VEnDR-Net: Voting Ensemble Classifier for Automated Diabetic Retinopathy Detection

Anju Mishra, Mrinal Pandey, Laxman Singh
Dept. of Computer Science and Technology, Manav Rachna University, Faridabad
Dept. of Computer Science and Engineering (AI & ML), KIET Group of Institutions, Ghaziabad
E-mail: aanjumishra2108@gmail.com, mrinalpandey@mru.edu.in, laxman.mehlawat01@gmail.com

Keywords: image preprocessing, diabetic retinopathy, segmentation, classification, voting ensemble, deep learning

Received: April 09, 2025

Diabetic Retinopathy (DR) is a significant eye disease, which is caused by the damage of retina. To provide the best timely treatment, it is necessary to detect DR in early stages. Firstly, advanced sequential image preprocessing and segmentation techniques are employed for accurate localizing and isolating the affected regions in retinal images. Secondly, a voting ensemble classifier is introduced using a deep neural network model, which combines the predictions of multiple CNN models i.e., ResNet50, VGG16, VGG19 and GoogLeNet to enhance the overall classification performance of the proposed model. Our proposed model, named VEnDR-Net (Voting Ensemble for Diabetic Retinopathy classification using deep neural networks), implements on the EyePACS dataset and achieves 0.97 sensitivity, 0.97 specificity, 0.98 accuracy, 0.98 precision, and 0.97 F1- Score, respectively. The enhancement of the performance is 1.49% in accuracy over the other existing models. Lastly, the research addresses the grading of diabetic retinopathy by aligning the classification results with a standardized grading system, providing clinicians with accurate severity assessment for effective treatment decision.

Povzetek: Opisan je nov model VEnDR-Net za avtomatizirano odkrivanje diabetične retinopatije z uporabo glasovalnega ansambla CNN modelov. Model združuje ResNet50, VGG16, VGG19 in GoogLeNet za izboljšano točnosti in občutljivost.

1 Introduction

The human eye is an essential organ among the five senses of a human body, which grants us the invaluable gift of sight. Unfortunately, vision deterioration not only darkens the patient's world but also ushers in the threat of blindness. Diabetes is the core cause of several critical complications in various organs of the human body, including kidneys, heart, blood vessels and eyes. Diabetic Retinopathy is one of the chronic conditions of eye, caused by diabetes. Diabetic retinopathy (DR) is a diabetes-induced eye disorder that can lead to vision impairment or blindness [1]. DR progresses through different stages of severity, ranging from mild to severe and proliferative forms. Prolonged high blood sugar levels in diabetic patients can impair retinal vasculature, causing diabetic retinopathy (DR), a leading cause of vision loss [2]. The retina is essential for vision as it captures and transmits visual information to the brain. When these blood vessels are compromised due to diabetes then various changes can occur which lead to DR. This condition is characterized by a gradual deterioration of the retina.

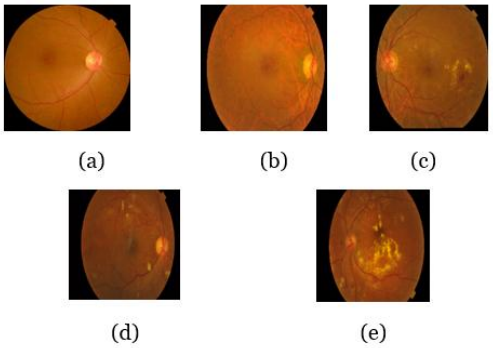

Figure 1: (a) No DR (b) Mild DR (c) Moderate DR (d) Severe NPDR (e) PDR

Factors such as the duration of diabetes, hereditary influences, and metabolic control are pivotal in the development of DR [3]. Figure 1 represents the five stages of DR. Initially, the arteries in the retina become weak and leaking starts which creates tiny hemorrhages. These leaky arteries often lead to the deposit of exudates (lipoproteins) in the retina.

