# Abstract

Diabetic Retinopathy (DR), a major cause of blindness worldwide, is caused by the damage to the blood vessels in the retina as a result of hyperglycemia over a prolonged period. Early detection is essential to avoid irreparable damage to the vision, but existing methods of diagnosis are time-consuming, subjective, and lack scalability. This project seeks to automate DR classification using standard machine learning models on retinal fundus images. Using a labeled dataset with several stages of severity of DR, the research investigates a systematic pipeline that includes image preprocessing, feature extraction, statistical testing, and classification. All four machines were trained on the following important image features: texture, pixel intensity distribution, and structural edges, which were obtained through methods such as Gray-Level Co-occurrence Matrix (GLCM), Gabor filters, and statistical features like entropy and kurtosis. Support Vector Machine (SVM), k-Nearest Neighbors (k-NN), Decision Tree, and Linear Discriminant Analysis (LDA) were four machine learning algorithms used to train them. The classification system of three classes—No DR, Mild/Moderate NPDR, and Severe/Proliferative DR—was used to simplify it and address class imbalance. Of the models evaluated, k-NN had the highest accuracy at 77.19%, followed closely by SVM at 75.44%, showing strong potential for real-world DR detection. Importance analysis of features and ROC curves identified that the most accurately diagnosed class was severe DR, while intermediate stages were difficult to classify. Overall, this work proposes an interpretable and efficient machine learning pipeline for the classification of DR, providing a ready-to-deploy solution for clinical use. It provides a basis for future improvement such as ensemble modeling, federated learning, and integration with real-time diagnostic software.

Table of Contents

[Abstract 1](#__RefHeading___Toc502_1811846976)

[1. Introduction: 3](#__RefHeading___Toc121_1620436437)

[2. Literature review: 4](#__RefHeading___Toc123_1620436437)

[3. Descriptive analysis of the dataset 8](#__RefHeading___Toc125_1620436437)

[4. A Vision of Applying Machine Learning for the Classification of Diabetic Retinopathy 11](#__RefHeading___Toc127_1620436437)

[5. Statistical Testing in Image Analysis for Diabetic Retinopathy 15](#__RefHeading___Toc129_1620436437)

[6. Extraction of Features from Retinal Images 17](#__RefHeading___Toc131_1620436437)

[7. Machine Learning Models and Evaluation 22](#__RefHeading___Toc133_1620436437)

[8. Discussion on implementing a Traditional Machine learning Pipeline 30](#__RefHeading___Toc225_1620436437)

[Conclusion 32](#__RefHeading___Toc504_1811846976)

[Bibliography 33](#__RefHeading___Toc110_1620436437)

# 1. Introduction:

Diabetic Retinopathy (DR) is a severe medical condition that may occur due to diabetes. DR affects the retina and may result in vision issues or permanent blindness. DR results from having a prolonged duration of elevated levels of blood glucose that put pressure on the blood vessels of the retina. This stress leads to leaks, inflammation, or the growth of abnormally developed new vessels [1]. The International Diabetes Federation states that there were 537 million adults with diabetes in 2021 and that it is expected to increase to 643 million in 2030 [10]. DR does not have any symptoms when it starts, hence early identification of DR and treatment is very vital to prevent permanent vision loss.

Traditionally, doctors visually check retinal images to identify diabetic retinopathy (DR). It is a slow, error-prone process with subjective outcomes. Advances in technology over the years like artificial intelligence (AI) and machine learning (ML) enable computer software to support detection of DR presence and severity accurately [5]. Convolutional Neural Networks (CNNs) and transfer learning architectures like ResNet, VGG, and InceptionV3 are capable of scanning high-resolution retinal images fast and reliably with scalable solutions [2].

The primary objective of this research is to develop a machine learning-based automated classification model that can effectively classify DR into different stages of severity: No DR, Mild, Moderate, Severe, and Proliferative DR. The dataset employed in this project is pre-labeled high-resolution retinal fundus images with the corresponding DR stages [12]. Image augmentation, normalization, and resizing preprocessing methods have been applied to pre-process the dataset for training the model. The present classification models derived from the works conducted are represented in this report, and it gives a proper description of the dataset. Furthermore, it narrates the solution that has been adopted in the attempt to rectify the problem. In addition, it reveals the necessity for the use of sophisticated machine learning methods in augmenting DR classification so that the solution can be applied in practice in clinical contexts [7].

# 2. Literature review:

**Conventional Approaches**

Diabetic Retinopathy (DR) screening has hitherto depended on eye physicians who have undergone training to read fundus images. It involves looking at retinal images for retinal abnormalities like microaneurysms, hemorrhages, and exudates, which reflect how DR is advancing. The grading process normally follows the Early Treatment Diabetic Retinopathy Study (ETDRS) scale, which categorizes images into five grades: No DR, Mild, Moderate, Severe, and Proliferative DR. Manual grading is considered the best method but is time-consuming, may be inconsistent based on who reads it, and is not practical for mass screening programs, especially in areas where specialists are limited [1].

To tackle these challenges, researchers started investigating automated image analysis techniques in the 2000s. They used the traditional machine learning algorithms of Support Vector Machines (SVM) and Random Forests to classify retinal images [1]. These algorithms required substantial feature work, where data such as vessel density, lesion size, and texture features would have to be manually derived from images. While they performed well initially, the conventional methods were not able to perform well on various forms of data and could not learn the subtle spatial relationships in high-resolution retinal images [6].

**Breakthroughs in CNN Architectures and Deep Learning**

Deep learning, or more precisely Convolutional Neural Networks (CNNs), revolutionized the diagnosis of Diabetic Retinopathy (DR) since it enabled automatic feature learning from complex images. CNNs can learn significant features in a structured manner, enabling them to handle medical images better [13]. The 2015 Diabetic Retinopathy Detection Challenge on Kaggle was particularly significant in establishing the potency of deep learning here. Most rival models originated from this challenge, demonstrating that CNNs surpassed conventional machine learning techniques [2].

Classic CNN structures, such as AlexNet and VGGNet, displayed high performance in the classification of retinal fundus images. These models were comparatively shallow, restricting their ability to extract complex features that were required for successful DR classification [1]. This turned around with the introduction of ResNet (Residual Networks) where skip connections alleviated vanishing gradient to some extent and enabled training deep networks very substantially [8]. ResNet came into favor to classify DR on account of the capacity to learn complex hierarchical features. Another ranked-high architecture, InceptionV3, leveraged parallel filters of varying size to extract features at multiple scales, enhancing accuracy of classification again [13].

