

Efficient Detection of Diabetic Retinopathy Through Deep Learning

Santha Rajani N, M.E. Students

Department of Computer Science and Engineering, MRK Institute of Technology, kattumannarkoil, India

Rajapriya K, Assistant Professor

Department of Computer Science and Engineering, MRK Institute of Technology, kattumannarkoil, India

Akilan A, HOD

Department of Computer Science and Engineering, MRK Institute of Technology, kattumannarkoil, India

Abstract- Diabetic retinopathy (DR) is a major microvascular complication of diabetes and a leading cause of avoidable visual impairment. Manual screening of retinal fundus images is clinically useful but is slow, specialist-dependent, and difficult to scale in rural or resource-limited healthcare environments. This article presents a hybrid framework for DR severity classification. The framework uses preprocessing, data augmentation, GoogleNet and ResNet deep feature extraction, hybrid feature fusion, Adaptive Particle Swarm Optimization (APSO) for feature selection, and machine learning classifiers for final prediction. The experimental results reported in the thesis show that the Random Forest classifier using APSO-optimized hybrid features achieved 94% accuracy, 93% precision, 92% recall, and 93% F1-score. The article also provides engineering-oriented discussion of design choices, system modules, evaluation metrics, implementation details, limitations, and future research directions.

Index Terms- Adaptive Particle Swarm Optimization, convolutional neural network, diabetic retinopathy, fundus image analysis, GoogleNet, medical image classification, ResNet, Random Forest.

I. INTRODUCTION

Diabetic retinopathy is a progressive retinal disease caused by prolonged hyperglycemia and damage to the microvascular structures of the retina. Because the retina converts light into neural signals, vascular leakage, capillary blockage, edema, hemorrhage, ischemia, and abnormal neovascularization can directly reduce visual function. The public health importance of this disease is high because many patients remain asymptomatic during early non-proliferative stages, while treatment is most effective when lesions are detected before irreversible structural damage occurs [1]-[4].

The thesis on which this article is based identifies automated screening as a practical engineering response to the shortage of ophthalmologists and the burden of large-scale manual grading. Conventional retinal examination, fundus photography interpretation, fluorescein angiography, and optical coherence tomography remain essential clinical tools; however, large population screening requires repeatable image-based decision support that can prioritize referrals and reduce expert workload [2], [10], [18].

Computer-aided DR detection has evolved from hand-crafted lesion descriptors to deep learning. Early machine learning systems used segmentation, color thresholding, vessel analysis, and manually selected descriptors for microaneurysms, hemorrhages, exudates, and blood vessels. Although such models are

interpretable, they are sensitive to illumination changes, camera artifacts, retinal pigmentation, class imbalance, and manual feature design. Deep convolutional networks reduce this dependence by learning hierarchical image features directly from data [6], [10], [11].

The proposed thesis framework combines GoogleNet and ResNet as complementary feature extractors. GoogleNet contributes multi-scale feature capture through inception modules, while ResNet improves deep representation learning through residual connections. The hybrid features are further optimized with Adaptive Particle Swarm Optimization before classification by Random Forest, Support Vector Machine, Decision Tree, and Linear Regression models. The purpose is to improve both accuracy and engineering efficiency by selecting the most relevant feature subset for DR severity discrimination.

This work is useful because it integrates several topics studied in advanced engineering programs: image preprocessing, convolutional architecture selection, evolutionary optimization, feature fusion, supervised classification, model evaluation, and software system deployment. The article is therefore written not only as a research report but also as an engineering design case study that shows how artificial intelligence modules can be assembled into a complete medical image analysis pipeline.

II. DIABETIC RETINOPATHY BACKGROUND

DR generally progresses through mild non-proliferative diabetic retinopathy (NPDR), moderate NPDR, severe NPDR, and proliferative diabetic retinopathy (PDR). In mild NPDR, microaneurysms may be the dominant signs. Moderate and severe stages may include increasing hemorrhages, hard exudates, cotton wool spots, venous beading, and intraretinal microvascular abnormalities. PDR is characterized by neovascularization and can lead to vitreous hemorrhage, tractional retinal detachment, and severe visual loss.

From an image processing viewpoint, the challenge is that the clinically meaningful structures vary in scale, contrast, and spatial location. Microaneurysms are small and subtle, exudates may be bright and irregular, hemorrhages can appear as dark red regions, and advanced vascular abnormalities may alter vessel morphology. These variations explain why a single lesion detector is not sufficient for reliable multi-stage DR classification.

Fundus images introduce additional engineering constraints. Image acquisition can produce blur, sensor noise, non-uniform illumination, color imbalance, and dark circular background areas. Standardization through resizing, normalization, contrast enhancement, and augmentation is therefore necessary before deep learning features are extracted. The thesis emphasizes that preprocessing is not a minor step; it directly affects the visibility of lesions and the generalization of the classifier.

The diagnostic classes used in the study are No DR, Mild NPDR, Moderate NPDR, Severe NPDR, and Proliferative DR. These classes correspond to a severity grading problem rather than a simple binary disease detection problem. Multi-class classification is more useful for clinical triage because mild cases, advanced NPDR, and PDR require different levels of urgency and follow-up.

TABLE I. DATASET PROFILE AND DIABETIC RETINOPATHY SEVERITY CATEGORIES

Item	Description
Image modality	Color retinal fundus image
Input classes	No DR, Mild NPDR, Moderate NPDR, Severe NPDR, PDR
Major lesions	Microaneurysms, hemorrhages, hard exudates, cotton wool spots, venous beading, IRMA, neovascularization
Image	Different illumination, contrast,

conditions	and acquisition variations
Split strategy	70% training, 15% validation, 15% testing
Learning goal	Severity classification from retinal image patterns

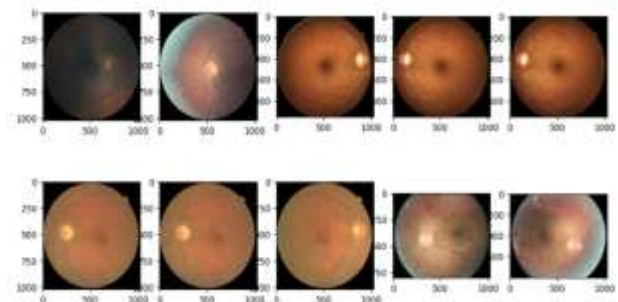


Fig. 1. Representative retinal fundus images before preprocessing.

III. RELATED WORK

A. Machine Learning Methods

Traditional machine learning approaches normally depend on hand-crafted features or segmented lesion descriptors. Studies summarized in the thesis include approaches that combine fundus image analysis with SVM, Random Forest, Decision Tree, and other supervised classifiers. These methods can be effective when lesion segmentation is accurate, but their performance often depends heavily on preprocessing quality and domain-specific feature engineering [8], [9], [14].

A common limitation of purely hand-crafted systems is that they focus on a limited group of early lesions. Models trained to detect only microaneurysms, exudates, or hemorrhages may not capture more severe disease indicators such as venous beading, IRMA, capillary degeneration, diffuse hemorrhages, and neovascularization. This can reduce performance when the task is severity classification rather than isolated lesion detection.

Machine learning classifiers are still valuable in a hybrid framework. Once robust deep features have been extracted, classifiers such as Random Forest and SVM can provide strong performance with lower computational cost than end-to-end retraining of very deep networks. This motivates the thesis design, where deep networks are used for representation learning and machine learning algorithms are used for final decision making.

B. Deep Learning Methods

Deep learning has become the dominant approach for retinal image classification because convolutional layers can learn local edges, vessel patterns, lesion textures, and higher-level disease representations. CNN-based methods have been applied to referable DR detection, lesion localization, and multi-stage severity grading. Reviews and comparative studies show that deep models can reduce manual feature design and improve automated grading performance when enough labeled images are available [1], [6], [10], [11].