Fragile new blood vessels may form, rupture, and leak into the eye, damaging the macula and hindering the retina's ability to transmit images to the brain, ultimately leading to vision loss. As a result, DR is a leading cause of blindness worldwide. The progression from diabetes to DR can take over a decade, making it a significant microvascular complication of diabetes mellitus globally [1]. While DR cannot be reversed, early diagnosis and appropriate treatment substantially lower the risk of vision impairment. Regular monitoring of diabetic patients is imperative for the early detection of DR. This process involves capturing fundus images of the eye, which are analyzed by an ophthalmologist to identify microaneurysms and exudates. Early detection of DR is critical to minimize the risk of blindness. Fundus images are used for identifying DR. Ophthalmologists visually examine a vast number of these images, making the process susceptible to errors, expensive, and time-consuming. The accuracy of interpreting retinal fundus images depends on the ophthalmologist's training, experience, and subjective criteria. Many cases of DR go undetected and missed during routine screenings, even by experts [4]. Advancements in technology, such as telemedicine and artificial intelligence-based image analysis, hold promises in enhancing the efficiency and accuracy of DR diagnosis and management. Deep learning has revolutionized the field of diabetic retinopathy (DR) by offering powerful capabilities in image analysis, diagnosis, screening, and management [3]. Convolutional neural networks (CNNs), a core component of deep learning, have shown outstanding performance in analyzing retinal fundus images for DR detection and classification. Moreover, deep learning techniques contribute to the early detection and monitoring of DR progression [4][5]. The aforesaid limitations and the advancement in healthcare technology and artificial intelligence motivated us to make a significant contribution to the field of diabetic retinopathy (DR), which are summarized as follows:

I. The research performs image preprocessing to eliminate noise and improve representative features from retinal fundus images.
II. The research explores and develops segmentation approach using CLAHE, Otsu thresholding approach and other morphological operations, for accurately segmenting the damaged areas in the retinal images. These segmentation techniques enable precise localization and isolation of the regions associated with DR, providing valuable insights for analysis and diagnosis.
III. The research introduces a novel voting ensemble deep neural network model that integrates the predictions of multiple learning algorithms using a soft voting weighted average technique.

2 Related work

Numerous methods to detect DR classification have been proposed. For multiclass classification, deep neural network has been used by the researchers.

Balasubramanian et al. [6] utilized median filtering for preprocessing, followed by super-pixels (SAS) for image segmentation. CNN was then applied to the segmented images, where features are extracted from segmented blood vessel regions by pooled hidden layers, shared parameters, and local connections. SVM is used to classify the extracted features, resulting in accuracy, specificity, and sensitivity of 97.43 %, 98.09%, and 97.12% respectively.

Kandhasamy et al. [7] employed two main methods. Initially, structural operations locate the clustered regions such as optical disk, exudates, hemorrhages, and microaneurysms. Subsequently, multiple-level set segmentation was employed, involving gradient computation, Neuman boundary conditions, and coefficient estimation using curvature central. Features were then extracted using Local Binary Pattern, color moments, and statistical features. These features were further refined using a generic algorithm (GA) before SVM classification, achieving accuracy: 97.3%, sensitivity: 97.14% and 98% specificity.

Wan et al. [8] using transfer learning approach via employing CNNs and tuning the hyperparameters. The different CNN architectures like AlexNet, VggNet, GoogleNet, and ResNet, applied and attained accuracy of 95.68.

Zhang et al. [9] suggested an automated diagnosis architecture for DR detection and grading i.e., Deep DR. CNN and DNN are combined with transfer learning and ensemble learning procedures and used high-quality DR image datasets. The optimal model achieved sensitivity: 97.5%, specificity: 97.7%, and AUC: 97.7%. Sensitivity of 98.1% and specificity of 98.8% are obtained by the grading system.

Table 1: Review of the recent state-of-art methods for detection of diabetic retinopathy