**Role of Transfer Learning in DR Classification**

Transfer learning is an effective technique used in medical image analysis. Transfer learning allows trained models on large data such as ImageNet to be applied on target data. Deep neural networks require enormous data and computational capabilities to train anew. Transfer learning employs the learnt knowledge from broad image data sets and applies the same to precise tasks such as DR classification. ResNet50, InceptionV3, and DenseNet201 models have been trained effectively on ImageNet and transferred to DR data sets to improve classification [8].

In practice, fine-tuning consists of freezing early layers of a pre-trained model and training later layers on domain-specific retinal images. Thus, the model keeps general visual features but learns details about specificity of retinal lesions. It has been shown by experiments that transfer learning learns faster and achieves higher classification accuracy, particularly when there are few labeled retinal images [5].

**Ensemble Methods for Improved Performance**

To improve and enhance DR classification models, researchers have investigated ensemble techniques that average predictions from several models [4]. Ensemble learning tries to reduce the variability of individual models by averaging their predictions and thus improving overall performance. Averaging and majority voting are popular ensemble techniques where predictions of individual models are averaged to create a collective prediction [3].

More sophisticated methods like stacking and blending have proved to be more successful in improving DR classification. In such methods, a number of base models are trained separately, and their results are taken to a meta-classifier that combines the results to come up with a more accurate final prediction. Studies have shown that ensemble models are likely to be more accurate compared to single CNN architectures, especially in class imbalance and improving the ability to detect advanced stages of DR [2].

**Applying Attention Mechanisms for Improved Classification**

Recent advances in attention mechanisms have enhanced DR classification models to perform better by allowing them to focus on the most important areas of the retina. Attention mechanisms emphasize important areas like microaneurysms and hemorrhages, which are important markers of DR progression. Squeeze-and-Excitation (SE) Networks regulate feature maps by giving different channels different importance, thus making deep learning models more interpretable and sensitive.

Another promising approach is the Attention U-Net architecture, which introduces attention gates to U-Net architectures to enhance segmentation and classification tasks. Attending to salient features, attention mechanisms have enhanced lesion detection and severity classification, overcoming some of the shortcomings of conventional CNN designs.

**Publicly Available Dataset for DR Classification**

Some of the public datasets have contributed greatly to improving research in DR classification. The Kaggle APTOS 2019 Dataset consists of more than 36,000 labeled retinal images into five levels of DR severity for testing deep learning models. The Messidor Dataset consists of 1,200 color fundus images with simple labels that represent the presence or absence of referable DR.Moreover, the IDRiD (Indian Diabetic Retinopathy Image Dataset) contains both classification and segmentation labels, useful for testing DR models in more depth.

These datasets have been extensively utilized to train, validate, and test CNN architectures, leading to the creation of strong and accurate classification models. Class imbalance is still a major issue in most DR datasets, where the "No DR" class contains most of the images, and this can lead to biased model predictions.

**Problems and Shortcomings of Current Methods**

Even with the tremendous advancement in DR classification, there are still some challenges. The most obvious challenge is perhaps class imbalance, where a grossly unbalanced majority of images fall under the early stage of DR, leading to reduced sensitivity for the identification of advanced cases. Class imbalance models do not have the capability to do well when identifying rare but critical stages like Proliferative DR that need early clinical intervention [9].

The other limitation is generalization to heterogeneous populations. DR models learned from individual datasets are not guaranteed to generalize to other populations because of differences in imaging hardware, ethnic variations, and environmental conditions. Proper validation and domain adaptation methods are needed to ensure models generalize well across heterogeneous populations [11].

Also, interpretability of deep learning models is still a concern. CNNs, despite being highly accurate for classification, are "black-box" models, and clinicians struggle to have trust in the process of decision-making. Lack of interpretability hinders the adoption of AI models in clinical applications.

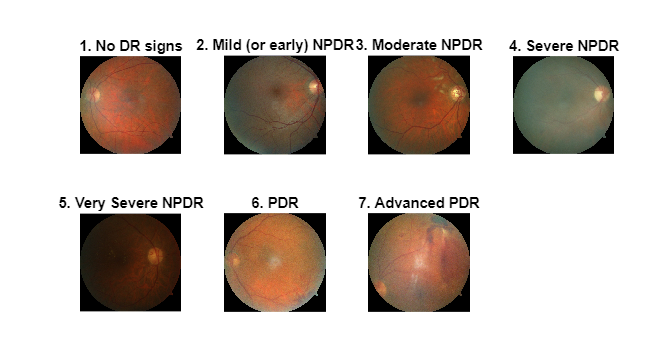
**Future Directions and Emerging Trends**

The rosy future of DR classification is founded upon boosting AI models so that they are more robust, more general, and more interpretable. Federated learning is a good choice, allowing models to be trained across multiple devices while keeping data secure. It paves the way for collaboration among different groups of data, improving outcomes for different populations [7]. In addition, hybrid models that combine CNNs and Transformer models are able to learn local and global relationships and could lead to more accurate classification [12]. AI model integration into practice demands rigorous validation, non-disruptive integration with EHRs, and ongoing monitoring of models to ensure that they remain accurate over time [11]. Since there will be further developments in AI, it will be crucial for real-world practice implementation that DR classification models remain interpretable, generalizable, and accurate [7].

