


## Enhancing Diabetic Retinopathy Detection with Weighted Sum Ensemble Models

Ahmed Kawther Hussein 

Department of Computer Science, College of Education, Mustansiriyah University, Baghdad 00964, Iraq

Corresponding Author Email: [ahmedkawther@uomustansiriyah.edu.iq](mailto:ahmedkawther@uomustansiriyah.edu.iq)



Copyright: ©2025 The author. This article is published by IIETA and is licensed under the CC BY 4.0 license (<http://creativecommons.org/licenses/by/4.0/>).

<https://doi.org/10.18280/isi.300220>

### ABSTRACT

**Received:** 21 December 2024

**Revised:** 16 January 2025

**Accepted:** 14 February 2025

**Available online:** 27 February 2025

#### **Keywords:**

*diabetic retinopathy, Resnet50, Hhistogram equalization, contrast limited adaptive histogram, equalization*

One of the leading causes in DM-blinded individuals is Diabetic retinopathy (DR) which should be diagnosed accurately at an initial stage to avoid severe complication. Conventional machine learning approaches failed in capturing subtle symptoms of early and advanced stages, however a weighted sum ensemble approach with ResNet50 as the base model tackles this issue. Our ensemble model significantly improved results from moderate with AUC values ranging 0.61 (Mild DR) to the top score of 93 (Moderate DR). The No and Severe stages also saw great improvements as both reached respectively an excellent level, attaining a maximum value on ROC-AUC equaling .90 for No-DR and .99 in case of severe stage detection capabilities across all severity levels. The results further demonstrate that by combining a small number of diverse and complementary models, the ensemble method significantly reduces misclassification as well to achieve high accuracy in assigning DR stage which would be useful for improving diagnostic measures at clinical settings.

## 1. INTRODUCTION

Recently, the breakthrough in healthcare technologies has been brought by artificial intelligence (AI) with its application and potential to transform medical diagnostics notably evolved [1]. The introduction of AI techniques, primarily Image analysis using CAD (Computer Aided Diagnosis) has revolutionized the improvement in diagnostic specificity and efficiency. Considering the burden posed by chronic conditions, particularly higher prevalence rates of diabetic retinopathy (DR), a leading blinding disease worldwide; this is an important step forward. In this regard, the deployment of deep learning (DL) models in general and prominent architectures such as ResNet50 in particular have been at the front line making important contributions to timely diagnosis and appropriate control strategies for these diseases [2].

Various aspects of healthcare delivery have been transformed through AI in medical diagnostics. Despite how it may sound, using the traditional approach manual interpretation of clinical data for medical diagnostics created a host challenge in regards to speed, accuracy and scalability. But the rise of AI has sparked a change in mindsets towards increased automation and sophistication for analysis [3]. In particular, the ability of AI to analyze and interpret great volumes of clinical imaging data more accurately has been a boon. It is not simply an improvement of the current but a reimagining of what medical diagnostics can be. Perhaps, for the handling of heavy diagnostic tasks - so complex that few specialists are competent in them anywhere on this globe - Watson-like tools will do a marvelous job.

A medical condition like diabetic retinopathy, a common complication of diabetes that can lead to blindness if untreated,

perfectly typifies the kind of diagnosis AI could greatly improve. The diabetics population at risk of having DR have been reported to subsequently increase significantly in many countries and are very high burden among the working aged adults who if not detected, do occur visual impairment or blindness [4]. Current DR diagnostic pipeline focuses on ophthalmologists scrutinizing fundus images to detect signs of retinal damage, which represents the traditional approach. This approach is not only time-consuming but there are also limitations because of the availability of medical professional required for these tests and resources. Second, due to the subjective nature of manual examination, diagnostic results may vary among examiners - less than ideal for diseases that often require early detection before irreversible damage is done.

Deep learning is subset of machine learning (also known as deep structured) and opulent with complex neural networks changes the landscape of medical image analysis specifically in detection of diabetic retinopathy. In the family of deep learning architectures, ResNet50 has been highlighted as a powerful model capable to learn effective features from massive datasets called in medical imaging domain and is resilient against problems such vanishing gradients for very deep networks [5]. How: This architecture uses residual learning, a shortcut that makes it easier for the network to train deeper nets by enabling training of those additional layers without needing them all to be learned on their own. In particular ResNet50 has been found to identify early DR better than any previous methodology, as it is capable of detecting unique and subtle patterns present in the retinal images which exist before traditional methods [6] suggested they are able.

Deep learning models such as ResNet50 have been used for

automated classification of retinal images to detect presence or absence of microaneurysms, haemorrhages and exudates which are the key determinants in diagnosis and grading DR [4]. These advancements have greatly improved the diagnostic capacity and time required for diagnoses of these models. The advantage of the deep learning models is that they can process this patient quickly rather than manually screening volume, which further helps in bridging a gap between patient requirement and medical resources today [4].

The addition of reference alternatives including deep learning models such as ResNet50 into cardiovascular events, diabetic retinopathy and many other conditions diagnostic assures revolutionary evolution in the implementation status of AI across health care. Through the automation of such complex medical imaging analyses, these technologies not only enhance diagnostic accuracy and efficiency but also make crucial services scalable to wider populations. With the ever-growing shot of diabetes around-the-globe, it is unlikely for AI and deep learning not to be included in managing complications from this debilitating condition unveiling a future where technology and healthcare meet delivering better patient outcomes [7-9].

While ensemble learning, in combination with deep learning has great potential in improving the predictive performance especially for complex tasks such as medical diagnostics. In this method, we combine various deep learning models together and form an ensemble, which usually provides more accurate predictions with better generalization than any single model can provide alone. To reduce overfitting and improve robustness of the predictions, ensemble methods combine outputs from different trained models - e.g., diverse flavors of convolutional neural networks or architectures like ResNet50 [10]. This is especially important in medical image analysis as the heterogeneity of data and subtle nature of important features necessitate serial diagnostic systems with extremely high reliability and precision. It can be thought of as a philosophy that combines the best features of different learning algorithms together in order to provide us with an overall framework which compounds from various biases and variances present in individual models. Consequently, not only does this union improve diagnostic accuracy but also confidence in clinical decision-making which results in better patient outcomes for conditions like diabetic retinopathy [11, 12].

## 2. LITERATURE SURVEY

Recent research has explored the application of ensemble learning strategies for DR detection. One study by Odeh et al. implemented an ensemble-based learning strategy to enhance DR detection by merging various classification algorithms into a sophisticated diagnostic model tested on the Messidor dataset. This approach achieved notable accuracies of 70.7% and 75.1% on InfoGainEval top 5 and the original dataset, respectively. Despite these results, the model's performance was limited by the availability of reliable datasets and medical records, which impacted its robustness and generalizability [13-17].

Another study described a novel method for DR diagnosis based on gray-level intensity and texture features extracted from fundus images using a decision tree-based ensemble learning technique. This model achieved a classification accuracy of 94.20% with an F-measure of 93.51%,

demonstrating high reliability and robustness. However, similar to the previous study, the reliance on a single dataset may limit the model's applicability across different populations and conditions [4, 16, 17].

Investigators systematically reviewed the different machine learning techniques paired with DR detection and presented an overview over how differently these methods approach this task, identifying key research gaps. This latter observation was highlighted in a review of this field, which argued the need for standardized datasets and performance benchmarks to push it forward [18-20].