S. No.	**Model**	**Brief Summary & Limitations**	**Strengths of the proposed approaches**	**Performance Metrics (in terms of Accuracy)**
1.	CNN-Vision Mamba model [6]	✓ Used CNN-Vision Mamba model ✓ Test conducted on smaller dataset.	✓ Achieved good performance in terms of accuracy ✓ Enhances its applicability and flexibility in diverse clinical scenarios.	0.906
2.	Inception V3 [7]	✓ Applied Inception V3 model ✓ Applied augmentation method due to small volume of data.	✓ Obtained good accuracy making an applied approach therapeutically reliable.	0.9164
3.	ResViT FusionNet [8]	✓ Used ResNet and Vision ✓ Transformers for disease diagnosis. This combination increases the computational complexity of the system	✓ Achieved accuracy more than 93% by integrating CNN and Vision Transformers (ViTs) models.	0.9301
4.	Integrated the DenseNet model into a Raspberry Pi 4 [9]	✓ Applied DenseNet architecture ✓ Used small volume of data-set.	✓ Integrated the model with Raspberry Pi IV ✓ Disease indication was done by blinking a LED	0.88
5.	Residual-based deep neural network model [10]	✓ Introduced Residual based DNN for DR diagnosis ✓ However, evaluated the model with small and imbalanced data-set.	N/A	0.83
6.	Grid Search Cross Validation (GSCV) [12]	✓ Applied generic CNN model, and performed optimized hyperparameter tuning using GSCV.	✓ All the hyperparameter are computed and estimated with GSCV technique.	0.89
7.	Optimized Deep learning-based technique [13]	✓ Applied Cuckoo search for feature optimization.	✓ Improvement of accuracy using the given model is 10.46%.	0.97

Table 1 represents the review of some research papers. M. K. Jabbaz et al. [10] develed a customized model based on transfer learning and VGGNET-16 for recognition of diabetic retinopathy. Before building the suggested model, the authors employed several data augmentation techniques to improve the generalization capabilty of models.

N. Gundluru et al. [11] presented a model based on PCA and Harris hawk's optimization algorithm for feature selection and optimization, eventually leading to classification of DR disease using UCI machine learning repository dataset.

S. Rajmani et al. [12] used applied generic deep learning model and perform optimized hyoerparameter tuning using GSCV and achieved 89% accuracy. Zhang. Q. et. al. [13] followed the methodology wherein they combined preprocessing and segmentation approach before applying Cuckoo search for feature optimization and obtained commendable results with accuracy metrics surpassing 97.55%.

A multi-stage DL framework for DR grading is suggested by P.S. Silva et al. [14]. The suggested model includes a lesion segmentation followed by a classification, achieving an encouraging result in terms of accuracy. The results achieved on a large DR dataset demonstrate the potential of DL in accurately diagnosing DR. C. Mohanty et al. [15] developed a hybrid model combining VGG16 and XGBoost classifier; DenseNet 121 architecture and achieved the accuracy of 79.50%.

W. L. Alyoubi et. al. [16] applied CNN for classification and YOLOv3 for lesion localization. This methodology achieved 89% accuracy. M.K. Yaqoob et. al. [17] utilized ResNet-50 for feature extraction and Random Forest for classification and obtained 75.09% accuracy. G. Zhang et al. [18] developed a Multi-Model Domain Adaptation (MMDA) with transfer learning using weighted pseudo-labeling and clustering-based approaches and achieved 90.6% accuracy on APTOS dataset.

3 Methodology

This research aims to develop an automated system for the detection and classification of diabetic retinopathy through retinal fundus image analysis. This study includes three basic stages shown in Figure 2. i.e., image preprocessing, segmentation and classification. In the image preprocessing, images have gone through gray scale conversion and resizing processing technique.

The updated image is segmented by CLAHE, Otsu binary thresholding and adaptive gaussian thresholding segmentation techniques in the next stage. Third stage is classification, which includes four advanced deep learning models, specifically ResNet50, VGG16, VGG19 and GoogLeNet. Additionally, soft voting approach is applied to evaluate the decision level of classification of diabetic retinopathy. All the operations were performed on EyePACS dataset, provided by Kaggle.

3.1 Proposed approach

The research involves several key steps, starting with the loading of data from standard benchmark datasets consisting of retinal fundus images with annotations of DR severity. To ensure reliable training and evaluation of the classification models, the dataset is split into training and test sets based on the severity levels of DR.

To improve model performance, image preprocessing techniques are applied prior to training. After image preprocessing, segmentation approaches viz., Otsu thresholding and Morphological operations are employed to extract the affected regions from the retinal images. The proposed approach begins with performing image preprocessing. In this step, the image is converted into gray scale from RGB. Image is then resized to 512 X 512 from 1024 X 1024 to save space. In the next step, image segmentation is implemented using CLAHE followed by otsu binary thresholding technique. After this Adaptive Gaussian Thresholding is applied in small regions of image. Step 3 dealt with the soft voting ensemble learning. This approach ensembles four deep CNN models ResNet50, VGG16, VGG19, GoogLeNet. Algorithm represents the VEnDR-Net in detail.