GoogleNet is attractive because inception modules process the image with multiple filter sizes. This allows the network to capture small lesions and broader retinal structures in parallel. ResNet is attractive because residual connections reduce the vanishing gradient problem and allow deeper networks to learn stable representations. The thesis combines both architectures so that their strengths become complementary rather than competitive.

Deep learning systems also require careful engineering. Without augmentation, balancing, and validation monitoring, models may overfit benchmark datasets and fail on unseen clinical images. The thesis therefore uses image transformations, hybrid feature fusion, and APSO-based feature selection to improve model robustness and reduce unnecessary feature dimensionality.

C. Transfer Learning and Hybrid Models

Transfer learning has been widely used in medical imaging because annotated clinical datasets are often smaller than general image datasets. Pre-trained networks can learn generic visual structures and then be adapted to retinal images. Studies cited in the thesis include VGG-based, CNN-based, and fusion-based systems for DR, glaucoma, and age-related macular degeneration classification [5], [20]-[22].

The present framework is hybrid in three ways: it fuses GoogleNet and ResNet features, optimizes feature subsets using APSO, and compares multiple classical machine learning classifiers. This modularity supports experimentation and makes the system easier to analyze than a single black-box network.

The literature gap addressed by the thesis is the limited attention given to comprehensive feature engineering for multiple DR stages. Many studies focus on model selection, ensemble averaging, or preprocessing improvements. The proposed approach emphasizes optimized hybrid features, which are

intended to capture a broader range of retinal abnormalities and support better severity classification.

IV. PROPOSED METHODOLOGY

The proposed system follows a sequential pipeline. First, retinal fundus images are collected and labeled according to severity. Second, images are preprocessed to reduce acquisition artifacts and enhance lesion visibility. Third, data augmentation generates additional training variations. Fourth, GoogleNet and ResNet extract deep feature vectors. Fifth, feature fusion combines representations from both networks. Sixth, APSO selects the most informative feature subset. Finally, optimized features are classified by multiple machine learning algorithms and evaluated using standard metrics.

This modular design makes the framework easier to implement, debug, and extend. Each stage can be improved independently: preprocessing can be replaced with stronger contrast enhancement, feature extractors can be replaced with EfficientNet or Vision Transformers, APSO can be compared with other evolutionary methods, and the classifier can be substituted with gradient boosting or neural classifiers.

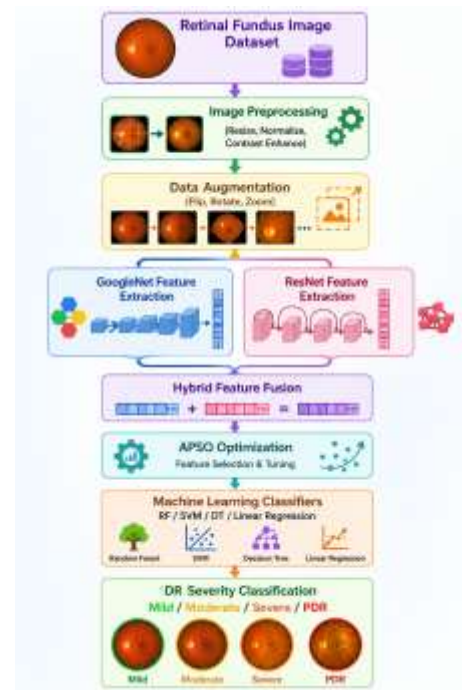


Fig. 2. Overall system architecture for DR severity classification.

A. Image Preprocessing

Preprocessing converts raw fundus images into a standardized and informative input for the feature extractors. The thesis uses resizing to match the 224 x 224 input format required by GoogleNet and ResNet. This reduces computational load and ensures that each

image is compatible with the selected network architecture.

Noise removal is applied to suppress sensor noise and small image distortions. Gaussian or median filtering can smooth irrelevant variations while preserving clinically meaningful structures. Because fundus photographs may have low contrast, contrast enhancement is used to highlight faint lesions and blood vessels. Histogram equalization and Contrast Limited Adaptive Histogram Equalization (CLAHE) are appropriate choices for enhancing local retinal details.

Normalization scales pixel intensity values from the original 0-255 range to a fixed range such as 0-1. This improves numerical stability and makes gradient-based model training more consistent. Green-channel extraction is also useful because retinal vessels and many DR lesions appear with stronger contrast in the green channel than in the red or blue channels. Morphological transformations, including top-hat and bottom-hat operations, further emphasize bright and dark structures related to lesions and vessels.

TABLE II. PREPROCESSING AND AUGMENTATION OPERATIONS USED IN THE FRAMEWORK

Stage	Operation	Purpose
Standardization	Resize to 224 x 224	Match GoogleNet and ResNet input dimensions
Noise control	Gaussian or median filtering	Suppress acquisition noise while retaining retinal structures
Contrast	Histogram equalization / CLAHE	Improve lesion and vessel visibility
Normalization	Scale pixel values to 0-1	Stabilize training and feature extraction
Color emphasis	Green-channel extraction	Enhance vessel and lesion contrast
Morphology	Top-hat / bottom-hat transforms	Highlight bright and dark retinal abnormalities
Augmentation	Flip, rotate, zoom, crop, shear, brightness change	Increase diversity and reduce overfitting

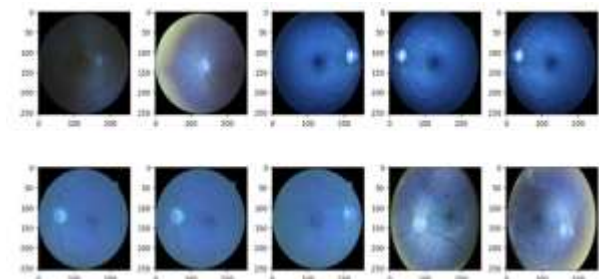


Fig. 3. Retinal fundus images after resizing.

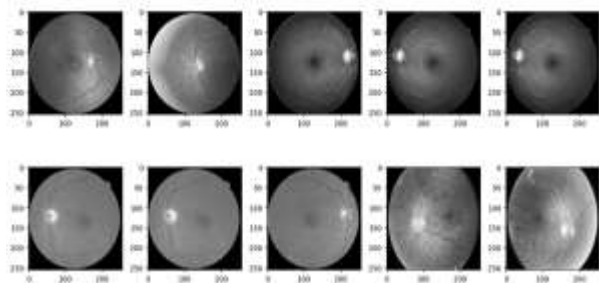


Fig. 4. Green-channel extraction output for retinal images.

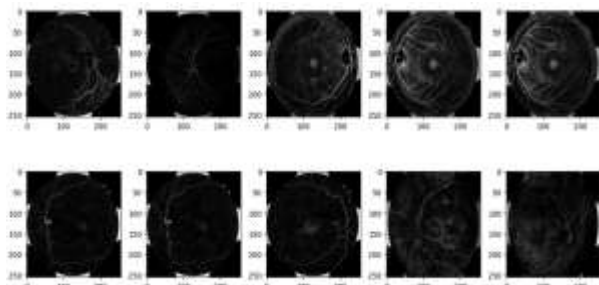


Fig. 5. Morphological transformations used to emphasize retinal structures.

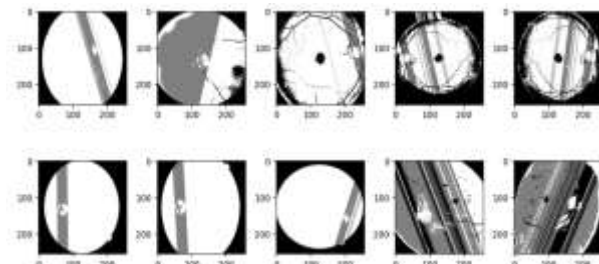


Fig. 6. Data augmentation examples after geometric transformations.