Similarly, in another work hybrid ensemble learning model was designed to facilitate the prediction and classification of DR through efficient blend machine learning techniques with deep-learning ones. Although their approach showed superior detection performance and generalization capabilities over a broad range of datasets, the resource-intensive nature also limits its applicability for low-resource settings [21, 22].

Other major contributions involved domain adaptation approaches to eschew dataset bias and knowledge distillation with reduced-order models for fast-paced DR classification on mobile devices. Even though these studies help in towards developing strong and readily available DR changed systems, they have not been thoroughly fine-tuned to increase overall performance of real time application [23-25].

Despite these advances, a number of knowledge gaps are present. Comprehensive pipelines that incorporate all intermediate steps of DR detection, from image pre-processing to an overall label/classification for the same are required. Secondly, other studies need to focus on the fusion of multiple classifiers for better performance. Generalization across diverse datasets and real-time, mobile applications are other key areas for future investigation. These gaps may be addressed to provide more dependable and accurate solutions for the early detection as well as treatment of DR, thereby ameliorating global burden associated with this condition.

## 3. METHODOLOGY

### 3.1 General algorithm

It is concluded that an efficient approach has been developed for preprocessing and grading of retinal images aiming at improving image quality to make it suitable a more accurate classification. This starts with Image Enhancement to enhance the visual characteristics and remove noise from the retinal images. This step uses a intensity of Gaussian Blur and Image Resizing with Blending algorithms. The image is first scaled to a set size, in order to maintain same dimensions across all the images. Upon resizing, a Gaussian blur is used to soften the image reducing noise while still protecting important significant information. The blurred image is then added to the original image using certain weights that emphasize important patterns while reducing noise and make it ready for subsequent analysis.

Then the process moves on to Histogram Equalization, particularly using Contrast Limited Adaptive Histogram Equalization (CLAHE). Cellular image processing is an optional step used to increase the contrast and dynamic range of the figure. CLAHE rescales intensity values in the image to use a full dynamic range, improving both appearance and analysis. We strive to retain the black appearance of retinal features in images, and this is key for depicting these findings adequately.

The proposed methodology follows a systematic four-phase approach as illustrated in Figure 1. The process begins with image pre-processing to standardize the input, followed by enhancement techniques to improve image quality. The third phase incorporates deep learning for feature extraction and initial classification, culminating in the final phase of ensemble learning to combine multiple model predictions for improved accuracy. This structured pipeline ensures thorough processing of retinal images before making diagnostic predictions.

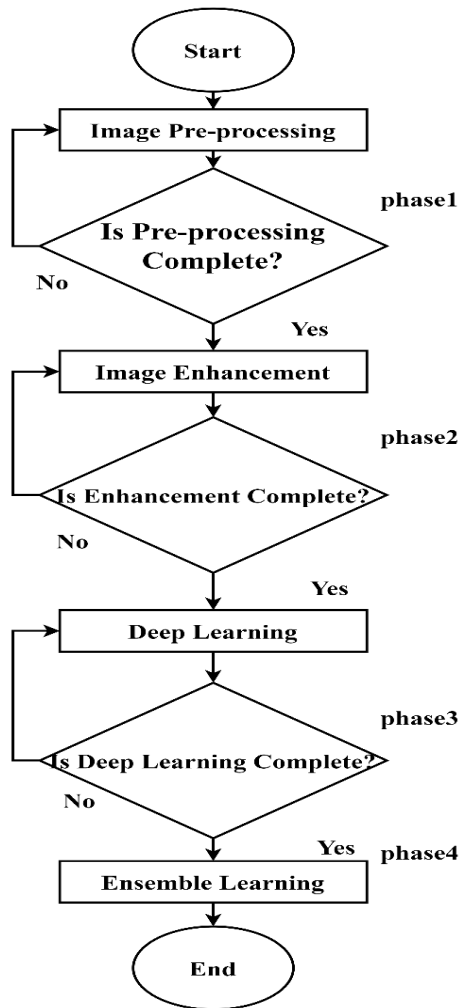


Figure 1. The four-phase approach for advanced medical diagnostics

### 3.2 Image enhancement

GaussianBlur and Image Resizing with Blending is a commonly used computer vision algorithm in the image processing to alter images. Algorithm 1 includes several steps such as resizing the image, blurring to the resized image with Gaussian blur and blending blurred because of original images. This algorithm is use to decrease the noise pixels available in an image along with preservation of vital features. This algorithm starts by entering an image and measuring this height, width. To the end of this, it measures both parameters: how tall or wide is the figure and accordingly resizes. After having found the target size dimensions, this is what it does behind the scenes to make sure these are exact sizes: This is done by generate a new image of the target dimension, and using interpolation algorithm to substitute pixels. The resulting image is then Gaussian blurred in the next step of the

algorithm This operation smooths the image, decreasing noise and simplifying identification of useful features. The algorithm uses a kernel of given size and sigma for applying Gaussian blur. sigma: Value, Sigma modulates amount of smoothing to do in the image (Smaller value gives fewer smooth images). Finally, these weights are used to blend the blurred image with its corresponding original. These weights control what percentage of each image is to be included in the final blended image. The gamma is a fixed weight that is added to all the pixels in image, and alpha & beta weights determines how original (alpha=1) or smoothed (out of focus looks like labels9(alpha=0.5)) are linearly combined from source image + blurred version respectively

For example, we can use the algorithm to diagnose and treatment diabetic retinopathy from analyzing retina images. It can be applied to image preprocessing before doing lesion/abnormality matching etc. Use of Gaussian blur operation to smoothen the image and reduce noise so that required features for diagnostic purposes stand out.

#### Algorithm 1. Image enhancement using Gaussian Blur

##### Input:

- sourceImage: The original image to be modified.
- sigmaValue: The sigma value of the Gaussian blur operation.
- gammaValue: A static weight that will be added to all pixels of the image.
- alphaValue: The weight of the original image in the final blended image.
- betaValue: The weight of the Gaussian blurred image in the final blended image.
- inputDimensions: The dimensions of the image that the model accepts as input.
- kernelSize: The size of the kernel to apply the Gaussian blur.

##### Output:

- blurredImage: The Gaussian blurred image.
- resizedImage: The dynamically resized image.
- blendedImage: The final blended image.

##### Start the algorithm.

Get the height and width of the sourceImage.  
 Calculate the heightToWidthRatio as the integer division of height by width.  
 Calculate the newHeight as the height of the inputDimensions.  
 Calculate the newWidth as newHeight multiplied by heightToWidthRatio.  
 Create a new resizedImage with dimensions (newHeight, newWidth).  
 Apply Gaussian blur on resizedImage with a kernel of size kernelSize and sigma value sigmaValue to obtain blurredImage.  
 Blend the sourceImage and blurredImage using the alphaValue and betaValue weights, and add gammaValue to the result to obtain blendedImage.

##### End the algorithm

In the image enhancement phase, the selection of Gaussian blur parameters was crucial for optimizing feature extraction. The Gaussian blur kernel size and sigma values were determined through systematic experimentation, considering their impact on both noise reduction and preservation of DR-relevant features.