The flow chart of the VEnDR-Net is presented in Figure 2. To further enhance the accuracy and reliability of the classification, a novel cascaded voting ensemble deep neural network model is constructed and developed. The model with four architecture is defined as,

$$\mathcal{E}soft = \{ResNet50, VGG16, VGG19, GoogLeNet\}$$

be the set of pre-trained models, each fine-tuned using the Fundus Images dataset (I, C); where I the set of N images, each of size, 512 × 512, and C contain the corresponding classes:

$$C = \{c \mid c \in \{Normal, Mild, Moderate, Severe, PDR\}\}$$

The training set (Itrain, Ctrain) is divided into mini batches, each of size n = 8, such that minibatches (Ij, Cj) $\in$ (Itrain, Ctrain), i = 1, 2, . . . N/n. The CNN model, e $\in$ $\mathcal{E}$soft, is iteratively optimized to minimize the empirical.

Algorithm: VotEnDR

Input: Set of Fundus Images (I, C); where C= {c | c ∈ {No DR, Mild DR, Moderate DR, Severe NPDR, PDR}}

Output: The trained model that classifies the s image i ∈ I

Step1: Perform Preprocessing

{

- Convert the image into gray scale from the other color space i.e., RGB by using:

  Gray Scale= 0.299 R + 0.587 G + 0.114 B

- Resize the image from 1024x1024 to 512x512 to save space and enhance visualization.

}

Step2: Implement Segmentation

{

- Apply CLAHE to eliminate artificial boundaries and limiting contrast to avoid amplifying noise in homogeneous areas.
- Calculates a threshold value from the image histogram, that sets at the middle of the peaks in the histogram using Otsu thresholding technique by using:

  $\sigma^2 b(T) = \omega_1(T)\ \omega_2(T)\ (\mu_1(T) - \mu_2(T))^2$

  Where, $\sigma^2 b(T)$ = Minimization of weighted variance of classes

- Apply Adaptive Gaussian Thresholding to calculates threshold values for small regions in the image by using:

  $\mathcal{T}h(x,y)=\sum(i,j)\in Nw(i,j)\ I(i,j)$

  Where, Th(x,y) = Threshold value of pixel at location (x,y)

}

Step3: Import the set of pre-trained models

$\mathcal{E}_{soft}$ = {ResNet50, VGG16, VGG19, GoogLeNet}

Fully Connected (FC) layer of each model = (5 X 1) dimension.

{**foreach** e ∈ $\mathcal{E}_{soft}$ do

α = 0.01

for epochs =1 to 50 **do**

foreach minibatch(I_j, C_j) ∈ (I_{train}, C_{train})**do**

Modify the parameters of the model e(*m*)

If validation error ! = improving for five epochs **then**

α = α X 0.01

endif

endforeach

endfor

endforeach}

{**foreach** i ∈ I_{test} **do**

Ensemble the output of every model, e ∈ $\mathcal{E}_{soft}$, using eq. (1)

