, 2003, doi: [https://doi.org/10.1016/S0161-6420\(03\)00475-5](https://doi.org/10.1016/S0161-6420(03)00475-5).
- [7] B. Antal and A. Hajdu, "An ensemble-based system for microaneurysm detection and diabetic retinopathy grading," *IEEE Transactions on Biomedical Engineering*, vol. 59, no. 6, pp. 1720-1726, 2012, doi: <https://doi.org/10.1109/tbme.2012.2193126>.
- [8] G. Pinedo-Diaz *et al.*, "Suitability classification of retinal fundus images for diabetic retinopathy using deep learning," *Electronics*, vol. 11, no. 16, p. 2564, 2022, doi: <https://doi.org/10.3390/electronics11162564>.
- [9] K. Ashwini and R. Dash, "Grading diabetic retinopathy using multiresolution based CNN," *Biomedical Signal Processing and Control*, vol. 86, p. 105210, 2023, doi: <https://doi.org/10.1016/j.bspc.2023.105210>.
- [10] A.-O. Asia *et al.*, "Detection of Diabetic Retinopathy in Retinal Fundus Images Using CNN Classification Models," *Electronics*, vol. 11, no. 17, p. 2740, 2022, doi: <https://doi.org/10.3390/electronics11172740>.
- [11] S. Yan, Z. Gu, J. H. Park, X. Xie, and W. Sun, "Distributed cooperative voltage control of networked islanded microgrid via proportional-integral observer," *IEEE Transactions on Smart Grid*, vol. 15, no. 6, pp. 5981-5991, 2024, doi: <https://doi.org/10.1109/tsg.2024.3407770>.
- [12] J. Wang, Y. Bai, and B. Xia, "Simultaneous diagnosis of severity and features of diabetic retinopathy in fundus photography using deep learning," *IEEE Journal of Biomedical and Health Informatics*, vol. 24, no. 12, pp. 3397-3407, 2020, doi: <https://doi.org/10.1109/jbhi.2020.3012547>.
- [13] P. Zang *et al.*, "DcardNet: diabetic retinopathy classification at multiple levels based on structural and angiographic optical coherence tomography," *IEEE Transactions on Biomedical Engineering*, vol. 68, no. 6, pp. 1859-1870, 2020, doi: <https://doi.org/10.1109/tbme.2020.3027231>.
- [14] S. Sunkari, A. Sangam, R. Raman, and R. Rajalakshmi, "A refined ResNet18 architecture with Swish activation function for Diabetic Retinopathy classification," *Biomedical Signal Processing and Control*, vol. 88, p. 105630, 2024, doi: <https://doi.org/10.1016/j.bspc.2023.105630>.
- [15] Y.-H. Kao and C.-L. Lin, "Enhancing Diabetic Retinopathy Detection Using Pixel Color Amplification and EfficientNetV2: A Novel Approach for Early Disease Identification," *Electronics*, vol. 13, no. 11, p. 2070, 2024, doi: <https://doi.org/10.3390/electronics13112070>.
- [16] H. Jiang *et al.*, "Eye tracking based deep learning analysis for the early detection of diabetic retinopathy: A pilot study," *Biomedical Signal Processing and Control*, vol. 84, p. 104830, 2023, doi: <https://doi.org/10.1016/j.bspc.2023.104830>.
- [17] A. He, T. Li, N. Li, K. Wang, and H. Fu, "CABNet: Category attention block for imbalanced diabetic retinopathy grading," *IEEE Transactions on Medical Imaging*, vol. 40, no. 1, pp. 143-153, 2020, doi: <https://doi.org/10.1109/tmi.2020.3023463>.
- [18] M. D. Alahmadi, "Texture attention network for diabetic retinopathy classification," *IEEE Access*, vol. 10, pp. 55522-55532, 2022, doi: <https://doi.org/10.1109/access.2022.3177651>.