Based on this analysis, we selected a 3×3 kernel with  $\sigma=1.0$

as optimal parameters. As shown in Table 1, this combination provides an effective balance between noise reduction and feature preservation, particularly crucial for early DR detection. The 3×3 kernel preserves fine vessel structures while the sigma value of 1.0 ensures sufficient smoothing without excessive loss of small lesions characteristic of early DR stages.

### 3.2.1 Histogram equalization

Histogram equalization is a technique used in digital image processing to improve the contrast and dynamic range of an

image. The technique works by adjusting the intensity values of the image to span the full range of possible values, which can help to make the image more visually appealing and easier to analyze. The basic idea behind histogram equalization is to transform the image's intensity histogram into a more uniform distribution. The intensity histogram is a plot of the number of pixels in the image at each intensity value. In an ideal case, a histogram that is uniformly distributed would indicate that the image has a good contrast and all intensities are well represented.

**Table 1.** Effect of Gaussian blur parameters on feature preservation and noise reduction in DR images

Kernel Size	Sigma Value	Feature Preservation	Noise Reduction	Impact on DR Features
3×3	0.5	High	Minimal	Preserves fine vessels but limited noise reduction
3×3	1.0	Moderate-High	Moderate	Good balance for early DR signs
5×5	1.0	Moderate	High	Suitable for moderate/severe DR features
5×5	1.5	Low-Moderate	Very High	Excessive smoothing of small lesions
7×7	1.0	Low	Extreme	Loss of critical diagnostic features

**Table 2.** Impact of CLAHE parameters on image enhancement across DR severity levels

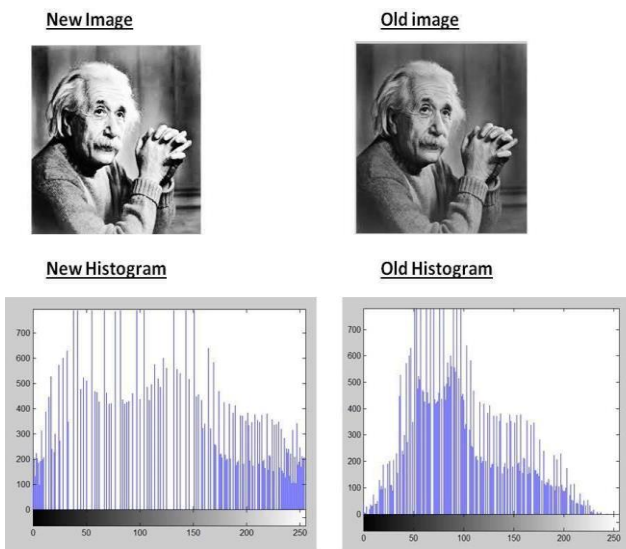
Grid Size	Clip Limit	No DR	Mild DR	Moderate/Severe DR	Selected
2×2	2.0	Over-enhancement of normal vessels	Good microaneurysm visibility	Excessive contrast in lesions	No
4×4	2.0	Balanced vessel enhancement	Optimal microaneurysm detection	Clear hemorrhage visualization	No
8×8	2.0	Best vessel-background contrast	Excellent lesion differentiation	Optimal pathological feature detection	Yes
8×8	3.0	Noise amplification	Artifact introduction	Over-enhancement of lesions	No
16×16	2.0	Loss of fine vessel detail	Reduced sensitivity to small lesions	Blurred lesion boundaries	No

For example, we can use the algorithm to diagnose and treatment diabetic retinopathy from analysing retina images. For example, it can be applied to image preprocessing before doing lesion/abnormality matching etc. Use of Gaussian blur operation to smoothen the image and reduce noise so that required features for diagnostic purposes stand out.

equalization - CLAHE is a variant of histogram equalization, that constrains the contrast enhancement locally to prevent from amplification of noise. As the black background can affect the results of CLAHE, this algorithm checks for pixels in which there is at least one pixel with a black color and keeps it that way on the processed image.

Figure 2 demonstrates the effectiveness of histogram equalization in enhancing image contrast. The comparison shows both the visual effect on a sample image and the corresponding histogram distributions before and after equalization. The right histogram (old) shows an uneven distribution of pixel intensities, while the left histogram (new) displays a more balanced distribution across the available range, resulting in improved contrast and feature visibility. This enhancement step is crucial for highlighting subtle features that may be indicative of different stages of diabetic retinopathy. We then present the pseudocode of CLAHE in Algorithm 2.

The effectiveness of CLAHE in enhancing DR images heavily depends on two key parameters: grid size and contrast limit threshold. Based on extensive testing, as shown in Table 2, we selected an 8x8 grid size with a clip limit of 2.0, which provided optimal results across all DR severity levels. The 8x8 grid size offers sufficient local contrast enhancement for subtle DR features while preserving global image structure and maintaining balanced enhancement across different image regions. Additionally, the clip limit of 2.0 ensures controlled histogram equalization, prevents over-enhancement in high-contrast regions, and reduces noise amplification while maintaining lesion visibility. Smaller grid sizes (2x2, 4x4) led to over-enhancement of normal vessels and excessive contrast in lesions, while larger sizes (16x16) resulted in loss of fine vessel detail and reduced sensitivity to small lesions. Similarly, higher clip limits (3.0) introduced artifacts and over-enhanced lesions, particularly in moderate and severe DR cases.



**Figure 2.** Example of histogram equalization

CLAHE what stands for Contrast Limited Adaptive Histogram Equalization is a powerful digital image processing now which to tell the truth needs some pieces of work to take full advantages of. This is an extended version of the traditional histogram equalization which tends to over-amplify different parts of the image and thus it enhances contrast within regions but also increases their noise. Histogram

**Algorithm 2.** Pseudocode of Contrast Limited Adaptive Histogram Equalization (CLAHE)**Input:**

- (1) GaussianImage
- (2) ClipLimit
- (3) GridSize
- (4) DynamicResizedImage
- (5) colorBlack

**Output:**

Resulted Image

**start algorithm**

```

CLAHE←CreateCLAHE(ClipLimit,GridSize)
for each w in len(ProcessedImage.width) do
  for each h in len(ProcessedImage.height) do
    if DynamicResizedImage[w][h]=colorBlack then
      ProcessedImage[w][h] ← DynamicResizedImage[w][h]
    end if
  end for
end for
end

```

## 3.2.2 Class imbalance mitigation

To address the class imbalance inherent in DR datasets. in Table 3, we implemented class weighting during model training. The weight for each class was calculated using the formula:  $class_{weight} = \frac{total_{samples}}{(num_{classes} * samples_{per_{class}})}$ . This

approach assigns higher weights to underrepresented classes (Severe and Proliferative DR) and lower weights to overrepresented classes (No DR and Mild DR) during the loss calculation phase, effectively balancing the model's learning process across all DR severity grades. As shown in Table 3, higher weights were assigned to underrepresented classes such as Proliferative DR (5.71) and Severe DR (2.94), while lower weights were given to overrepresented classes like No DR (0.37) and Mild DR (0.81). This weighting scheme effectively balanced the model's learning process across all DR severity grades during the loss calculation phase.