endforeach}

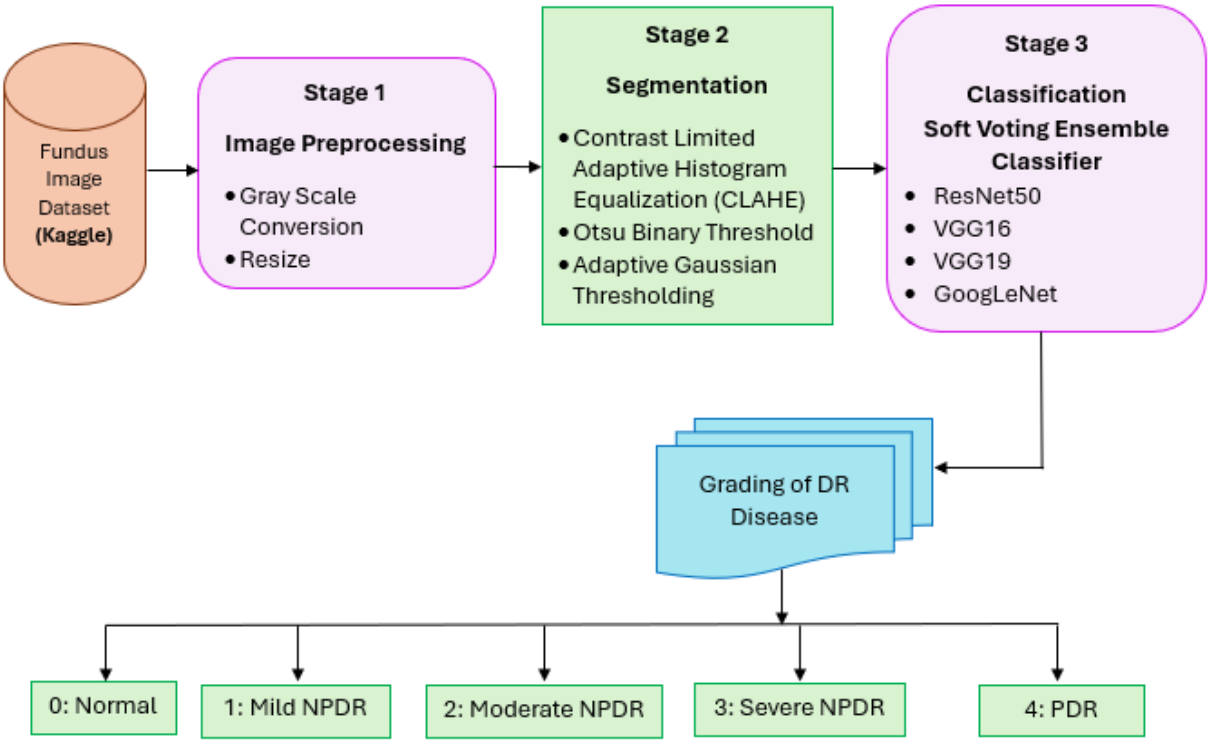

Figure 2: Structure of VEnDR-Net

loss:

$$L(m, Ij) = \frac{1}{n} \sum_{i \in Ij, c \in Cj} l(e(i, m), c) \ldots\ldots\ldots\ldots (1)$$

Where, e (x, m) is the CNN model, which predicts y for input x given m and l(.) is the categorial cross entropy loss penalty function.

3.2 Dataset used

In this study, the open-source dataset named, EyePACS has been used, which is publicly available on Kaggle. The dataset was developed in collaboration with the California-based EyePACS screening service. It contains total 88,000 high-resolution color fundus images with size ranging from 1,500 × 1,000 pixels to over 3,000 × 2,000 pixels.

Out of these 88,000 fundus images, 35126 retinal fundus images were utilized in our experiment to train and test the VEnDR-Net model. It comprises of 5 Classes namely, No DR, Mild DR, Moderate DR, Severe NPDR and PDR. The bifurcation of classes in the dataset is listed out in Table 2.

Table 2 illustrates the class imbalance in the original dataset provided by EyePACS. To mitigate this imbalance, both sampling and data augmentation techniques were employed. From the No DR class 3,500 images were selectively sampled from the total of 25,376 available. The images from the Mild DR and Moderate DR classes were utilized without modification. For the underrepresented Severe NPDR and PDR classes, data augmentation techniques—specifically rescaling and resizing—were applied to increase the image count from 1,013 to 2,500 and from 765 to 2,200, respectively. Out of these 16,171 images, 80% (i.e., 12937 images) were used for training, while 20% (i.e., 3234) of images were used for testing purposes.

Table 2: Class distribution of EyePACS dataset

Class Label	Stage of DR/ Classes	Image Count	Image count after Down/Up Sampling
0	No DR	25376	3500
1	Mild DR	2495	2495
2	Moderate DR	5476	5476
3	Severe NPDR	1013	2500
4	PDR	765	2200

3.3 Image preprocessing

Image preprocessing converts raw image data into an organized and informative representation, facilitating subsequent processing and feature extraction.