B. Deep Feature Extraction

GoogleNet is used as the first feature extractor. Its inception modules combine 1 x 1, 3 x 3, and 5 x 5 convolutions with pooling operations. The 1 x 1 filters reduce dimensionality and mix channel information, while larger filters detect spatial patterns at different scales. This makes GoogleNet suitable for retinal images because the lesions vary from small microaneurysms to larger hemorrhagic and exudative regions.

ResNet is used as the second feature extractor. A residual block learns a mapping $f(x, W_i)$, and the output can be represented as $y = f(x, W_i) + x$ when input and output dimensions match. The skip connection allows gradient information to flow through deeper layers, reducing vanishing gradient effects and enabling the network to learn more discriminative representations of advanced retinal changes.

The thesis framework combines GoogleNet features and ResNet features to form a hybrid vector. This is based on the assumption that one network may be more sensitive to multi-scale local lesion patterns, while the other may better encode deeper hierarchical structure. Feature fusion therefore increases representation diversity and improves the ability of the downstream classifier to separate DR severity stages.

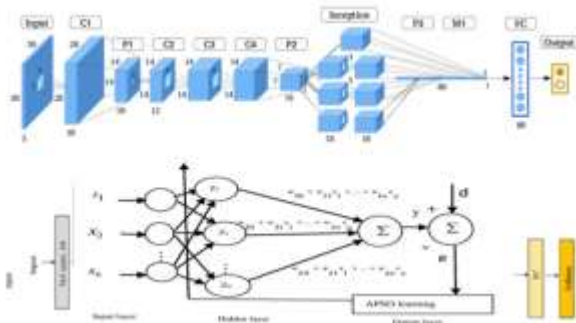


Fig. 7. Hybrid feature extraction model based on GoogleNet, ResNet, and APSO learning.

C. Adaptive Particle Swarm Optimization

After feature fusion, the vector may contain redundant, noisy, or weakly informative dimensions. Feature optimization reduces this problem. APSO is an adaptive version of Particle Swarm Optimization inspired by the movement of particles in a search space. Each particle represents a candidate subset of features. The particle changes its position using its personal best solution and the global best solution found by the swarm.

A simplified velocity update can be written as $v_i(t+1) = wv_i(t) + c_1r_1(p_i - x_i) + c_2r_2(g - x_i)$, where v_i is velocity, x_i is position, p_i is the personal best, g is the global best, c_1 and c_2 are acceleration coefficients, r_1 and r_2 are random values, and w is the inertia weight. In APSO, control parameters are adjusted adaptively to improve convergence and avoid premature trapping in local optima.

For DR classification, the fitness function can be linked to validation accuracy, F1-score, or a weighted objective that balances classification performance with feature subset size. The optimized subset is then passed to the classifiers. This step is important for engineering

efficiency because it can reduce computation while increasing classification reliability.

D. Machine Learning Classifiers

Four classifiers are compared in the thesis. Random Forest uses an ensemble of decision trees and combines their outputs to reduce variance and improve robustness. It is particularly effective for high-dimensional optimized features and performed best in the reported results. SVM separates classes through maximum-margin hyperplanes and is useful for complex decision boundaries. Decision Tree offers simple rule-based interpretation but may overfit when not properly controlled. Linear Regression is used as a simpler predictive baseline and produced lower performance than the nonlinear classifiers.

The comparison is pedagogically useful because it shows that feature extraction alone is not the complete solution. The classifier must match the feature distribution and classification problem. In this study, the strongest result is obtained by the combination of GoogleNet, ResNet, APSO, and Random Forest.

V. SYSTEM DESIGN AND IMPLEMENTATION

The system design contains seven major modules: dataset collection, preprocessing, augmentation, feature extraction, feature optimization, classification, and performance evaluation. Each module receives an input, performs a well-defined operation, and passes its output to the next module. This modular structure supports testing and replacement of components without rewriting the entire system.

The dataset collection module organizes labeled fundus images into severity folders and splits them into training, validation, and testing sets. Proper labeling is essential because the model learns severity-specific visual patterns. A balanced dataset improves training fairness and reduces bias toward majority classes.

The preprocessing module resizes images, removes noise, enhances contrast, normalizes intensity, and prepares green-channel or morphology-enhanced images when required. The augmentation module generates additional examples by flipping, rotating, cropping, zooming, shearing, and changing brightness. These transformations improve generalization without changing the clinical label of the image.

The feature extraction module forwards each processed image through GoogleNet and ResNet. Instead of using only final class predictions, the system extracts feature vectors from the networks. The feature fusion module concatenates or combines these vectors

into a single hybrid representation. APSO then selects the most relevant subset, and the classification module predicts the DR severity class.

The evaluation module compares predictions with ground-truth labels and calculates accuracy, precision, recall, F1-score, sensitivity, specificity, confusion matrix values, and AUC. These metrics are necessary because medical systems should not be evaluated by accuracy alone. False negatives can delay treatment, while false positives can increase unnecessary referrals.

TABLE III. EXPERIMENTAL AND SOFTWARE CONFIGURATION

Parameter	Specification
Programming language	Python
Deep learning framework	TensorFlow / Keras
Machine learning library	Scikit-learn
Image processing	OpenCV
Hardware	GPU or high-performance CPU
Operating system	Windows / Linux
Input image size	224 x 224
Dataset split	70% training, 15% validation, 15% testing
Models	GoogleNet, ResNet, APSO, RF, SVM, DT, Linear Regression

A. Prototype Interface

The thesis also includes a Flask-based web interface for practical demonstration. The prototype contains registration and login screens, an upload page, and a result page where a retinal image can be selected and passed to the trained model. Although the article focuses on the classification framework, the interface shows how an engineering prototype can connect a trained model to a user-facing workflow.

For deployment, the prototype would need user authentication, secure image handling, model versioning, prediction logging, and clinical disclaimers. It should also include explainability outputs such as heatmaps so that clinicians can verify which retinal regions contributed to the prediction.



Fig. 8. Prototype web interface for uploading a fundus image.



Fig. 9. Uploaded retinal image shown in the prototype interface.

B. Detailed Module Architecture

A research article should describe not only the final architecture but also the reason for each module. The thesis implementation is best understood as a chain of transformations that convert a raw color fundus image into a clinically meaningful severity label. The input image is not directly trusted because it may contain noise, illumination imbalance, and background pixels. Each subsequent stage reduces uncertainty or adds discriminative information.

The dataset collection module is responsible for obtaining benchmark retinal images, verifying labels, and organizing images into folders or database records. The quality of this module affects every later result. When images are mislabeled or stored without class balance, a high-performing model may still learn incorrect patterns. The thesis therefore treats dataset structure as an engineering requirement rather than an administrative detail.

The preprocessing module is the first active transformation. It standardizes image size, removes acquisition noise, improves contrast, extracts useful color channels, and prepares images for network input. In medical imaging, preprocessing must preserve clinical meaning. A transformation is useful only if it improves visibility without creating artificial lesions or removing subtle disease indicators.

The augmentation module increases the number and variety of training samples. This is important because medical datasets are often limited and expensive to label. Augmentation does not replace real data, but it helps the model become less sensitive to rotation, scale, brightness, and minor positioning differences. For DR grading, this is acceptable because lesion identity should not change when the image is slightly rotated or cropped within safe limits.

The feature extraction module is the computational core. GoogleNet extracts multi-scale features through inception blocks, while ResNet learns deeper patterns through residual connections. In the proposed framework, each image is converted into two feature vectors. These vectors are then combined so that the final representation contains both local lesion information and deeper retinal structure information.