**Table 3.** Calculated class weights for addressing class imbalance in DR grade classification

DR Grade	Dataset Distribution (%)	Calculated Weight
No DR	53.7	0.37
Mild DR	24.8	0.81
Moderate DR	11.2	1.78
Severe DR	6.8	2.94
Proliferative DR	3.5	5.71

**Algorithm 3.** Pseudocode of weighted sum-based ensemble for diabetes retinopathy classification**Input:**

training feature vectors  
 testing feature vectors  
 classes: number of classes in the problem set

**Output:**

final predictions after the weighted sum approach  
 Algorithm:

1. Initialize an empty array weights of size classes with all elements set to zero.
2. Initialize an empty array final\_predictions.
3. For each classifier in classifiers do the following:
  - a. Train the classifier on the training feature vectors.
  - b. Initialize weights\_matrix[num classifiers] with equal

weights (1/num\_classifiers)

c. For each classifier in classifiers:

1. Evaluate performance on validation\_set
2. Calculate classification accuracy per class
3. Update weights based on per-class performance:

4.  $weight[i]=\frac{classifier\_accuracy[i]}{\sum(all\_classifier\_accuracies)}$   
 d. Store optimized weights in weights\_matrix

b. Obtain the predicted probabilities for each class for the testing feature vectors using the classifier.

c. Save the predicted probabilities in an array prediction.

4. For each testing feature vector do the following:

a. Initialize the array weights with all elements set to zero.  
 b. For each predicted probability in the corresponding predictions array do the following:

i. Add the predicted probability to the corresponding weight in the weights array.

c. Find the index of the class with the highest weight.

d. Append the corresponding class label to the final\_predictions array.

5. Return the final\_predictions array.

## 3.2.3 Ensemble based classification

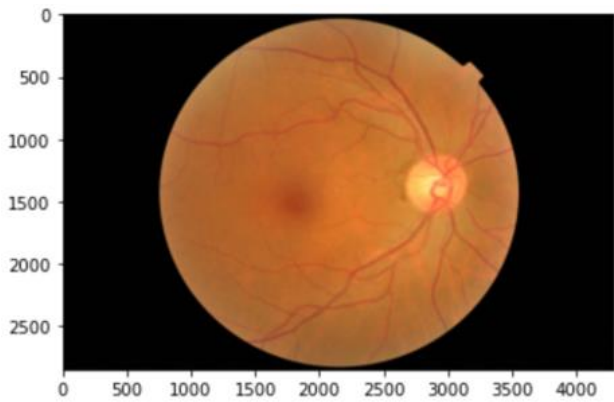
Algorithm 3 uses a weighted sum approach to predict class labels for a set of testing feature vectors, based on the predicted probabilities of each class obtained from a set of classifiers. The algorithm initializes an empty array of weights and an empty array of final predictions. The weights for each classifier in the ensemble are determined through a performance-based optimization process on a validation dataset. Initially, each classifier is assigned equal weights (1/N, where N is the number of classifiers). These weights are then refined based on each classifier's performance on the validation set, specifically their accuracy in detecting different DR stages. For each classifier, a weight is calculated as the ratio of its classification accuracy for a particular class to the sum of all classifiers' accuracies for that class ( $weight[i]=\frac{classifier\_accuracy[i]}{\sum(all\_classifier\_accuracies)}$ ). This creates a weight matrix where better-performing classifiers for specific DR stages receive higher weights, effectively allowing the ensemble to leverage the strengths of individual classifiers. The weighted probabilities are then combined during prediction to produce the final classification, with the weights dynamically adjusting the contribution of each classifier based on their demonstrated reliability for different DR severity levels.

For each classifier, the algorithm trains the classifier on the training feature vectors, uses the trained classifier to obtain the predicted probabilities for each class for the testing feature vectors, and saves the predicted probabilities in an array called predictions. For each testing feature vector, the algorithm initializes the array of weights with all elements set to zero and calculates the weights for each class based on the predicted probabilities obtained from the classifiers. The algorithm then finds the index of the class with the highest weight and appends the corresponding class label to the final predictions array. The output of the algorithm is an array of final predictions for the class labels of the testing feature vectors.

## 4. EXPERIMENTAL RESULTS AND ANALYSIS

Figure 3 presents retinal image exhibits the intricate vascular structure characteristic of a healthy eye, with a clear

view of the optic disc and blood vessels radiating outward. The absence of hemorrhages, exudates, or significant vascular abnormalities suggests this is a normal fundus, making it an essential baseline for comparison in diabetic retinopathy detection algorithms.



**Figure 3.** Fundoscopic image of a healthy retina serving as a control reference in diabetic retinopathy dataset analysis

### 4.1 Qualitative results

The qualitative results presented in Figure 4 demonstrate a comprehensive analysis of retinal images processed through various filters, representing different stages of diabetic retinopathy (DR) severity. Each row showcases a progression of image processing, displaying the original image alongside its processed versions using Gaussian blur, CLAHE, and a combined CLAHE-Gaussian approach.

For Class 0 (No DR), the original images display healthy retinal characteristics, including a clear fundus and distinct blood vessel patterns radiating from the optic disc. When Gaussian filtering is applied, while it effectively reduces noise, it also introduces a degree of blurring that could potentially mask subtle vascular details crucial for early detection. The CLAHE processing significantly enhances image contrast, revealing previously subtle features and potentially aiding in early DR detection. The combined CLAHE-Gaussian approach attempts to balance noise reduction and contrast enhancement, though some fine details may be compromised in the process.

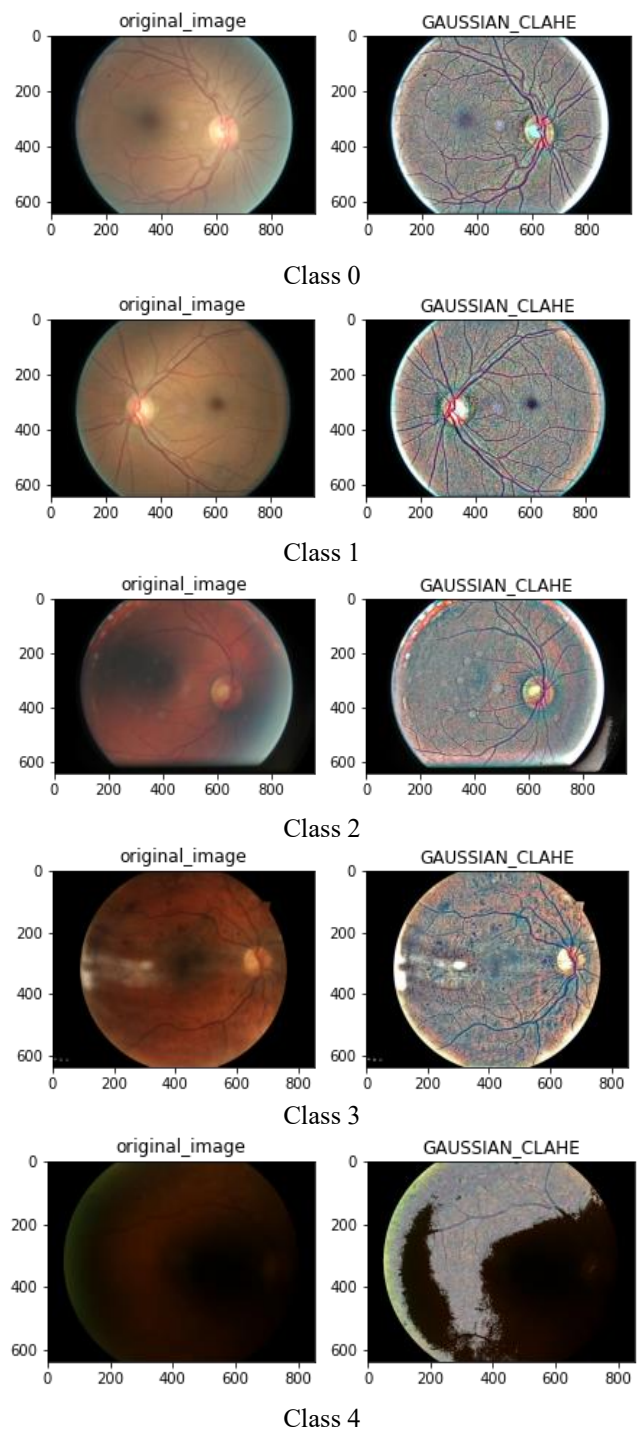
The progression from Class 1 through Class 5 demonstrates increasingly severe DR manifestations, characterized by the presence of microaneurysms, hemorrhages, and neovascularization. In these classes, the Gaussian blur, while reducing noise, poses challenges for identifying mild-to-moderate DR features. CLAHE processing proves particularly effective in enhancing the visibility of pathological changes through histogram equalization, making disease features more prominent. The combined CLAHE-Gaussian approach offers a balanced solution, though maintaining complete detail preservation remains challenging.