The goal is to enhance image quality, extract useful information, and make images suitable for further analysis or interpretation [7, 8, 9]. The primary goal is to augment the data by down-sampling for No DR Class and up- sampling for Severe NPDR and PDR classes. Figure 3. represents the results of the preprocessing techniques. The following two preprocessing techniques are applied in sequence to produce enhanced results:

- **Gray scale conversion:** gray scale conversion is a low-level image preprocessing operation that transforms a multichannel color image (typically in the RGB color space) into a single-channel intensity image. From a computational perspective, this operation significantly reduces data dimensionality, enabling faster processing, lower memory usage, and often more efficient learning in computer vision models, especially convolutional neural networks (CNNs). In clinical contexts, RGB retinal fundus images are frequently converted to gray scale to focus on structural information such as blood vessels, optic disc boundaries, and retinal lesions. The most widely used gray scale conversion method is based on the ITU-R BT.601 standard, which approximates human visual luminance sensitivity using equation (2) represents the gray scale conversion.

$$\text{Gray Scale} = 0.299\,R + 0.587\,G + 0.114\,B \quad \ldots.(2)$$

This perceptually weighted transformation helps preserve brightness and contrast features critical for detecting subtle pathologies in fundus image. Nevertheless, gray scale conversion may not always be optimal. Certain pathological indicators, such as microaneurysms or hemorrhages, exhibit chromatic characteristics better captured in RGB or transformed color spaces (e.g., LAB, HSV). Thus, selective gray scale conversion, hybrid channel processing, or even multi-modal approaches are being integrated into modern deep learning pipelines to balance interpretability and performance.

- **Resizing:** In this model, images are resized from 1024x1024 to 512x512 to save space and enhance visualization. Its primary purpose is to normalize input dimensions, ensuring uniformity across diverse datasets and compatibility with network architectures that require fixed-size input tensors. From a computational standpoint, resizing enables memory-efficient training, batch processing, and streamlined deployment on GPU-based infrastructures. It directly impacts the spatial resolution at which the model learns to identify features, which is especially critical in detecting small-scale anomalies such as microaneurysms or haemorrhagic spots in fundus images. Image resizing is achieved using interpolation algorithms that estimate intensity values at new pixel locations. In DR detection using fundus images, resizing must preserve critical diagnostic features like microaneurysms and vessel structures. Excessive downscaling risks losing small but clinically significant details, while upscaling can introduce artifacts or noise. Thus, the choice of target resolution and interpolation technique should be aligned with the model's receptive field and the visual scale of diagnostic features. Resizing serves as a normalization step, helping to control variation in input dimensions due to differing acquisition devices or protocols. These methods attempt to minimize information loss in critical areas while still standardizing input dimensions. This standardization enhances the robustness and generalizability of models deployed across multiple clinical settings. This streamlined deployment on GPU-based infrastructures. It directly impacts the spatial resolution at which the model learns to identify features, which is especially critical in detecting small-scale anomalies such as microaneurysms or hemorrhagic spots in fundus images. Image resizing is achieved using interpolation algorithms that estimate intensity values at new pixel locations. In DR detection using fundus images, resizing must preserve critical diagnostic features like microaneurysms and vessel structures. Excessive downscaling risks losing small but clinically significant details, while upscaling can introduce artifacts or noise.

Thus, the choice of target resolution and interpolation technique should be aligned with the model's receptive field and the visual scale of diagnostic features. resizing serves as a normalization step, helping to control variation in input dimensions due to differing acquisition devices or protocols. These methods attempt to minimize information loss in critical areas while still standardizing input dimensions. This standardization enhances the robustness and generalizability of models deployed across multiple clinical settings.