The optimization module applies APSO to the fused feature vector. This step is similar to selecting the most useful sensors in an engineering measurement system. More sensors do not always improve a decision if some are noisy or redundant. APSO searches for a compact subset of features that improves classification performance while lowering dimensionality.

The classification module receives optimized features and assigns a severity label. The use of several classifiers provides comparative evidence. A single reported accuracy could be misleading; multiple classifiers show whether the feature set is generally discriminative or only works with one specific algorithm. In this thesis, Random Forest is the strongest final classifier.

The performance evaluation module closes the loop. It compares predicted labels to ground truth and produces numerical evidence. Accuracy alone is not enough, especially in medical diagnosis. Sensitivity, specificity, precision, recall, F1-score, confusion matrix behavior, and AUC must all be examined to understand whether the model is safe and useful for screening support.

TABLE IV. MODULE-LEVEL INPUT, PROCESS, AND OUTPUT SPECIFICATION

Module	Input	Processing task	Output
Dataset collection	Fundus image files and labels	Class organization, validation, train/validation/test split	Structured dataset

Preprocessing	Raw fundus image	Resize, denoise, enhance contrast, normalize, green-channel extraction	Standardized image
Augmentation	Training image	Flip, rotate, crop, zoom, shear, brightness adjustment	Expanded training set
GoogleNet feature extraction	Preprocessed image	Inception-based multi-scale convolution	GoogleNet feature vector
ResNet feature extraction	Preprocessed image	Residual-block hierarchical convolution	ResNet feature vector
Hybrid fusion	Two feature vectors	Concatenation or feature-level fusion	Hybrid feature vector
APSO optimization	Hybrid feature vector	Fitness-based subset selection	Optimized feature subset
Classification	Optimized features	RF, SVM, DT, and regression-based prediction	DR severity label
Evaluation	Predictions and ground truth	Metric and confusion-matrix analysis	Performance report

C. Algorithmic Workflow

The following algorithm blocks convert the thesis workflow into implementation-ready steps. They are intentionally written in a compact form so that students can translate them into Python code using OpenCV, TensorFlow/Keras, and Scikit-learn. The main loop should be executed only after the dataset has been checked for class labels and directory consistency.

Algorithm 1 focuses on image preparation. The key engineering principle is repeatability: the same

preprocessing functions must be applied to training, validation, and testing images, except that augmentation should be applied only to the training set. Applying augmentation to validation or testing data can produce misleading performance estimates.

ALGORITHM 1. RETINAL IMAGE PREPROCESSING AND AUGMENTATION

```

Input: labeled retinal fundus image dataset D
Output: processed training, validation, and testing sets
1: for each image I in D do
2:   resize I to 224 x 224 pixels
3:   remove acquisition noise using Gaussian or median filtering
4:   enhance local contrast using histogram equalization or CLAHE
5:   normalize pixel intensity values to the range [0, 1]
6:   optionally extract green channel for lesion-sensitive representation
7: end for
8: split data into training, validation, and testing subsets
9: for each training image do
10:   generate flip/rotate/zoom/crop/shear/brightness samples
11: end for

```

Algorithm 2 describes the hybrid deep feature extraction process. The feature vectors can be extracted from penultimate layers or global average pooling layers depending on the implementation. The important requirement is that the same feature layer is used consistently for all images so that the classifier receives vectors with fixed dimensionality.

The fusion operation may be simple concatenation or a more advanced weighted fusion strategy. The thesis describes hybrid feature fusion as combining GoogleNet and ResNet representations. In practical experiments, concatenation is usually the easiest baseline because it preserves all extracted values before optimization.

ALGORITHM 2. HYBRID GOOGLNET-RESNET FEATURE EXTRACTION

```

Input: processed retinal image set  $X$ 
Output: fused feature matrix  $F$ 
1: load GoogleNet feature extractor  $G$ 
2: load ResNet feature extractor  $R$ 
3: for each image  $x$  in  $X$  do
4:    $fG = G.extract\_features(x)$ 
5:    $fR = R.extract\_features(x)$ 
6:    $f = concatenate(fG, fR)$ 
7:   append  $f$  to  $F$ 
8: end for
9: return  $F$ 

```

Algorithm 3 summarizes APSO-based feature selection. The particle position can be represented as a binary mask where 1 means that a feature is selected and 0 means that it is excluded. The fitness score may be validation accuracy or a weighted combination of accuracy and subset size. A penalty for very large subsets is useful when computational efficiency is important.

After convergence, the best mask is applied to training, validation, and testing features. This avoids a common error: the feature subset must be selected using training and validation data only, and then applied to the untouched test set. Using test labels during optimization would leak information and inflate performance.

ALGORITHM 3. APSO FEATURE SELECTION

Input: fused feature matrix F and labels y
Output: optimized feature subset F^*

- 1: initialize swarm particles as candidate feature masks
- 2: initialize velocity and adaptive control parameters
- 3: repeat until maximum iterations or convergence
- 4: for each particle mask m do
- 5: select features F_m using m
- 6: train temporary classifier on F_m
- 7: compute fitness from validation score and subset size
- 8: update personal best and global best masks
- 9: end for
- 10: update particle velocities and positions adaptively
- 11: end repeat
- 12: return features selected by the global best mask

ALGORITHM 4. CLASSIFICATION AND PERFORMANCE EVALUATION

Input: optimized features F^* and labels y
Output: best classifier and performance report

- 1: train RF, SVM, DT, and linear baseline on training features
- 2: validate models and tune basic parameters
- 3: predict labels for the test set
- 4: compute accuracy, precision, recall, F1, sensitivity, specificity, AUC
- 5: generate confusion matrix for all five DR classes
- 6: compare models and select best classifier
- 7: report results with tables, curves, and interpretation

D. Engineering Design Decisions

The first important design decision is the use of fundus images rather than multimodal clinical data. This makes the system easier to deploy in screening programs because fundus photography is already common in tele-ophthalmology. However, image-only prediction also limits the model because it cannot use patient history, blood glucose level, diabetes duration, or OCT measurements.

The second design decision is the selection of 224×224 input resolution. This value is compatible with many standard CNN architectures and reduces computation. The trade-off is that very small lesions may be harder to detect if resizing removes fine detail. Students can investigate this trade-off by comparing 224×224 , 299×299 , and higher-resolution patch-based methods.

The third design decision is hybrid feature extraction. Rather than relying on a single network, the framework combines GoogleNet and ResNet. The engineering justification is that retinal abnormalities appear at multiple levels of abstraction. Some lesions are small and local; others reflect broader structural disease. A hybrid representation is more likely to encode both types of information.

The fourth design decision is APSO feature optimization. Feature fusion increases dimensionality, and high dimensionality can create redundant or noisy variables. APSO provides a search-based method for selecting useful features. This is similar to dimensionality reduction, but unlike PCA it can directly optimize a classification-related fitness value.

The fifth design decision is the use of Random Forest as the final best classifier. Random Forest is robust, relatively easy to tune, and works well with tabular feature vectors. It also provides some interpretability through feature importance, which can be useful in engineering analysis even if it does not replace clinical explainability methods.

The sixth design decision is the inclusion of a web prototype. A model that runs only in a notebook is not a complete system. A Flask interface demonstrates user interaction, file upload, preprocessing automation, and prediction display. This helps students understand the difference between algorithm development and application deployment.