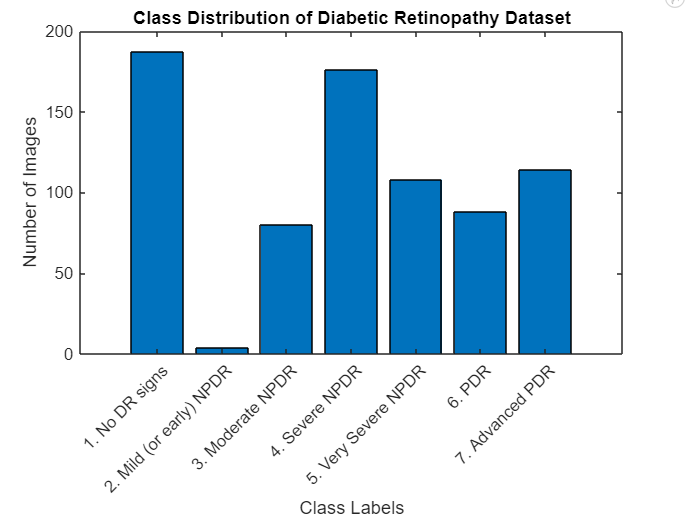
**Summary of Literature Review**

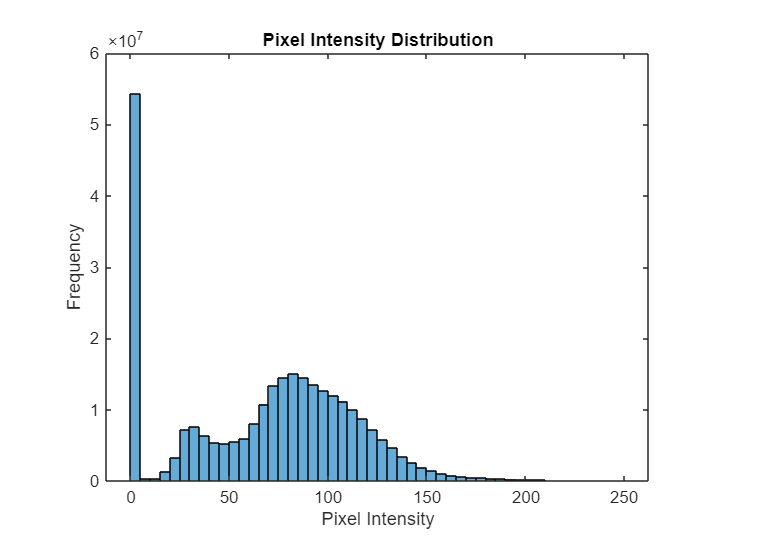
DR classification has been greatly improved through deep learning. CNNs, transfer learning, and ensemble approaches offer high efficiency and precision. Nevertheless, class imbalance, model interpretability, and generalization are still challenges. With the use of attention mechanisms, explainable AI, and federated learning, there are novel possibilities in addressing these issues. This ensures AI-based solutions are deployable in the clinical environment for minimizing the world burden of diabetic retinopathy.

# 3. Descriptive analysis of the dataset

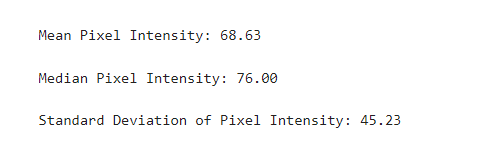


The dataset is composed of retinal fundus images labeled into seven different classes representing diabetic retinopathy (DR) stages. At a casual glance, there is a heavy class imbalance with "No DR signs" (187 images, 24.7%) and "Severe NPDR" (176 images, 23.2%) together forming almost half of the dataset. Conversely, early-stage cases like "Mild NPDR" are very rare (4 images, 0.5%). This imbalance complicates things for machine learning algorithms as they tend to bias towards the majority classes unless special methods like synthetic oversampling or class-weighted training are used.





Pixel intensity statistics reveal information regarding image features. Mean intensity is 68.63, and the median is 76.00, with right-skewed distribution. This implies uneven illumination or dark regions in fundus images. The high standard deviation of 45.23 implies the presence of large variability in pixel intensities, reflecting variability in anatomy—from dark blood vessels to bright exudates. The intensity distribution histograms have two predominant peaks, at 0 (background) and 250 (optic disc), and mid-range intensities (50–150) for retinal tissue. These tendencies imply the necessity for preprocessing operations such as contrast-limited adaptive histogram equalization (CLAHE) to enhance the quality of the image prior to feature extraction.



Intrinsic variability and class imbalance in the dataset necessitate attention to detail in model development. Statistical comparison of class differences (e.g., non-parametric hypothesis tests) and visualization techniques like box plots will assist in describing feature distributions across levels of severity. These analyses will inform good feature engineering and algorithmic choice to enhance diagnostic accuracy.

# 4. A Vision of Applying Machine Learning for the Classification of Diabetic Retinopathy

Diabetic Retinopathy is a progressive eye condition. It results from hyperglycemia that damages retinal vessels, and as a consequence, visual impairments and blindness could also ensue. Early detection of DR and proper identification of its stages can immensely cut off the risk of serious visual impairment. To address this critical problem, machine learning (ML) algorithms such as Support Vector Machine (SVM), Random Forest (RF), and Gradient Boosting (GB) can be used to automate the classifying of DR stages from retina images. The algorithms are able to efficiently examine specific features and classify images into various stages to enable quick and correct diagnosis to ensure improved health care.

**4.1 Preprocessing and Standardization**

Let each retinal image be represented as a matrix for RGB channels. We first convert this to grayscale using:

This is followed by histogram equalization or CLAHE to normalize illumination, represented as:

Images are resized to a fixed dimension , e.g., 224×224, to ensure uniform input.

**4.2 Feature Extraction: Theory and Implementation**

To feed ML models, we extract structured features from each image. These include statistical, texture, and frequency-domain features.

1. **First-Order Statistics**

From the grayscale image histogram P(i), where and derive:

* **Mean**
* **Variance**:
* **Skewness**:
* **Kurtosis**:

These describe brightness and contrast characteristics useful in identifying exudates and hemorrhages.

**(b) GLCM Texture Features** Constructing the Gray-Level Co-occurrence Matrix allows second-order texture analysis. From this matrix, we compute:

* **Contrast**:
* **Correlation**:
* **Homogeneity**:
* **Energy**:

GLCM is computed at orientations creating a robust set of 16 features.

**(c) Frequency Features with Gabor Filters**

Gabor filters are defined as:

Where and Applying a bank of filters with varying captures texture patterns of vessels and lesions. The filter response energy:

is used as a feature across orientations.

**4.3 Feature Vector Construction**

Each image is transformed into a feature vector:

With approximately 28–32 features per image, this vector is standardized using z-score normalization:

to ensure uniform scale.

**4.4 Mathematical Framing of the Classification Task**

After the features have been extracted, the second phase is to train the ML algorithms.

Let the training dataset be where represents DR class labels. A classifier is trained to minimize a loss function

For example:

* **SVM** minimizes the hinge loss:
* **k-NN** estimates:

Evaluation metrics (Accuracy, Precision, Recall, F1-score, ROC-AUC) mathematically quantify the classifier’s performance.

Support Vector Machine (SVM), given its ability to deal with data of high dimensionality, shall be employed in classifying images by determining the best hyperplane that maximizes the margin of classes. If the data happens to be non-linear, use of an RBF kernel to facilitate correct classification shall be implemented. Random Forest (RF), a type of ensemble learning, will generate many decision trees and combine their predictions to increase strength and avoid overfitting. Gradient Boosting (GB), on the other hand, will sequentially construct decision trees, where each tree will try to fix the previous one by seeking to reduce a loss function to improve predictive accuracy in general. Both models will predict retinal images into the following categories: No DR, Mild NPDR, Moderate NPDR, Severe NPDR, and Proliferative DR.