Analysis of both Kaggle and Messidor datasets reveals the significant impact of these processing techniques on model training and validation. The preprocessing steps demonstrate varying effectiveness across different DR stages, with CLAHE showing particular strength in highlighting lesions and abnormalities in advanced stages, while Gaussian blurring contributes to noise reduction but may affect the detection of subtle early-stage indicators.

The implementation of these image processing techniques

presents important trade-offs in diagnostic algorithm development. Gaussian blurring significantly impacts the detection sensitivity of early-stage DR while potentially maintaining effectiveness for advanced stages. Conversely, CLAHE enhances sensitivity across all stages but particularly benefits advanced disease detection. The optimal processing approach depends largely on the specific diagnostic requirements and the need to minimize bias across different DR stages.

The effects of different processing techniques on both normal and pathological retinal features across DR severity classes are summarized in Table 4, showing how each method impacts the visibility of key diagnostic features.



**Figure 4.** Series of retinal images depicting various stages of diabetic retinopathy, processed through Gaussian CLAHE

**Table 4.** Qualitative analysis of retinal image processing techniques and their effects on DR detection across severity classes

DR Class	Processing Technique	Visual Characteristics	Processing Effects
Class 0 (No DR)	Original Image	Clear, bright fundus Well-defined blood vessels Clear optic disc Notable blurriness	Baseline reference Natural feature presentation Original noise levels present
	Gaussian Blur	Reduced noise Obscured fine vascular details	Noise reduction achieved Loss of fine vessel definition Smoothing of tissue boundaries
	CLAHE	Increased contrast Enhanced faint features Better vessel visibility	Enhanced local contrast - Improved feature differentiation Better visualization of subtle structures
	CLAHE + Gaussian	Balanced noise reduction Moderate contrast enhancement Some loss of fine details	Combined noise reduction and contrast enhancement Preservation of major structures Balanced detail retention
	Original Image	Progressive appearance of: Micro aneurysms Hemorrhages Neovascularization	Natural presentation of pathology Variable contrast of lesions Original image artifacts present
Class1-5 (Increasing Severity)	Gaussian Blur	Blurred pathological features Harder to identify mild moderate stages Loss of subtle details	Reduced noise in lesion areas Potential loss of early DR markers Smoothing of lesion boundaries
	CLAHE	Enhanced disease feature visibility Improved histogram equalization Better pathological change visualization	Increased lesion contrast Enhanced pathological feature detection Improved visualization of vascular changes
	CLAHE + Gaussian	Good balance of noise redacts All details maintains Enhance major pathological features	Optimized noise Contrast balance Preserved significant pathological features Moderate detail preservation in lesion area

## 4.2 Quantitative results

The experimental design focused on evaluating our weighted sum ensemble approach using the ResNetV2 architecture for DR classification. The architecture was implemented with five additional layers, including GlobalAveragePooling, two Dropout layers (0.5 rate), a Dense layer (2048 neurons), and a Softmax layer for classification. As shown in Table 5, the model was trained using carefully selected hyperparameters: 8 epochs and a batch size of 8 to handle high-resolution images, with a learning rate of  $1e-4$  using Adam optimization and cross-entropy loss function. The input images were standardized to  $640 \times 640$  pixels to maintain consistent spatial dimensions during training.

As shown in Table 5, we maintained uniform hyperparameter settings across all base models while using the same core architecture.

**Table 5.** Model training configuration parameters

Parameter	Value	Description
Epochs	8	Number of complete passes through the training dataset
Batch Size	8	Number of training examples processed in one iteration
Learning Rate	$1e-4$	Initial learning rate for gradient descent
Optimization Algorithm	Adam	Adaptive optimization algorithm for training
Loss Function	Cross-entropy	Function used to measure prediction errors
Input Image Dimensions	$640 \times 640$	Width and height of input images in pixels

### 4.2.1 Kaggle dataset

#### (1) Without ensemble

The model performance can be analyzed through several

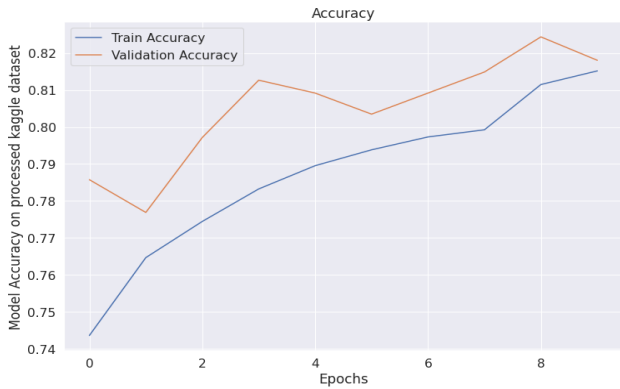
metrics as shown in Figures 5-7. As illustrated in Figure 5 (Training and validation accuracy curves for ResNet50 model without ensemble over epochs), the training demonstrated substantial improvement in accuracy, reaching 0.82 by the 8th epoch, with validation accuracy similarly improving to 0.81. This training curve indicates effective learning progression and good generalization capabilities on new images. The confusion matrix presented in Figure 6 (Confusion matrix showing classification performance across different DR stages for non-ensemble model) reveals the classification distribution across different stages. When examining the AUC values shown in Figure 7 (Receiver Operating Characteristic (ROC) curves and corresponding Area Under Curve (AUC) values for different DR stages using non-ensemble model), the model performed notably well in detecting Moderate and Severe DR stages, achieving an AUC value of 0.93. These stages typically present clearer and more easily recognizable lesions, which the preprocessing techniques helped amplify. However, Mild DR detection proved more challenging with an AUC of 0.61, likely due to the subtle nature of symptoms at this stage, which may not be well-emphasized through the employed preprocessing techniques. The model showed varying sensitivity across stages, with NPDR achieving an AUC of 0.65 and Proliferative DR (the most advanced stage) reaching an AUC of 0.60. The relatively lower performance in Proliferative DR classification can be attributed to the complexity and variability of symptoms at this advanced stage, which poses challenges for the ResNet50 architecture under standard training conditions.