TABLE V. DESIGN TRADE-OFFS IN THE PROPOSED DR DETECTION SYSTEM

Decision	Benefit	Possible drawback	Student experiment
Image-only input	Simpler deployment and lower data requirement	Ignores clinical variables	Compare image-only and multimodal models
224×224 resizing	Faster training and	May lose tiny lesion detail	Test higher resolution or patches

	standard CNN input		
Google Net + ResNet fusion	Combines multi-scale and deep features	Increases feature dimension	Compare individual and fused features
APSO feature selection	Reduces redundancy and improves accuracy	Adds training-time optimization cost	Compare APSO with GA or PCA
Random Forest classifier	Strong performance and robust behavior	Less end-to-end than neural classifier	Compare RF with XGBoost or MLP
Flask prototype	Demonstrates deployable workflow	Needs security and validation for real use	Add authentication and logging

E. Implementation Notes for Reproducible Experiments

Reproducibility is important for student projects. The dataset split should be saved so that repeated experiments use the same training, validation, and testing images. Random seeds should be fixed for Python, NumPy, TensorFlow, and Scikit-learn where possible. The versions of TensorFlow, Keras, OpenCV, and Scikit-learn should be recorded in the report.

Preprocessing should be implemented as reusable functions rather than manually applied steps. A common source of error is applying one preprocessing pipeline during training and a slightly different pipeline during testing. The safest method is to create a single preprocessing function and call it from both training and prediction scripts.

Model checkpoints should be stored during training. If validation accuracy improves until a certain epoch and then decreases, the best checkpoint should be used rather than the final epoch. The thesis reports stable improvement up to epoch 50, but real experiments may require early stopping when validation loss stops decreasing.

Evaluation should be performed on an untouched test set. Hyperparameter tuning, APSO feature selection, and classifier selection should rely on training and validation data. The test set should be used only once for the final reported performance. This protects the credibility of the article and avoids optimistic estimates.

When reporting results, students should include not only final numbers but also learning curves, confusion matrix, class-wise results, and an explanation of errors. For example, mild and moderate DR can be confused because their visible lesion patterns overlap. Interpreting such errors is more valuable than simply listing accuracy.

All figures and tables should have clear captions and should be cited in the text. In IEEE writing, tables are usually captioned above, while figures are captioned below. Numerical values should be consistent across abstract, results, conclusion, and tables. This article follows that structure.

VI. EVALUATION METRICS

Performance evaluation uses complementary metrics. Accuracy measures the percentage of total correct predictions. Precision measures how many images predicted as a given class are actually correct. Recall, also called sensitivity in medical diagnosis, measures how many true disease cases are detected. F1-score



Fig. 10. Prototype sign-up screen used in the web application.



Fig. 11. Prototype login screen used in the web application.

balances precision and recall and is especially useful when class distributions are not perfectly balanced.

In binary notation, $\text{Accuracy} = (\text{TP} + \text{TN}) / (\text{TP} + \text{TN} + \text{FP} + \text{FN})$, $\text{Precision} = \text{TP} / (\text{TP} + \text{FP})$, $\text{Recall} = \text{TP} / (\text{TP} + \text{FN})$, and $\text{F1-score} = 2 \times \text{Precision} \times \text{Recall} / (\text{Precision} + \text{Recall})$. For multi-class DR grading, these metrics can be calculated class-wise and averaged. Specificity = $\text{TN} / (\text{TN} + \text{FP})$ is also important because it indicates how well the model avoids false positives among normal or non-target cases.

Confusion matrix analysis is used to identify which severity levels are most often confused. In DR, mild and moderate stages may overlap visually because both contain early lesions that can be subtle. AUC summarizes the discrimination capability of the classifier, with higher values indicating stronger separation of disease classes.

TABLE VI. TRAINING AND VALIDATION LEARNING HISTORY

Epoch	Train Acc.	Val. Acc.	Train Loss	Val. Loss
10	88%	85%	0.45	0.52
20	91%	89%	0.33	0.38
30	94%	92%	0.24	0.29
40	95%	93%	0.18	0.22
50	96%	94%	0.12	0.17

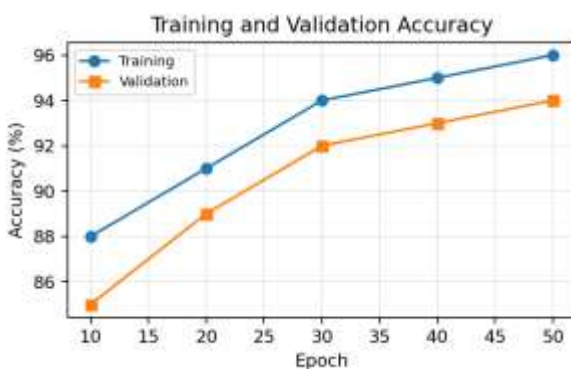


Fig. 12. Training and validation accuracy across epochs.

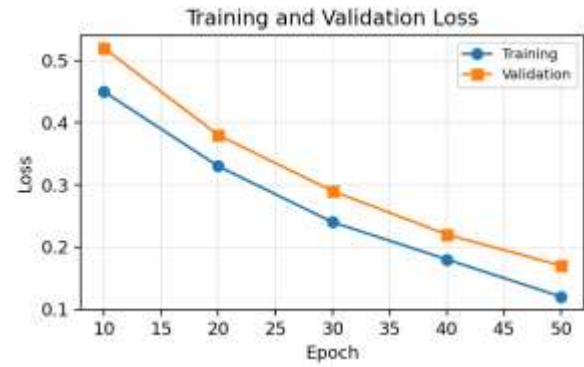


Fig. 13. Training and validation loss across epochs.

VII. RESULTS AND DISCUSSION

The reported results show that the proposed framework learned discriminative retinal representations and generalized well to validation data. Training accuracy increased from 88% at epoch 10 to 96% at epoch 50. Validation accuracy increased from 85% to 94% over the same range. The small gap between training and validation accuracy suggests that augmentation and feature optimization helped control overfitting.

Training loss decreased from 0.45 to 0.12, while validation loss decreased from 0.52 to 0.17. This consistent loss reduction indicates stable convergence. In an engineering implementation, such curves should be monitored continuously because divergence between training and validation behavior can indicate overfitting, underfitting, data leakage, or unstable preprocessing.

The classifier comparison demonstrates that Random Forest achieved the strongest overall results with 94% accuracy, 93% precision, 92% recall, and 93% F1-score. SVM achieved 90% accuracy and also performed strongly. Decision Tree and Linear Regression were less effective, which is expected because standalone trees can be unstable and linear models may not capture nonlinear class boundaries in fused deep feature spaces.

TABLE VII. CLASSIFIER PERFORMANCE COMPARISON

Classifier	Accuracy	Precision	Recall	F1-score
Decision Tree	86%	84%	85%	84%
SVM	90%	89%	88%	89%
Random Forest	94%	93%	92%	93%
Linear Regression	82%	80%	81%	80%

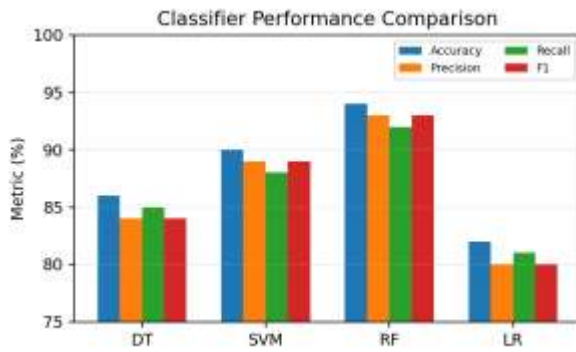


Fig. 14. Bar chart of classifier accuracy, precision, recall, and F1-score.

A. Class-wise Analysis

Class-wise results show that No DR achieved 96% accuracy and PDR achieved 95% accuracy. These classes are often visually more separable because normal images lack pathological lesions, while proliferative disease contains advanced structural abnormalities. Mild DR achieved 92% accuracy, which is slightly lower because early lesions are small, sparse, and sometimes visually similar to normal retinal variations.

The moderate and severe classes achieved 93% and 94% accuracy, respectively. Minor errors between adjacent classes are expected because DR severity is ordinal and lesion characteristics can overlap. From a clinical decision-support perspective, the most concerning errors are those that classify a severe case as normal or mild. The reported confusion matrix indicates that most errors occur between neighboring stages rather than across distant classes.