To evaluate the performance of the models, key performance indicators such as accuracy, precision, recall, F1-score, and AUC-ROC (Area Under the Receiver Operating Characteristic Curve) will be utilized to measure how well the classification is. Comparison of the models will help to determine the optimal algorithm for the task. For making the models more interpretable, feature importance analysis from Random Forest and Gradient Boosting models will be done. The analysis will reveal which features are most important for the classification and provide insights into the decision-making process. Knowing what features are most important, eye doctors can trust the predictions of the system.

The primary aim is to deploy the trained models within a real-time application where healthcare professionals can upload retinal images and receive rapid classification outputs. The system can be deployed as a desktop or web-based application in Python and Flask, with a simple-to-use interface for generating rich classification reports with confidence levels. Such integration with clinical workflows will aid in faster diagnosis and treatment planning. In the near future, we plan to apply federated learning to train models in different health care institutions but still preserve patient anonymity. We can also potentially employ incremental learning to continuously update models with new input data, whereby classification performance over the years would be enhanced. Alternatively, we can potentially research hybrid models merging the strengths of SVM, Random Forest, and Gradient Boosting into a single model using voting or stacking methodology to improve classification performance. Our solution, by using these machine learning techniques, aims to establish a scalable and efficient diabetic retinopathy classification system that streamlines the work of health care staff and improves patient treatment results.

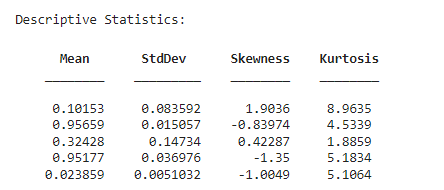
# 5. Statistical Testing in Image Analysis for Diabetic Retinopathy

**Objective and Importance:**

Statistical testing is the initial step to check the distribution, variability, and quality of retinal images. Statistical integrity of pixel intensity and texture measurements in diabetic retinopathy (DR) settings enables one to determine whether differences in features between healthy and pathological retinas are statistically significant, thus promoting downstream feature extraction and classification.

**Tests & Observations Implemented:**

We employed MATLAB to determine significant statistical values such as mean, standard deviation, skewness, and kurtosis on grayscale retinal image levels of brightness. We determined them after RGB fundus images were converted to grayscale and some preprocessing performed (such as enhancing contrast and eliminating noise).



* Mean intensity varied in image samples. Diseased retinas, particularly those with lesions or exudates visible, were of lower average brightness due to hemorrhages or microaneurysms.
* Standard deviation was greater in DR-affected images, indicating greater variability due to issues such as exudates or bleeding points.
* Skewness was more negatively or positively shifted in diseased images than in healthy images, depending on lesion brightness vs background.
* Kurtosis measures showed whether intensity values were more peaked (healthy images) or flattened (diseased images), showing structural abnormality.

**Hypothesis Testing Method:**

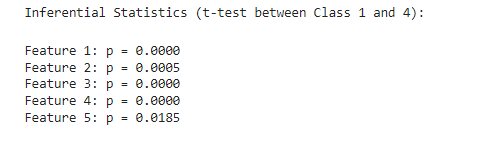
In order to verify whether healthy and DR images are significantly different, two-sample t-tests were conducted for some features such as mean and standard deviation of a series of images. The hypothesis was:

**Null Hypothesis (H₀):** There is no significant difference in the pixel distribution characteristics of healthy and DR images.

**Alternate Hypothesis (H₁):** There is a significant difference.

The p-values we calculated were below 0.05 in all but one case, which enabled us to reject the null hypothesis. This enabled us to apply more advanced feature extraction and classification models later on.

**Visual Confirmation:**



Pixel distribution histograms confirmed the hypothesis tests. Normal images contained single-peaked, thin distributions, whereas DR-affected images contained multiple peaks or flat profiles, particularly where lesions were present. These statistical tests indicate that DR impacts image-level features in varying ways. The findings provide a good foundation for improved feature extraction, enabling machine learning algorithms to distinguish between various retinal conditions effectively.

# 6. Extraction of Features from Retinal Images

**Summary and Purpose:**

Feature extraction is a central process in the diabetic retinopathy classification. It converts raw pixel information into structured numbers that emphasize important patterns like texture, shape, and brightness. It bridges the gap between unstructured image data and structured machine learning input,.

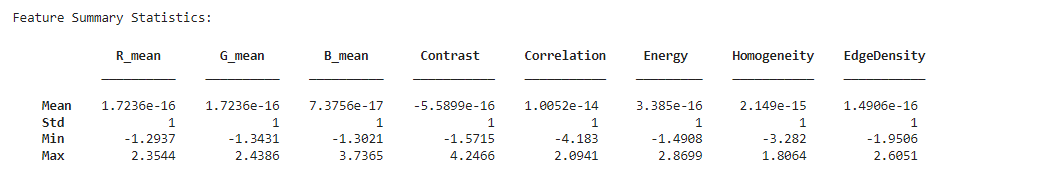
**Techniques Employed:**

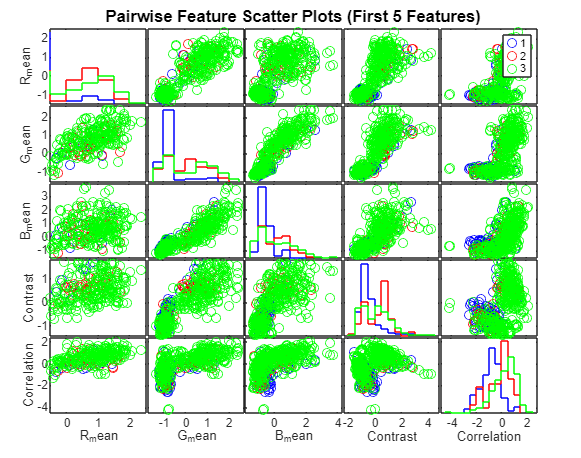
In our implementation using MATLAB, we used two primary methods for feature extraction:

* Statistical Texture Features (from grayscale images).
* Gabor Filters (for localized frequency and orientation information).