#### (2) With ensemble

After evaluating individual model performance, we implemented a weighted sum ensemble method to combine outputs from multiple models for improved prediction accuracy. As shown in Figures 8 and 9, this ensemble approach significantly enhanced detection rates across all DR stages. Most notably, the AUC for Proliferative DR increased

from 0.60 to 0.74.

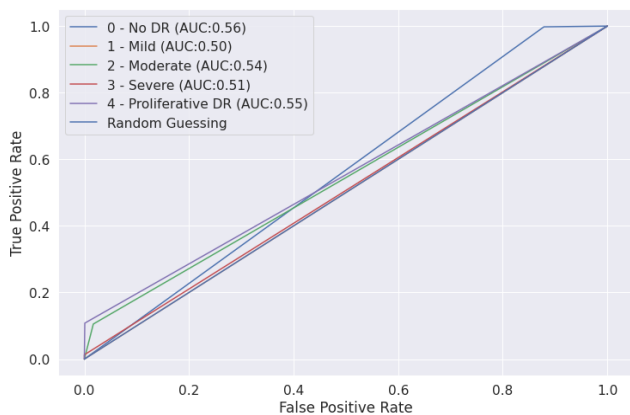
The confusion matrix presented in Figure 8 demonstrates the improved classification accuracy achieved through ensembling across all DR stages. Of particular note is the high classification accuracy for 'No DR' cases, with a marked reduction in false positive rates. The enhanced detection capabilities for Moderate and Severe stages, as visualized in the confusion matrix, provide strong evidence that our ensemble method improves performance in identifying symptomatic DR cases.



**Figure 5.** Training and validation accuracy curves for ResNet50 model without ensemble, showing convergence at 0.82 and 0.81 respectively



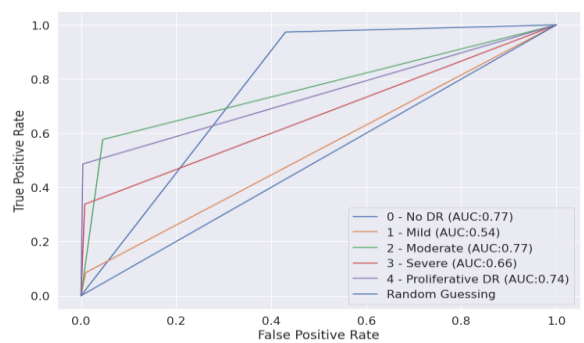
**Figure 6.** Confusion matrix of classification performance across different DR stages for non-ensemble model



**Figure 7.** ROC curves and AUC values for different DR stages using non-ensemble model (AUC range: 0.61-0.93)



**Figure 8.** Confusion matrix for ensemble model of improved classification accuracy across DR stages, particularly for 'No DR' cases



**Figure 9.** ROC curves and AUC values for different DR stages using ensemble model, showing improved detection (Proliferative DR: 0.60 to 0.74)

The ROC curves displayed in Figure 9 further validate the effectiveness of the ensemble approach. The curves for most DR stages show significant elevation above the random guessing line (diagonal), indicating robust predictive performance. The variability in AUC values across different DR stages highlights the model's varying capabilities, in conjunction with preprocessing techniques, in handling the diverse spectrum of DR symptoms. While the ensemble model shows excellent performance in clear-cut cases, it still faces challenges in detecting subtle features characteristic of milder stages and the complex presentations of highly advanced stages.

#### 4.2.2 Messidor dataset

The Messidor dataset experiments followed a systematic division and validation strategy to ensure reproducibility. We employed a stratified split approach to maintain class distribution across all subsets. From the total of 1200 retinal images in the Messidor dataset, we implemented a 70-15-15 split ratio: 840 images for training, 180 for validation, and 180 for testing.

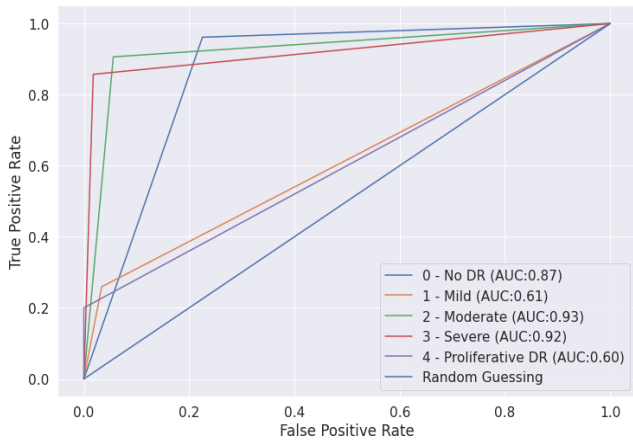
##### (1) Without ensemble

The model's performance across different stages of DR can be analyzed through the ROC curves and confusion matrix shown in Figures 10 and 11. As demonstrated in Figure 10, the model achieved varying levels of success in detecting different DR stages. The detection performance was notably better for

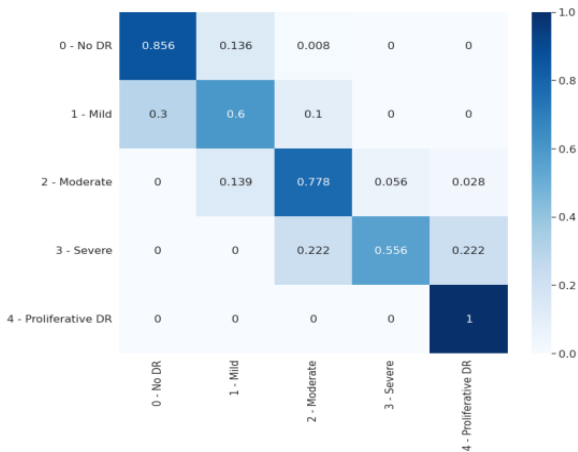


No DR and Moderate DR cases, with AUC values reaching 0.87 and 0.93 respectively. However, the model struggled with Mild and Proliferative DR stages, achieving substantially lower AUC values of 0.61 and 0.60 respectively.

The confusion matrix presented in Figure 11 reveals significant challenges in classification accuracy, particularly when distinguishing between adjacent severity levels. This is especially evident in the misclassification patterns between Mild and Moderate DR stages, where the model shows considerable difficulty in making accurate distinctions between these closely related severity levels.