TABLE VIII. CLASS-WISE ACCURACY RESULTS

DR severity class	Accuracy
No DR	96%
Mild DR	92%
Moderate DR	93%
Severe DR	94%
Proliferative DR	95%

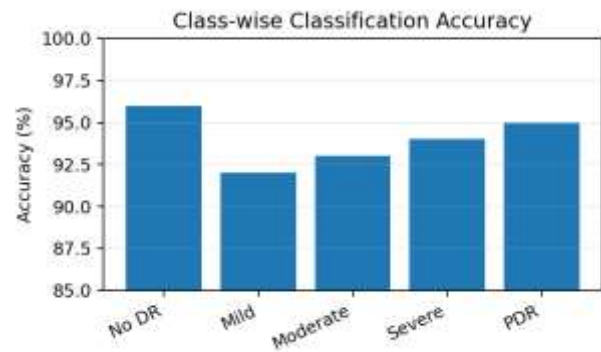


Fig. 15. Class-wise accuracy for five DR severity categories.

TABLE IX. CONFUSION MATRIX FOR DR SEVERITY CLASSIFICATION

Actual / Pred.	No DR	Mild	Moderate	Severe	PDR
No DR	58	1	0	0	0
Mild	2	54	3	0	0
Moderate	1	2	56	1	0
Severe	0	1	2	57	1
PDR	0	0	1	2	58

B. Sensitivity, Specificity, and ROC

Sensitivity and specificity are critical in medical image analysis. The thesis reports 93% sensitivity and 95% specificity. High sensitivity indicates that most DR cases are detected, reducing the risk of missed referrals. High specificity indicates that normal or less severe cases are not excessively overcalled, reducing unnecessary clinical workload.

The ROC analysis also supports the result. Random Forest achieved an AUC score of 0.96, higher than SVM, Decision Tree, and Linear Regression. This indicates strong separation capability when optimized hybrid features are used. The strong AUC is consistent with the accuracy and F1-score results and supports the use of Random Forest as the final classifier in the proposed pipeline.

TABLE X. SENSITIVITY, SPECIFICITY, AND AUC RESULTS

Metric / Classifier	Value
Sensitivity	93%
Specificity	95%
Decision Tree AUC	0.87
SVM AUC	0.91

Random Forest AUC	0.96
Linear Regression AUC	0.83

C. Comparison with Baseline Methods

The proposed method outperformed the baseline models reported in the thesis. Traditional CNN achieved 84% accuracy, GoogleNet achieved 88%, ResNet achieved 90%, and combined GoogleNet plus ResNet achieved 92%. Adding APSO optimization and using Random Forest increased the result to 94%. This trend supports the hypothesis that feature fusion and feature selection both contribute to better classification performance.

The gain from GoogleNet to ResNet and from ResNet to fused features shows the value of complementary representations. The final improvement after APSO indicates that feature dimensionality and redundancy matter. In engineering terms, the model does not simply become better by adding more features; it becomes better by selecting features that contribute to classification while reducing noise and computational burden.

TABLE XI. COMPARATIVE PERFORMANCE WITH EXISTING MODEL VARIANTS

Method	Accuracy
Traditional CNN	84%
GoogleNet	88%
ResNet	90%
GoogleNet + ResNet	92%
Proposed GoogleNet + ResNet + APSO + RF	94%

VIII. ENGINEERING INTERPRETATION FOR M.E. STUDENTS

The project is a useful example of end-to-end engineering design. It begins with a healthcare problem, converts that problem into an image classification task, selects a data representation, designs a computational pipeline, chooses optimization and classification methods, and evaluates performance with measurable criteria. This sequence mirrors the engineering design

cycle: define the problem, model the system, implement the solution, validate results, and identify constraints.

The feature extraction stage illustrates why architecture selection should match the image content. GoogleNet is well suited for multi-scale features, while ResNet supports deeper feature learning. A student can test whether the two networks are complementary by comparing individual and fused feature performance. This encourages experimental thinking rather than blind model selection.

The APSO stage illustrates the role of optimization beyond neural network training. Many students associate optimization only with backpropagation, but feature subset selection is also an optimization problem. APSO searches the feature space for a compact and informative subset, which can improve accuracy, reduce computation, and make classifiers more stable.

The classifier comparison teaches that no algorithm is universally best. Random Forest performed best in this thesis, but the result depends on the features, data split, class balance, and tuning. SVM, Decision Tree, and linear models remain useful baselines because they help quantify the value of the final design. Without baselines, it is difficult to prove that a proposed hybrid framework provides a real improvement.

The implementation prototype demonstrates how an algorithm becomes a usable system. A complete project should not stop at model accuracy; it should consider user login, image upload, preprocessing automation, prediction display, storage, security, and clinical review. For real deployment, the system would also require validation on external datasets, ethical review, regulatory compliance, explainability, and continuous monitoring.

IX. EXTENDED TECHNICAL DISCUSSION

The proposed framework can be examined from four engineering perspectives: signal preparation, representation learning, optimization, and decision making. Signal preparation deals with the quality of the retinal image. Representation learning deals with the transformation of pixels into informative numerical descriptors. Optimization deals with reducing unnecessary information and selecting a compact feature subset. Decision making deals with assigning the final severity label and measuring reliability.

In many student projects, the deep learning model receives most of the attention, but the surrounding stages determine whether the model can be trusted. A noisy training set, an inconsistent preprocessing function, or a poorly chosen validation protocol can

produce a model that looks accurate but fails in practice. The thesis avoids this by describing dataset splitting, image enhancement, augmentation, feature fusion, and multiple evaluation metrics.

A retinal fundus image contains both useful and irrelevant regions. Useful regions include the optic disc, macula, blood vessel network, and lesion-like structures. Irrelevant regions include the black circular border, camera artifacts, and illumination gradients. A good preprocessing pipeline should increase the ratio of useful signal to irrelevant variation. This is why contrast enhancement, green-channel extraction, and morphological operations are included in the workflow.

The deep network is not explicitly told what a microaneurysm or hemorrhage looks like. Instead, it learns filters that respond to patterns in the training images. Lower layers may detect edges, color transitions, and vessel boundaries. Middle layers may detect textures and small lesion-like regions. Higher layers may encode disease-stage patterns. This hierarchical behavior is one reason CNNs outperform hand-crafted feature methods in many retinal image tasks.

GoogleNet and ResNet approach feature learning differently. GoogleNet widens the network through inception modules so that filters of different sizes operate in parallel. ResNet increases depth while maintaining trainability through skip connections. In the hybrid framework, these differences are treated as complementary. The final feature vector should capture both multi-scale local detail and deeper contextual patterns.

Feature fusion must be followed by careful dimensionality control. Concatenating two deep feature vectors may improve representation power, but it can also introduce redundant values. Redundant features can make classifiers slower, increase memory requirements, and sometimes reduce accuracy. APSO addresses this problem by searching for an optimal or near-optimal subset rather than forcing the classifier to use every extracted value.

The Random Forest classifier performed best because ensemble trees can model nonlinear interactions among optimized features. Each tree learns a different partition of the feature space, and the ensemble aggregates their decisions. This reduces the risk that a single unstable tree will dominate the result. In contrast, a linear model may be too simple for the nonlinear structure of retinal disease features.

The SVM result is also important. SVMs often perform well with high-dimensional feature vectors, especially when the classes are separable with appropriate kernels. However, SVM training and parameter selection can be sensitive to kernel choice, scaling, and class imbalance. Random Forest can be easier to tune for a student implementation while still providing strong accuracy.