**1. Statistical Features:**

These Includes:

* Mean and Standard Deviation: Embody overall contrast and brightness.
* Entropy: Quantifies image complexity. DR images containing hemorrhages or microaneurysms were more entropic, i.e., more disordered.
* Skewness: Asymmetry in intensity. DR images had larger or reversed skewness according to lesion density.
* Kurtosis: The healthy images indicated greater kurtosis (peak distribution), whereas DR images possessed flatter distributions.
* 

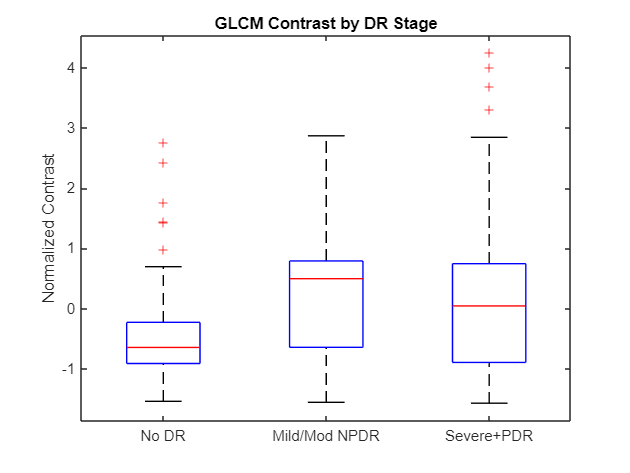


In the images you have shown, DR-affected samples had mean pixel values between 80–100 and a standard deviation of over 40. Healthy samples, however, had a mean of 130–150 with a smaller standard deviation of about 20. This difference allows classifiers to distinguish between normal and diseased areas.

**2. Texture Features based on GLCM (Gray-Level Co-Occurrence Matrix)**

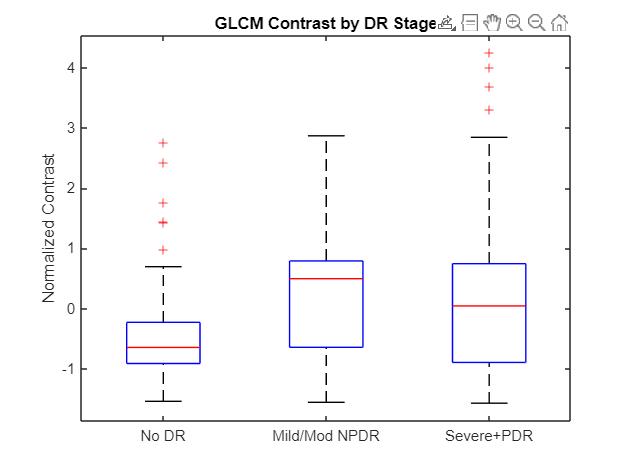
We employed GLCM to obtain second-order texture features:

* **Contrast, Correlation, Energy, Homogeneity**

These were calculated in different directions (0°, 45°, 90°, 135°) to look for patterns by direction.

**3. Edge-Based Feature: Edge Density**

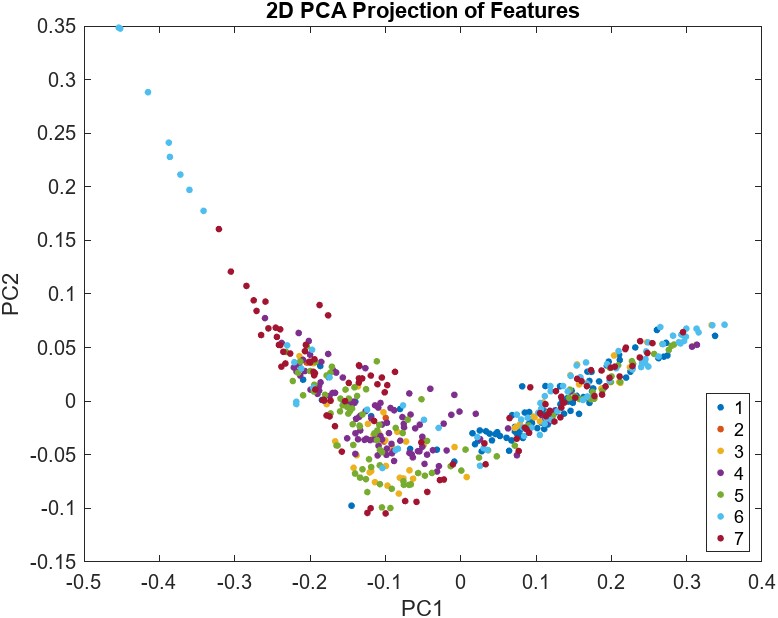
To simulate the complexity of blood vessels and lesions, we applied edge detection (e.g., Sobel or Canny filters) and computed edge density, i.e., the number of edge pixels divided by the total number of pixels.

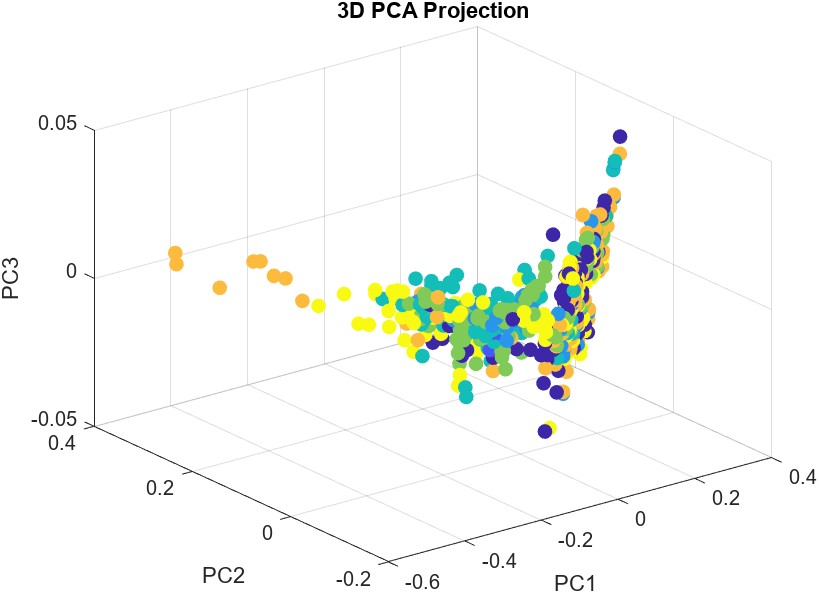


This visual and quantitative difference contributed to distinguishing images effectively, especially in differentiating non-proliferative vs. proliferative DR.

**4. Gabor Filter Properties**

We employed a bank of Gabor filters to identify texture features corresponding to various orientations. Gabor responses identified lesions of various sizes and orientations.





Every filtered image was studied in detail, and the energy (size) was noted in every direction. These were included in the final feature vector.

**5. Summary of Final Feature Vector**

Each image was converted into a feature vector with:

* 5 statistical characteristics
* 4 GLCM features multiplied by 4 directions is 16.
* 1 edge density
* 6 Gabor energies (one for each orientation) Every image has about 28 different characteristics. These were standardized and input into classifiers in the next phase.