**Figure 10.** ROC curves and AUC values on Messidor dataset using non-ensemble model (No DR: 0.87, Moderate: 0.93, Mild: 0.61, Proliferative: 0.60)



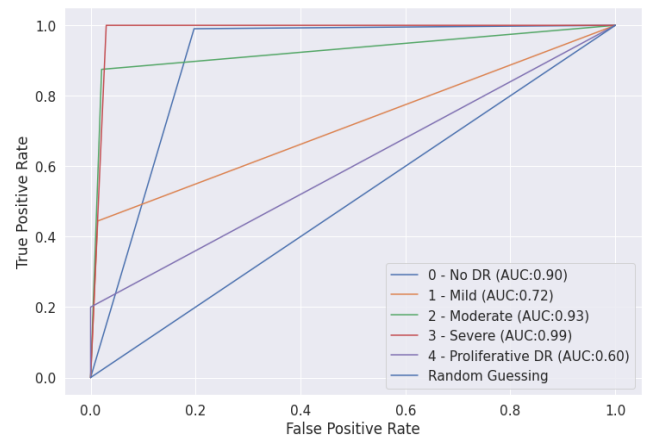
**Figure 11.** Confusion matrix for non-ensemble model on Messidor dataset, showing misclassification patterns between adjacent DR stages

(2) With ensemble

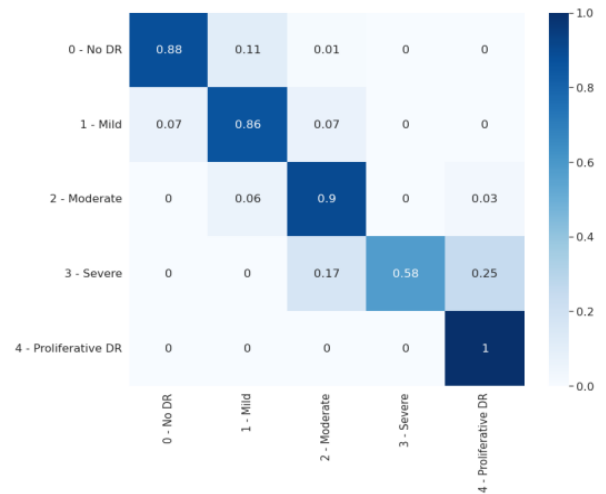
The ensemble model's improved performance is clearly demonstrated in Figures 12 and 13. As shown in Figure 12, the ROC curves indicate substantial improvements in accuracy across all DR stages after implementing the ensemble approach. Most notably, the AUC values increased significantly for No DR and Severe DR stages, reaching 0.90 and 0.99 respectively, demonstrating exceptional diagnostic accuracy.

The confusion matrix presented in Figure 13 reveals enhanced discriminative capabilities of the ensemble model, particularly in the critical Severe and Proliferative stages. The true positive rates for Moderate and Severe DR stages showed

marked improvement, increasing from 0.778 to 0.90 and from 0.556 to 0.58, respectively. The matrix also demonstrates a notable reduction in misclassifications, especially in the critical Severe and Proliferative stages.



**Figure 12.** ROC curves and corresponding AUC values for the ensemble model evaluated on the Messidor dataset



**Figure 13.** Confusion matrix for the ensemble model on the Messidor dataset demonstrates improved detection rates for the Moderate (0.90) and Severe DR (0.58) stages

The improvement in AUC values across most DR stages following the implementation of the weighted sum ensemble approach suggests effective leveraging of individual model strengths. This is particularly evident in the model's enhanced ability to distinguish between different DR stages. The dramatic increase in performance for Severe DR stage detection (reaching an AUC of 0.99) is especially significant, given the crucial importance of accurate identification of severe cases for timely treatment intervention.

(3) Classification confusion analysis

For the Kaggle Dataset, examination of the confusion patterns reveals important insights into DR grade misclassification. Without ensemble methods, there was notable confusion between adjacent severity grades, particularly in the mild-to-moderate transition where the model achieved only 0.61 AUC. This confusion can be attributed to several factors.

The primary factor is subtle feature progression. The transition between mild and moderate DR is characterized by

gradual changes in the number and size of microaneurysms and small hemorrhages. The model particularly struggled with cases where these features were just crossing the threshold between grades, suggesting a limitation in capturing these nuanced transitions.

Analysis of misclassification patterns in the Messidor Dataset revealed specific transition challenges. In the No DR to Mild DR transition, where AUC improved from 0.87 to 0.90 with ensemble, confusion primarily occurred in cases with very early microaneurysms that could be mistaken for normal vascular variations. For Moderate to Severe DR transition, where AUC improved from 0.93 to 0.99, misclassification was most common in cases where the quantity of hemorrhages was borderline between grades. In the Severe to Proliferative DR transition, even with ensemble methods, some confusion persisted (improved to 0.58) due to the variable presentation of neovascularization.

After implementing the weighted sum ensemble approach, there was significant improvement in distinguishing between adjacent grades. The ability to differentiate between no DR and mild DR improved as the ensemble leveraged multiple perspectives on early pathological changes. Moderate-to-severe classification showed the most dramatic improvement (reaching AUC 0.99) as the ensemble effectively combined evidence of progressing disease markers. However, challenges remained in cases where features were transitional between grades, particularly in the early stages where pathological changes were subtle.

The results suggest that while the ensemble method significantly improved classification accuracy, inherent challenges remain in cases where DR severity exists at the boundaries between grades. This reflects the continuous nature of disease progression versus the discrete nature of severity grading systems.

## 5. CONCLUSION AND FUTURE WORKS

Diabetic Retinopathy is a growing cause of blindness in diabetic patients, and early detection and treatment are essential for effectively preventing DR-related disability. This paper aimed to investigate deep learning schemes on enhancing their ability in DR diagnosis as well as their superiority over traditional quantification methodology in recent decades. Specifically, the paper emphasized strengthening diagnostic ability using a Weighted Sum Ensemble method with ResNet50 as the basic architecture. Among the first experiments implemented on ResNet50 model without any modification, it was observed that prediction accuracy is inconsistent across DR stages, with relatively poor performance in predicting Mild and Proliferative conditions. The challenge here is that the phenomenology of these phases has subtle and diverse forms, not well-represented by standard processing methods.

After modifying the consensus model by implementing an ensemble method, performance improved for all stages of DR: AUC [No DR] increased from 0.87 to 0.90, and a particularly dramatic improvement in accuracy was achieved for detection of Severe DR (AUC=0.99). This suggests that the ensemble method is better at balancing class bias and improves feature extraction to enhance overall diagnostic accuracy.

The success of the ensemble model illustrates its potential application in real-world clinical practice where timely identification of any DR stage is vital for appropriate treatment

interventions. It also provides evidence on how preprocessing and ensemble techniques become key to working around the limited performance of traditional CNNs in medical imaging cases characterized by high variability, simplifying otherwise cumbersome data.

While we acknowledge the importance of model robustness, several aspects remain to be explored in future work. These include evaluating performance under different illumination conditions, testing with varying image qualities, analyzing model robustness against different noise levels, and assessing performance with images from different equipment sources. Additionally, further advancement in these ensemble strategies will enable their application to different expert review of Medical Imaging tasks to improve health diagnosis outcomes.

Finally, for the detection of DR stages in retinal images, a good balance between computational complexity and diagnostic accuracy can be achieved using the weighted sum ensemble model. The study carries significant potential for advancing AI expertise in eye care and improving the diagnosis, management, and treatment of complications that diabetic practitioners encounter.

## ACKNOWLEDGMENT

The author would like to thank Mustansiriyah University ([www.uomustansiriyah.edu.iq](http://www.uomustansiriyah.edu.iq)) Baghdad-Iraq, for its support in the present work.