The confusion matrix gives more insight than a single accuracy number. Errors between Mild and Moderate DR suggest that early disease stages require especially careful feature extraction. These classes contain subtle and partially overlapping signs. Future systems may improve this region by adding lesion localization, attention mechanisms, patch-based high-resolution analysis, or clinical metadata.

The high accuracy for No DR and PDR suggests that the model can identify clear normal and advanced abnormal patterns. This is clinically useful for triage because a screening system should confidently separate obviously normal cases from severe disease. However, mild disease detection remains important because early intervention and monitoring can prevent progression.

A. Preprocessing Rationale in Detail

Image resizing is necessary because standard CNN models require fixed input dimensions. However, resizing is not a neutral operation. If an image is compressed too strongly, tiny lesions may disappear. If it is enlarged without preserving details, interpolation artifacts may be introduced. Bicubic interpolation is often preferred because it produces smoother visual results than nearest-neighbor interpolation, while preserving more structure than overly aggressive smoothing.

Green-channel extraction is useful because retinal vessels and many lesions have higher contrast in the green channel. The red channel may be saturated or less informative in some fundus images, while the blue channel can be noisy. By examining the green channel, the system can emphasize vascular patterns, microaneurysms, and hemorrhagic structures that are important for DR detection.

Top-hat and bottom-hat transformations are morphological operations that highlight structures based on their brightness relative to the surrounding background. A top-hat operation can emphasize bright lesions such as exudates, while a bottom-hat operation can highlight darker structures such as vessels or hemorrhages. These transformations are useful when

lesion contrast is weak or when illumination varies across the image.

Contrast enhancement should be applied carefully. Global histogram equalization can over-enhance noise or create unnatural intensity distributions. CLAHE limits contrast amplification in small regions, making it more suitable for medical images where local lesion visibility is important. Students should visually inspect enhanced images to ensure that preprocessing improves interpretability rather than distorting retinal anatomy.

Normalization improves numerical stability. Deep networks often train better when inputs are scaled consistently. If one image has pixel values in a different range from another, feature activations may become inconsistent. Normalization also supports transfer learning because many pretrained networks expect a specific input scaling convention.

Data augmentation helps the model learn invariance. A lesion remains clinically meaningful if the image is slightly rotated or if brightness changes due to camera settings. However, augmentation should be realistic. Excessive rotation, extreme cropping, or unnatural brightness changes may create images that do not represent real clinical conditions. Engineering judgment is required when selecting augmentation ranges.

B. Feature Learning and Fusion in Detail

A CNN feature map can be interpreted as a response map for learned patterns. Early feature maps may respond to simple edges or color boundaries. Deeper feature maps respond to combinations of patterns that are more specific to disease-related texture and structure. Extracting features from a trained network converts each image into a numerical description suitable for machine learning classifiers.

Inception modules in GoogleNet reduce computational cost by using 1×1 convolutions to compress channel dimensions. This allows larger filters to operate efficiently. For retinal images, this is valuable because small filters capture micro-level lesions while larger filters can capture broader vessel and optic-disc context. Multi-scale processing is therefore aligned with the biological variability of DR lesions.

Residual learning in ResNet is based on the idea that it can be easier to learn a residual correction than a direct mapping. The shortcut connection allows information to bypass one or more layers. This stabilizes training and allows deep networks to achieve strong performance. In DR images, deeper residual features may capture structural combinations that are not visible through shallow texture descriptors alone.

Feature fusion increases the expressive power of the representation. If GoogleNet is more sensitive to small lesion scale and ResNet captures deeper contextual information, the fused vector can support more robust classification. However, fusion also increases dimensionality, which is why optimization is not optional in the proposed design. The system deliberately pairs fusion with APSO to avoid uncontrolled feature growth.

A student implementation can compare early fusion, late fusion, and decision fusion. Early fusion concatenates features before classification. Late fusion trains separate classifiers and combines decisions. Decision fusion averages or votes among model outputs. The thesis uses feature-level fusion because it allows APSO to optimize the combined representation before the final classifier is trained.

Feature extraction can be performed using pretrained networks or networks trained/fine-tuned on retinal images. Pretraining may reduce training time and improve performance when data are limited. Fine-tuning can adapt the filters more strongly to retinal structures. The best strategy depends on dataset size, computing resources, and the similarity between pretraining data and the target medical images.

C. Optimization and Classifier Selection in Detail

APSO is appropriate when the search space is large and a gradient is not available for the selection objective. Feature subset selection is a combinatorial problem because each feature can be included or excluded. Exhaustively testing all subsets is impossible when the feature vector has hundreds or thousands of dimensions. Swarm optimization offers a practical approximation.

The adaptive component improves the basic PSO process by changing parameters during the search. A high exploration tendency is useful early because the swarm needs to examine different areas of the feature space. A stronger exploitation tendency is useful later because the swarm should refine promising subsets. Balancing these phases helps convergence.

The fitness function must reflect the project goal. If the goal is maximum accuracy only, the selected subset may still be large. If the goal includes speed and simplicity, the fitness function can penalize large subsets. For medical screening, recall or sensitivity may deserve extra weight because missing a diseased case can be more serious than sending an additional case for review.

Random Forest, SVM, Decision Tree, and Linear Regression provide a useful spectrum of classifier complexity. Decision Tree is interpretable but can overfit. Random Forest improves stability through ensemble learning. SVM is strong for margin-based separation but may require careful kernel tuning. Linear Regression is simple and useful as a baseline, but it is not ideal for complex nonlinear severity grading.

Hyperparameter tuning affects all classifiers. Random Forest depends on the number of trees, maximum depth, feature sampling, and class weighting. SVM depends on kernel, C value, gamma, and scaling. Decision Tree depends on splitting criteria and pruning. A fair comparison should tune each model within reasonable limits rather than using default values blindly.

Because the thesis reports Random Forest as the strongest classifier, future experiments should test whether this result remains stable across different random splits and external datasets. A single split can be affected by the distribution of easy and difficult images. Cross-validation or repeated holdout evaluation can provide stronger evidence.

TABLE XII. COMMON IMPLEMENTATION ERRORS AND RECOMMENDED CONTROLS

Issue	Effect on results	Recommended control
Data leakage	Inflated performance and poor real-world validity	Separate training, validation, and test sets before optimization
Augmenting test images	Unfair evaluation and optimistic accuracy	Apply augmentation only to training data
Unbalanced classes	Bias toward majority severity class	Use class weights, balanced sampling, or targeted augmentation
Inconsistent preprocessing	Mismatch between training and inference	Use one shared preprocessing function
Overfitting	High training accuracy but weak validation accuracy	Use validation monitoring, dropout, augmentation,

		and early stopping
Large feature vector	Slow classifier and redundant information	Apply APSO or other feature selection methods
Accuracy-only reporting	Missed clinical risk information	Report precision, recall, F1, sensitivity, specificity, AUC, and confusion matrix

X. LIMITATIONS

The reported performance depends on the quality and size of the benchmark dataset. Poor-quality images, blur, illumination artifacts, and camera differences can reduce classification reliability. The dataset may also not fully represent all patient demographics, retinal pigment variations, imaging devices, and clinical environments.

Class imbalance is another limitation. If one severity class contains far fewer images, the model may learn biased decision boundaries. Augmentation can reduce this problem but cannot fully replace real labeled images. Future work should use larger, multi-center datasets and report per-class sensitivity to ensure fairness across all DR stages.

Deep learning systems require computational resources. GoogleNet and ResNet feature extraction can be demanding on low-cost hardware, especially when processing many images. APSO adds optimization cost during training, although it can reduce feature dimensionality during classification. Engineering deployment should therefore consider model compression, batch processing, and hardware acceleration.

The proposed model is a decision-support tool, not a replacement for ophthalmologists. It does not include treatment recommendation, patient history, blood glucose levels, duration of diabetes, OCT data, or other clinical variables. For clinical use, predictions should be reviewed by qualified professionals and validated through regulatory procedures.