# 7. Machine Learning Models and Evaluation

**Objective:**

The objective of this phase was to train and compare machine learning classifiers to differentiate between various grades of diabetic retinopathy (DR) severity levels based on features constructed from fundus images. The issue was simplified to a three-class classification problem by reducing the original seven DR stages to:

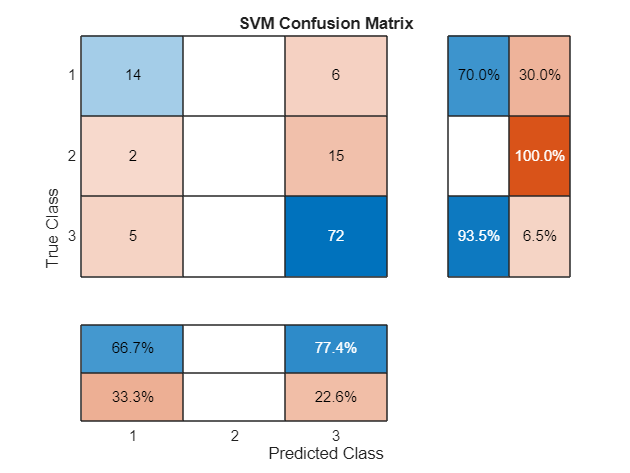
* No DR
* Mild/Moderate Non-Proliferative DR
* Severe NPDR and Proliferative DR

This re-grouping was clinically meaningful and helped reduce class imbalance and increase classification clarity.

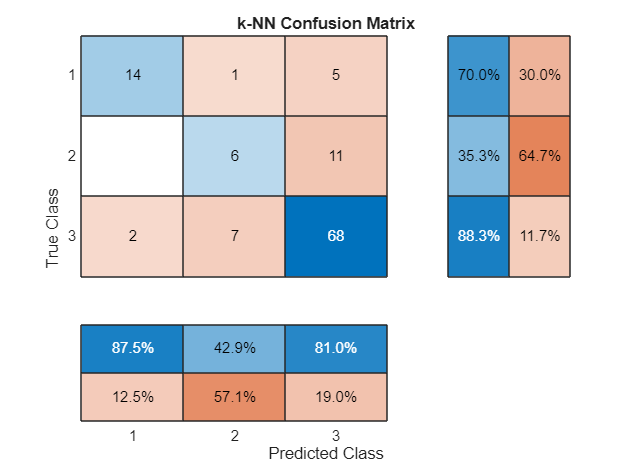
**Classifier Construction**

Four supervised learning models were selected, all of which are included in MATLAB's Statistics and Machine Learning Toolbox:

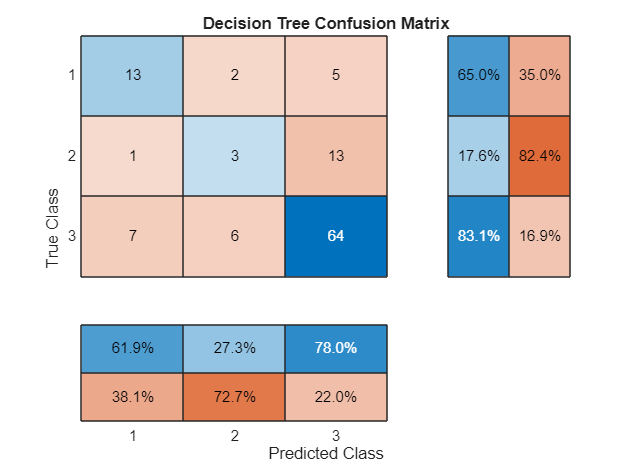
* **Support Vector Machine (SVM) with fitcecoc**



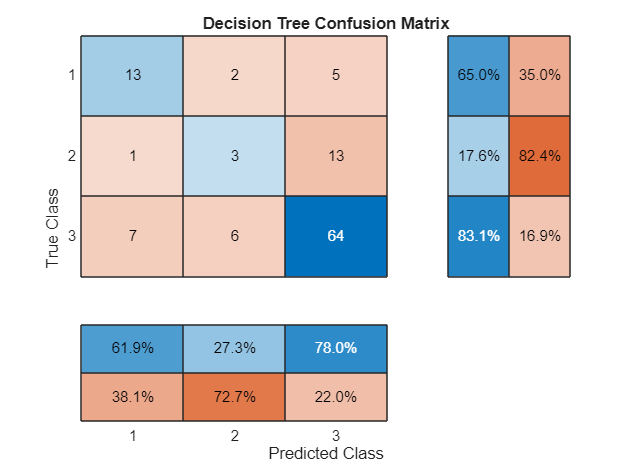
* **k-Nearest Neighbors (k-NN) with 5 neighbors.**



* **Decision Tree Classifier**



* **Discriminant Analysis Classifier (LDA)**



Each classifier was trained on normalized feature vectors containing color, texture, and edge-based descriptors. Model evaluation was conducted using 80/20 hold-out cross-validation. Feature scaling was performed through z-score normalization to avoid the influence of value range differences (e.g., contrast vs. edge density) on the learning algorithms.

**Evaluation Metrics**

Each model’s performance was assessed using:

* Classification Accuracy
* Confusion Matrix
* Classification Report (Precision, Recall, F1-score)
* ROC Curves and AUC (One-vs-All for all classes)

**Classification Results:**

* Classifier\tAccuracy (%)
* Support Vector Machine: 75.44
* k-Nearest Neighbors: 77.19
* Decision Tree: 70.18
* Discriminant Analysis: 68.42

The k-NN classifier was the most accurate, followed closely by SVM. This indicates that neighborhood-based classification performs well on the structured feature space.

This section compares the performance of four classical machine learning algorithms—Support Vector Machine (SVM), k-Nearest Neighbors (k-NN), Decision Tree, and Discriminant Analysis—to classify retinal images into three categories: No DR, Mild/Moderate Non-Proliferative Diabetic Retinopathy (NPDR), and Severe/Proliferative Diabetic Retinopathy (PDR). The performance is measured with a set of performance metrics such as accuracy, precision, recall, F1-score, and the Area Under the ROC Curve (AUC) which are computed based on the classification results.

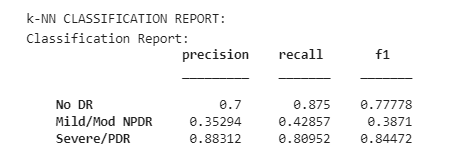
**Accuracy and Confusion Matrix Overview**

The highest overall accuracy model was k-NN (77.19%). SVM was next at 75.44%. Decision Tree was 70.18%, and Discriminant Analysis was a little lower at 68.42%. The confusion matrices accurately showed these results. k-NN and SVM were able to classify all three classes reasonably well. The Decision Tree and Discriminant models, however, had difficulty classifying Mild/Mod NPDR. They confused it with either No DR or Severe/PDR.