- [19] X. Luo *et al.*, "MVDRNet: Multi-view diabetic retinopathy detection by combining DCNNs and attention mechanisms,"

- Pattern Recognition*, vol. 120, p. 108104, 2021, doi: <https://doi.org/10.1016/j.patcog.2021.108104>.
- [20] S. Liu, W. Wang, L. Deng, and H. Xu, "Cnn-trans model: A parallel dual-branch network for fundus image classification," *Biomedical Signal Processing and Control*, vol. 96, p. 106621, 2024, doi: <https://doi.org/10.1016/j.bspc.2024.106621>.
- [21] G. Alwakid, W. Gouda, M. Humayun, and N. Jhanjhi, "Enhancing diabetic retinopathy classification using deep learning," *Digital health*, vol. 9, pp. 1-18, 2023, doi: <https://doi.org/10.1177/20552076231203676>.
- [22] W. Nazih, A. O. Aseeri, O. Y. Atallah, and S. El-Sappagh, "Vision transformer model for predicting the severity of diabetic retinopathy in fundus photography-based retina images," *IEEE Access*, vol. 11, pp. 117546-117561, 2023, doi: <https://doi.org/10.1109/access.2023.3326528>.
- [23] Y. Yang, Z. Cai, S. Qiu, and P. Xu, "Vision transformer with masked autoencoders for referable diabetic retinopathy classification based on large-size retina image," *Plos one*, vol. 19, no. 3, p. e0299265, 2024, doi: <https://doi.org/10.1371/journal.pone.0299265>.
- [24] N. S. Kumar and B. R. Karthikeyan, "Diabetic retinopathy detection using CNN, transformer and MLP based architectures," in *2021 International Symposium on Intelligent Signal Processing and Communication Systems (ISPACS)*, 2021: IEEE, pp. 1-2, doi: <https://doi.org/10.1109/ispacs51563.2021.9651024>.
- [25] J. Wu, R. Hu, Z. Xiao, J. Chen, and J. Liu, "Vision Transformer-based recognition of diabetic retinopathy grade," *Medical Physics*, vol. 48, no. 12, pp. 7850-7863, 2021, doi: <https://doi.org/10.1002/mp.15312>.
- [26] S. Sheikh and U. Qidwai, "Using MobileNetV2 to classify the severity of diabetic retinopathy," *International Journal of Simulation Systems Science and Technology*, vol. 21, no. 2, pp. 161 - 166, 2020, doi: <https://doi.org/10.5013/ijssst.a.21.02.16>.
- [27] J. Gao, C. Leung, and C. Miao, "Diabetic retinopathy classification using an efficient convolutional neural network," in *2019 IEEE International Conference on Agents (ICA)*, 2019: IEEE, pp. 80-85, doi: <https://doi.org/10.1109/agents.2019.8929191>.
- [28] S. Wang *et al.*, "Diabetic retinopathy diagnosis using multichannel generative adversarial network with semisupervision," *IEEE Transactions on Automation Science and Engineering*, vol. 18, no. 2, pp. 574-585, 2020, doi: <https://doi.org/10.1109/tase.2020.2981637>.
- [29] M. K. Yaqoob, S. F. Ali, M. Bilal, M. S. Hanif, and U. M. Al-Saggaf, "ResNet based deep features and random forest classifier for diabetic retinopathy detection," *Sensors*, vol. 21, no. 11, p. 3883, 2021, doi: <https://doi.org/10.3390/s21113883>.
- [30] K. S. Vani *et al.*, "An Enhancing Diabetic Retinopathy Classification and Segmentation based on TaNet," *Nano Biomedicine & Engineering*, vol. 16, no. 1, pp. 85 - 100, 2024, doi: <https://doi.org/10.26599/nbe.2023.9290041>.
- [31] N. Ibrahim. *DR_FOUR_CLASS_DATASET*, Kaggle, 2025, doi: <https://doi.org/10.34740/KAGGLE/DSV/12208994>.
- [32] N. Ibrahim. *DR_SEVEN_CLASS_DATASET*, Kaggle, 2025, doi: <https://doi.org/10.34740/KAGGLE/DSV/12204660>.