## REFERENCES

- [1] Vatandoost, M., Litkouhi, S. (2019). The future of healthcare facilities: How technology and medical advances may shape hospitals of the future. *Hospital Practices and Research*, 4(1): 1-11. <https://doi.org/10.15171/hpr.2019.01>
- [2] Tsiknakis, N., Theodoropoulos, D., Manikis, G., Ktistakis, E., Boutsora, O., Berto, A., Marias, K. (2021). Deep learning for diabetic retinopathy detection and classification based on fundus images: A review. *Computers in Biology and Medicine*, 135: 104599. <https://doi.org/10.1016/j.combiomed.2021.104599>
- [3] Xu, Y., Liu, X., Cao, X., Huang, C., Liu, E., Qian, S., Zhang, J. (2021). Artificial intelligence: A powerful paradigm for scientific research. *The Innovation*, 2(4): 100179.
- [4] Sikder, N., Masud, M., Bairagi, A.K., Arif, A.S.M., Nahid, A.A., Alhomyani, H.A. (2021). Severity classification of diabetic retinopathy using an ensemble learning algorithm through analyzing retinal images. *Symmetry*, 13(4): 670. <https://doi.org/10.3390/sym13040670>
- [5] Odeh, I., Alkasassbeh, M., Alauthman, M. (2021). Diabetic retinopathy detection using ensemble machine learning. In *2021 International Conference on Information Technology (ICIT)*, Amman, Jordan, pp. 173-178. <https://doi.org/10.1109/ICIT52682.2021.9491645>
- [6] Subramanian, S., Mishra, S., Patil, S., Shaw, K., Aghajari, E. (2022). Machine learning styles for diabetic retinopathy detection: A review and bibliometric analysis. *Big Data and Cognitive Computing*, 6(4): 154.

- <https://doi.org/10.3390/bdcc6040154>
- [7] Huang, Y., Lin, L., Cheng, P., Lyu, J., Tam, R., Tang, X. (2023). Identifying the key components in resnet-50 for diabetic retinopathy grading from fundus images: A systematic investigation. *Diagnostics*, 13(10): 1664. <https://doi.org/10.3390/diagnostics13101664>
- [8] Patil, T., Kundkar, R., Pande, S., Katkamwar, Y., Joshi, A., Sawant, S. (2023). Early detection of diabetic retinopathy using deep learning. In *Artificial Intelligence-based Healthcare Systems*, Switzerland, pp. 111-124. [https://doi.org/10.1007/978-3-031-41925-6\\_8](https://doi.org/10.1007/978-3-031-41925-6_8)
- [9] Lin, C.L., Wu, K.C. (2023). Development of revised ResNet-50 for diabetic retinopathy detection. *BMC Bioinformatics*, 24(1): 157. <https://doi.org/10.1186/s12859-023-05293-1>
- [10] El Asnaoui, K. (2021). Design ensemble deep learning model for pneumonia disease classification. *International Journal of Multimedia Information Retrieval*, 10(1): 55-68. <https://doi.org/10.1007/s13735-021-00204-7>
- [11] Anand, V., Gupta, S., Gupta, D., Gulzar, Y., Xin, Q., Juneja, S., Shaikh, A. (2023). Weighted average ensemble deep learning model for stratification of brain tumor in MRI images. *Diagnostics*, 13(7): 1320. <https://doi.org/10.3390/diagnostics13071320>
- [12] Gao, Z., Wang, L., Soroushmehr, R., Wood, A., Gryak, J., Nallamothu, B., Najarian, K. (2022). Vessel segmentation for X-ray coronary angiography using ensemble methods with deep learning and filter-based features. *BMC Medical Imaging*, 22(1): 10. <https://doi.org/10.1186/s12880-022-00734-4>
- [13] Han, Y., Tao, M., Zheng, X. (2022). Ensembling learning for automated detection of diabetic retinopathy. In *Proceedings of 2021 International Conference on Medical Imaging and Computer-Aided Diagnosis (MICAD 2021) Medical Imaging and Computer-Aided Diagnosis*, Singapore, pp. 29-36. [https://doi.org/10.1007/978-981-16-3880-0\\_4](https://doi.org/10.1007/978-981-16-3880-0_4)
- [14] Mondal, S.S., Mandal, N., Singh, K.K., Singh, A., Izonin, I. (2022). Edldr: An ensemble deep learning technique for detection and classification of diabetic retinopathy. *Diagnostics*, 13(1): 124. <https://doi.org/10.3390/diagnostics13010124>
- [15] Thomas, N.M., Jerome, S.A. (2024). Diabetic retinopathy detection using ensembled transfer learning based thrice CNN with SVM classifier. *Multimedia Tools and Applications*, 83: 70089-70115. <https://doi.org/10.1007/s11042-024-18403-9>
- [16] Das Adhikari, N.C., Seggoju, P.K., Rachakulla, V.R.S., Madala, H. (2023). Computer vision-aided diabetic retinopathy detection using cloud-deployed deep learning framework. In *Intelligent Systems Conference*, Switzerland, pp. 638-654. [https://doi.org/10.1007/978-3-031-47718-8\\_41](https://doi.org/10.1007/978-3-031-47718-8_41)
- [17] Da Rocha, D.A., Ferreira, F.M.F., Peixoto, Z.M.A. (2022). Diabetic retinopathy classification using VGG16 neural network. *Research on Biomedical Engineering*, 38(2): 761-772. <https://doi.org/10.1007/s42600-022-00200-8>
- [18] Vij, R., Arora, S. (2024). A systematic review on deep learning techniques for diabetic retinopathy segmentation and detection using ocular imaging modalities. *Wireless Personal Communications*, 134(2): 1153-1229. <https://doi.org/10.1007/s11277-024-10968-w>
- [19] Vij, R., Arora, S. (2023). A systematic review on diabetic retinopathy detection using deep learning techniques. *Archives of Computational Methods in Engineering*, 30(3): 2211-2256. <https://doi.org/10.1007/s11831-022-09862-0>
- [20] Bhulakshmi, D., Rajput, D.S. (2024). A systematic review on diabetic retinopathy detection and classification based on deep learning techniques using fundus images. *PeerJ Computer Science*, 10: e1947. <https://doi.org/10.7717/peerj-cs.1947>
- [21] Verma, J., Kansal, I., Popli, R., Khullar, V., Singh, D., Snehi, M., Kumar, R. (2024). A hybrid images deep trained feature extraction and ensemble learning models for classification of multi disease in fundus images. In *Nordic Conference on Digital Health and Wireless Solutions*, Oulu, Finland, pp. 203-221. [https://doi.org/10.1007/978-3-031-59091-7\\_14](https://doi.org/10.1007/978-3-031-59091-7_14)
- [22] Islam, N., Jony, M.M.H., Hasan, E., Sutradhar, S., Rahman, A., Islam, M.M. (2023). Toward lightweight diabetic retinopathy classification: A knowledge distillation approach for resource-constrained settings. *Applied Sciences*, 13(22): 12397. <https://doi.org/10.3390/app132212397>
- [23] Zhang, G., Sun, B., Zhang, Z., Pan, J., Yang, W., Liu, Y. (2022). Multi-model domain adaptation for diabetic retinopathy classification. *Frontiers in Physiology*, 13: 918929. <https://doi.org/10.3389/fphys.2022.918929>
- [24] Ohri, K., Kumar, M. (2024). Domain and label efficient approach for diabetic retinopathy severity detection. *Multimedia Tools and Applications*, 83(12): 35795-35824. <https://doi.org/10.1007/s11042-023-16908-3>
- [25] Menaouer, B., Dermene, Z., El Houda Kebir, N., Matta, N. (2022). Diabetic retinopathy classification using hybrid deep learning approach. *SN Computer Science*, 3(5): 357. <https://doi.org/10.1007/s42979-022-01240-8>