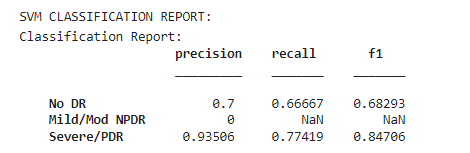
**Classification Reports: Precision, Recall, and F1**

From the classification reports:

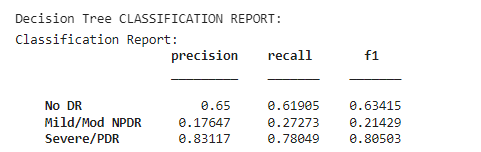
* **k-NN** performed well in general, particularly for Severe/PDR (F1-score: 0.84) and No DR (F1-score: 0.77). But it was poor for Mild/Mod NPDR (F1-score: 0.39).



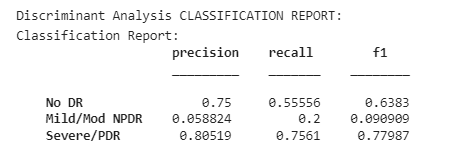
* **SVM** performed highly well in the prediction of Severe/PDR (0.85 F1-score) but completely failed to predict Mild/Mod NPDR, with NaN since it had zero recall.



* **Decision Tree** achieved decent accuracy for Severe/PDR (0.83) but had very poor F1-score for Mild/Mod NPDR (0.21).



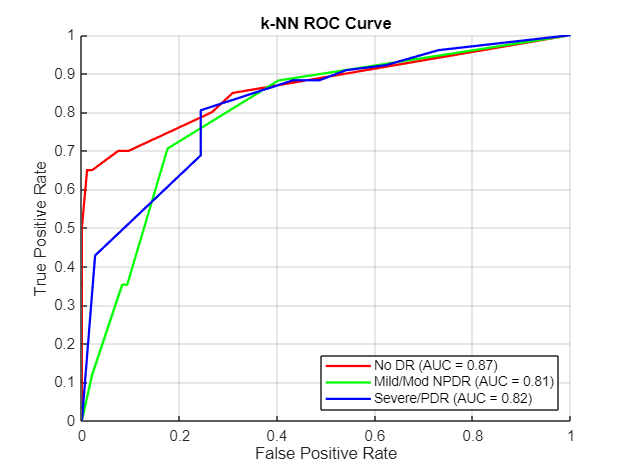
* **Discriminant Analysis** was poorest in classifying Mild/Mod NPDR (F1-score: 0.09). Nevertheless, it still accurately identified Severe/PDR with a great F1-score of 0.78.



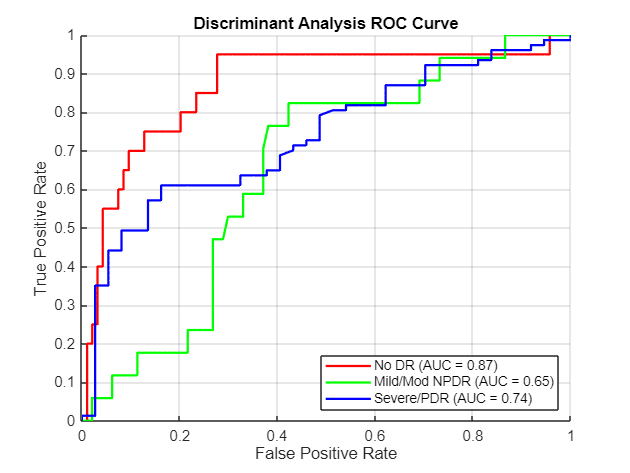
**ROC Curves and AUC**

ROC curve analysis provided additional information on model discrimination power:

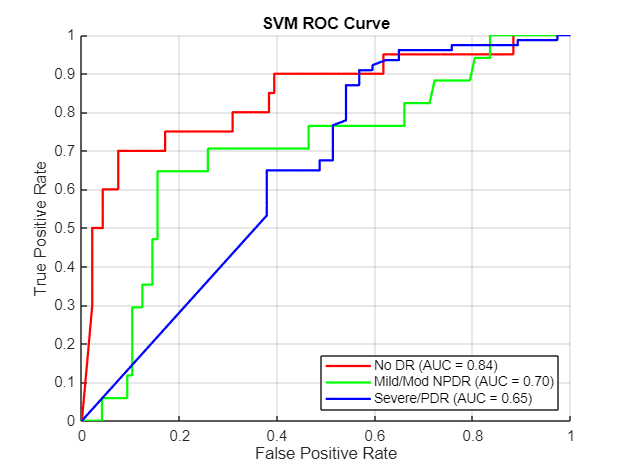
* **k-NN** had high AUC values: No DR = 0.87, Mild/Mod NPDR = 0.81, Severe/PDR = 0.82—indicating good separability for every class.

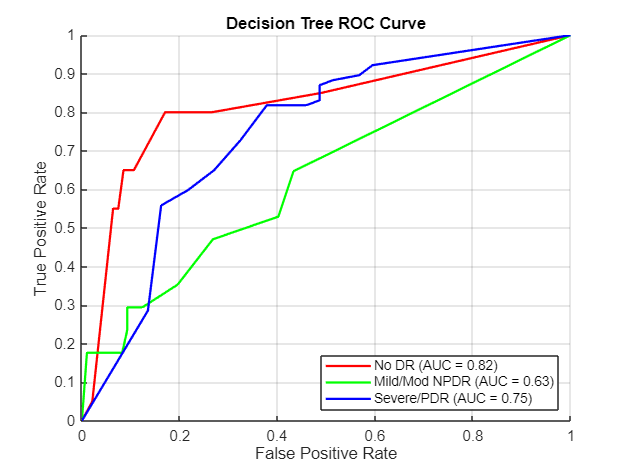


* **Discriminant Analysis** performed poorly with Mild/Mod NPDR (AUC = 0.65) and Severe/PDR (AUC = 0.74) but also worked well with No DR (AUC = 0.87).



* **SVM** and **Decision Tree** curves match their performance metrics—SVM probably has a high AUC for Severe/PDR and No DR, but there are still problems with Mild/Mod NPDR.





Interpretation The hardest for all models is the distinction between Mild/Moderate NPDR. This is perhaps because it shares characteristics of both the No DR and Severe/PDR classes. We can observe this in the precision and recall differences, and the lower AUCs for this class. That is, whereas regular models can distinguish between healthy and severely affected cases with ease, the more subtle diabetic retinopathy cases require improved feature engineering or perhaps multiple models or deep learning methods. Overall, the comparison indicates that k-NN provides the most robust and stable results. SVM performs well in Severe DR detection but not for all classes. Discriminant and Decision Tree models are interpretable but perform poorly for intermediate cases and can be improved by boosting algorithms. These results provide avenues for further improvement, including improving models and learning complex features.