XI. FUTURE WORK

Future research can improve the system in several directions. First, larger and more diverse datasets should be collected from multiple hospitals and imaging devices. External validation is necessary to measure generalization. Second, modern architectures such as

EfficientNet, DenseNet, Vision Transformers, attention-based CNNs, and hybrid CNN-transformer models can be compared with GoogleNet and ResNet.

Third, multimodal learning can be investigated by combining fundus images with OCT, blood glucose level, diabetes duration, patient age, and other clinical variables. Such fusion could improve risk prediction beyond image-only severity classification. Fourth, explainable AI techniques such as Grad-CAM and saliency maps should be integrated to show the retinal regions responsible for predictions.

Fifth, optimization algorithms such as Genetic Algorithms, Ant Colony Optimization, Bayesian optimization, and reinforcement learning can be compared with APSO. Sixth, deployment work should focus on mobile screening, telemedicine platforms, secure databases, clinician dashboards, and model monitoring after release. Finally, the framework can be extended to other retinal diseases such as glaucoma, age-related macular degeneration, cataract-related retinal changes, and hypertensive retinopathy.

XII. CONCLUSION

This research paper presented an efficient hybrid framework for diabetic retinopathy severity classification using retinal fundus images. The proposed system integrates preprocessing, augmentation, GoogleNet and ResNet feature extraction, hybrid feature fusion, APSO feature selection, and machine learning classification.

The experimental results reported in the thesis show that Random Forest performed best on APSO-optimized hybrid features, achieving 94% accuracy, 93% precision, 92% recall, and 93% F1-score. Class-wise results and confusion matrix analysis indicate strong discrimination across No DR, Mild DR, Moderate DR, Severe DR, and PDR classes, with most errors occurring between neighboring severity levels.

The study demonstrates that combining deep learning representation learning with evolutionary optimization and classical machine learning can improve automated DR screening performance.

REFERENCES

- [1] W. L. Alyoubi, W. M. Shalash, and M. F. Abulkhair, "Diabetic retinopathy detection through deep learning techniques: A review," *Informatics in Medicine Unlocked*, vol. 20, 2020, Art. no. 100377.
- [2] R. Rajalakshmi, R. Subashini, R. M. Anjana, and V. Mohan, "Automated diabetic retinopathy detection in smartphone-based fundus photography using artificial intelligence," *Eye*, vol. 32, no. 6, pp. 1138-1144, Jun. 2018.
- [3] World Health Organization, "Diabetes," 2022. [Online]. Available: WHO diabetes health topic.
- [4] S. Bhandari, S. Pathak, and S. A. Jain, "A literature review of early-stage diabetic retinopathy detection using deep learning and evolutionary computing techniques," *Archives of Computational Methods in Engineering*, vol. 30, no. 2, pp. 799-810, Mar. 2023.
- [5] A. K. Gangwar and V. Ravi, "Diabetic retinopathy detection using transfer learning and deep learning," in *Evolution in Computational Intelligence: Frontiers in Intelligent Computing: Theory and Applications*, vol. 1. Cham, Switzerland: Springer, 2020, pp. 679-689.
- [6] Z. Khan et al., "Diabetic retinopathy detection using VGG-NIN a deep learning architecture," *IEEE Access*, vol. 9, pp. 61408-61416, 2021.
- [7] E. Ozbay, "An active deep learning method for diabetic retinopathy detection in segmented fundus images using artificial bee colony algorithm," *Artificial Intelligence Review*, vol. 56, no. 4, pp. 3291-3318, Apr. 2023.
- [8] A. Panwar, G. Semwal, S. Goel, and S. Gupta, "Stratification of the lesions in color fundus images of diabetic retinopathy patients using deep learning models and machine learning classifiers," in *Proc. ADCOM*. Cham, Switzerland: Springer, 2020, pp. 653-666.
- [9] G. Parthasharathi, R. Premnivas, and K. Jasmine, "Diabetic retinopathy detection using machine learning," *Journal of Innovative Image Processing*, vol. 4, no. 1, pp. 26-33, 2022.
- [10] N. Tsiknakis et al., "Deep learning for diabetic retinopathy detection and classification based on fundus images: A review," *Computers in Biology and Medicine*, vol. 135, Jan. 2021, Art. no. 104599.
- [11] R. Pires, S. Avila, J. Wainer, E. Valle, M. D. Abramoff, and A. Rocha, "A data-driven approach to referable diabetic retinopathy detection," *Artificial Intelligence in Medicine*, vol. 96, pp. 93-106, May 2019.
- [12] Y. Yang, T. Li, W. Li, H. Wu, W. Fan, and W. Zhang, "Lesion detection and grading of diabetic retinopathy via two-stage deep convolutional neural networks," in *Proc. Int. Conf. Medical Image Computing and Computer-Assisted Intervention*. Cham, Switzerland: Springer, 2017, pp. 533-540.
- [13] P. Saranya and S. Prabakaran, "Automatic detection of non-proliferative diabetic retinopathy in retinal fundus images using convolution neural

- network," *Journal of Ambient Intelligence and Humanized Computing*, pp. 1-10, Sep. 2020.
- [14] T. M. Usman, Y. K. Saheed, D. Ignace, and A. Nsang, "Diabetic retinopathy detection using principal component analysis multi-label feature extraction and classification," *International Journal of Cognitive Computing in Engineering*, vol. 4, pp. 78-88, Jun. 2023.
- [15] M. Mujahid et al., "Pneumonia classification from X-ray images with Inception-V3 and convolutional neural network," *Diagnostics*, vol. 12, no. 5, p. 1280, May 2022.
- [16] I. Shafi et al., "An effective method for lung cancer diagnosis from CT scan using deep learning-based support vector network," *Cancers*, vol. 14, no. 21, p. 5457, Nov. 2022.
- [17] A. Fatima et al., "Deep learning-based multiclass instance segmentation for dental lesion detection," *Healthcare*, vol. 11, no. 3, p. 347, Jan. 2023.
- [18] V. Gulshan et al., "Performance of a deep-learning algorithm vs manual grading for detecting diabetic retinopathy in India," *JAMA Ophthalmology*, vol. 137, no. 9, p. 987, Sep. 2019.
- [19] X. Li, L. Shen, M. Shen, F. Tan, and C. S. Qiu, "Deep learning based early stage diabetic retinopathy detection using optical coherence tomography," *Neurocomputing*, vol. 369, pp. 134-144, Dec. 2019.
- [20] P. Saranya, K. M. Umamaheswari, M. Sivaram, C. Jain, and D. Bagchi, "Classification of different stages of diabetic retinopathy using convolutional neural networks," in *Proc. 2nd Int. Conf. Computing, Automation and Knowledge Management (ICCAKM)*, Jan. 2021, pp. 59-64.
- [21] M. Karakaya and R. E. Hacisoftoglu, "Comparison of smartphone-based retinal imaging systems for diabetic retinopathy detection using deep learning," *BMC Bioinformatics*, vol. 21, no. S4, pp. 1-18, Jul. 2020.
- [22] G. T. Zago, R. V. Andreao, B. Dorizzi, and E. O. T. Salles, "Diabetic retinopathy detection using red lesion localization and convolutional neural networks," *Computers in Biology and Medicine*, vol. 116, Jan. 2020, Art. no. 103537.
- [23] Kaggle, "Diabetic Retinopathy Detection: Identify signs of diabetic retinopathy in eye images," 2015. [Online]. Available: Kaggle competition dataset.
- [24] O. Perdomo, J. Arevalo, and F. A. Gonzalez, "Convolutional network to detect exudates in eye fundus images of diabetic subjects," *Proc. SPIE*, vol. 10160, pp. 235-240, Jan. 2017.