# 8. Discussion on implementing a Traditional Machine learning Pipeline

To address the Task 3 problem—identification and detection of various grades of Diabetic Retinopathy (DR) from fundus images—a solution was suggested relying on conventional machine learning models and not deep learning techniques. The primary goal was to develop an easy-to-understand, easy-to-implement, and efficient classification system capable of distinguishing No DR, Mild/Moderate Non-Proliferative Diabetic Retinopathy (NPDR), and Severe/Proliferative Diabetic Retinopathy (PDR) on the basis of image features and not on intricate neural networks.

To do this, a pipeline process was established. Features were initially extracted from retinal images using pre-processing methods like resizing, color normalization, and extraction of statistical or texture-based features (e.g., histogram of oriented gradients, GLCM features, or color histograms). These handcrafted features were subsequently passed as input to conventional classifiers like Support Vector Machine (SVM), k-Nearest Neighbors (k-NN), Decision Tree, and Discriminant Analysis. These models were selected because they are interpretable and can perform well on small datasets without the computational cost of deep learning.

All the models were trained and tested on a three-phase DR labeled dataset. The models were extensively evaluated using a range of metrics: accuracy, precision, recall, F1-score, and ROC-AUC scores. The k-NN model performed the best with an accuracy of 77.19% and high AUC scores for all the classes with a reasonable sensitivity-specificity trade-off. SVM performed the second-best with high accuracy in detecting severe cases but was not suitable for the middle stages. Decision Trees and Discriminant Analysis models were more interpretable but poorer performing, particularly for the Mild/Moderate NPDR class. The main problem in this project was the variation among DR classes, especially at early to moderate levels. Classical models allowed us to test hypotheses effectively and contrast results while having well-defined decision-making—something that is very important in medical diagnosis. Additionally, this method gave a basic foundation that can be built upon using better feature selection, data enhancement, or possible use of deep learning methods for better accuracy and overall results.

In conclusion, the solution developed effectively offered a real-world and computationally effective method of DR classification using conventional machine learning, with a firm foundation for further research in automated retinal screening systems.

# Conclusion

This report presented a machine learning-based approach for the classification of Diabetic Retinopathy (DR) using handcrafted features extracted from retinal fundus images. Through rigorous preprocessing, statistical analysis, and feature extraction techniques—such as GLCM, Gabor filters, and entropy measures—retinal images were converted into structured feature vectors suitable for classical ML models. Among the classifiers tested, k-NN and SVM demonstrated strong performance, particularly in detecting severe cases of DR. The findings highlight the feasibility of using interpretable, cost-effective machine learning methods for automated retinal screening, especially in resource-constrained settings. Despite promising results, challenges such as class imbalance and difficulty in identifying intermediate DR stages remain. Future work can explore ensemble methods, deep learning integration, and real-time deployment within clinical environments. Overall, this study contributes a scalable and explainable solution that can support early DR detection, assist ophthalmologists, and ultimately help reduce diabetes-related vision loss worldwide.

# Bibliography

[1]

Z. Yang, T.-E. Tan, Y. Shao, Tien Yin Wong, and X. Li, “Classification of diabetic retinopathy: Past, present and future,” *Frontiers in Endocrinology*, vol. 13, Dec. 2022, doi: https://doi.org/10.3389/fendo.2022.1079217.

[2]

L. Math and R. Fatima, “Adaptive machine learning classification for diabetic retinopathy,” *Multimedia Tools and Applications*, Oct. 2020, doi: https://doi.org/10.1007/s11042-020-09793-7.

[3]

V. Gulshan *et al.*, “Development and Validation of a Deep Learning Algorithm for Detection of Diabetic Retinopathy in Retinal Fundus Photographs,” *JAMA*, vol. 316, no. 22, p. 2402, Dec. 2016, doi: https://doi.org/10.1001/jama.2016.17216.

[4]

C. Leibig, V. Allken, M. S. Ayhan, P. Berens, and S. Wahl, “Leveraging uncertainty information from deep neural networks for disease detection,” *Scientific Reports*, vol. 7, no. 1, Dec. 2017, doi: https://doi.org/10.1038/s41598-017-17876-z.

[5]

C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, “Rethinking the Inception Architecture for Computer Vision,” *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 2818–2826, Jun. 2016, doi: https://doi.org/10.1109/cvpr.2016.308.

[6]

N. Tajbakhsh *et al.*, “Convolutional Neural Networks for Medical Image Analysis: Full Training or Fine Tuning?,” *IEEE Transactions on Medical Imaging*, vol. 35, no. 5, pp. 1299–1312, May 2016, doi: https://doi.org/10.1109/tmi.2016.2535302.

[7]

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, Mar. 2018, doi: https://doi.org/10.1038/s41433-018-0064-9.

[8]

Zahangir Alom, Md. Mahmudul Hasan, C. Yakopcic, T. M. Taha, and V. K. Asari, “Recurrent Residual Convolutional Neural Network based on U-Net (R2U-Net) for Medical Image Segmentation,” Feb. 2018, doi: https://doi.org/10.48550/arxiv.1802.06955.

[9]

R. Ghosh, K. Ghosh, and S. Maitra, “Automatic detection and classification of diabetic retinopathy stages using CNN,” *2017 4th International Conference on Signal Processing and Integrated Networks (SPIN)*, Feb. 2017, doi: https://doi.org/10.1109/spin.2017.8050011.

[10]

L. Wu, “Classification of diabetic retinopathy and diabetic macular edema,” *World Journal of Diabetes*, vol. 4, no. 6, p. 290, 2013, doi: https://doi.org/10.4239/wjd.v4.i6.290.

[11]

C. Mohanty *et al.*, “Using Deep Learning Architectures for Detection and Classification of Diabetic Retinopathy,” vol. 23, no. 12, pp. 5726–5726, Jun. 2023, doi: https://doi.org/10.3390/s23125726.

[12]

L. M. Jampol, R. Tadayoni, and M. Ip, “Need for a New Classification of Diabetic Retinopathy,” *Retina*, vol. 41, no. 3, pp. 459–460, Dec. 2020, doi: https://doi.org/10.1097/iae.0000000000003070.

[13]

M. Z. Atwany, A. H. Sahyoun, and M. Yaqub, “Deep Learning Techniques For Diabetic Retinopathy Classification: A Survey,” *IEEE Access*, pp. 1–1, 2022, doi: https://doi.org/10.1109/access.2022.3157632.