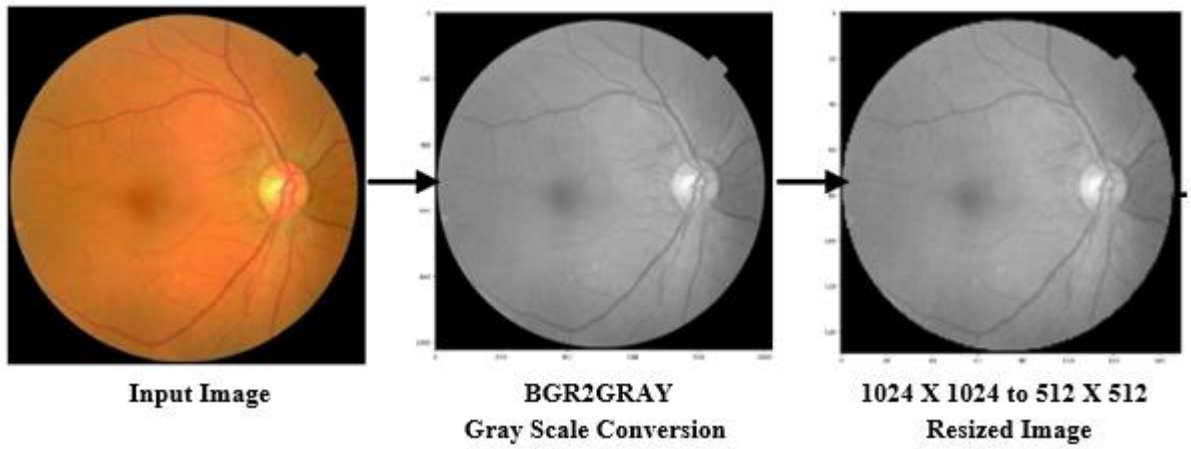

Figure 3: Image after preprocessing

3.4 Segmentation

Image segmentation divides a digital image into subgroups (called Image Objects) to reduce complexity, making analysis simpler [19]. Figure 4 represents the changes after multiple segmentation techniques. The segmentation techniques used in this model are:

Contrast limited adaptive histogram equalization (CLAHE): CLAHE enhances local contrast in an image, improving the visibility of fine-grained structures such as blood vessels, lesions, and optic disc boundaries—features crucial for diabetic retinopathy detection and grading. CLAHE operates by dividing the input image into non-overlapping contextual regions (tiles) and applying histogram equalization independently within each tile. Unlike global histogram equalization, which may over-amplify noise in homogeneous areas, CLAHE limits the amplification by clipping the histogram at a predefined threshold (contrast limit), thus preventing over-enhancement.

Bilinear interpolation is used to combine neighboring tiles, eliminating artificial boundaries and limiting contrast to avoid amplifying noise in homogeneous areas [20]. Since contrast enhancement is applied independently to each tile, there may be visible discontinuities at tile boundaries. Interpolation techniques are used to smooth out these transitions and create a visually consistent result across the entire image. CLAHE preprocessing improves the performance of CNNs and other machine learning models by amplifying subtle features like microaneurysms and hemorrhages that may otherwise be obscured [21]. Its contrast-limiting capability, tile-based processing, and adaptability to local context make it especially suitable for enhancing diagnostic features in retinal images while maintaining robustness against noise amplification.

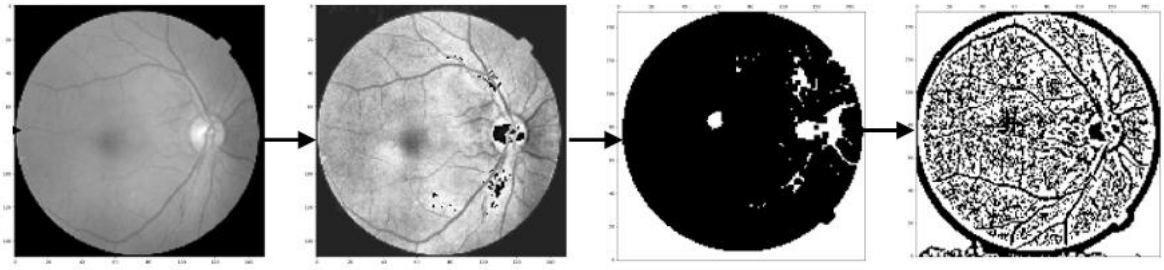

Figure 4: Conversion of updated image after segmentation

- **Otsu threshold:** Otsu's thresholding is a widely adopted binarization method used to segment regions of interest by converting gray scale images into binary form. It is particularly useful for isolating anatomical structures such as blood vessels, optic discs, or lesions in retinal images. Otsu's method assumes that an image contains two classes of pixels—foreground and background—and calculates the optimum threshold that minimizes the intra-class variance (or equivalently, maximizes the inter-class variance) between these two groups [22]. The process involves computing a histogram of image intensity levels and exhaustively searching for the threshold through equation (3) that yields the minimum weighted within-class variance:

$$\sigma_\omega^2 (T) = \omega_1(T)\, \omega_2(T)\, (\mu_1(T) - \mu_2(T))^2 \quad \text{.....(3)}$$

Where ω = class probabilities
σ^2 = class variances for threshold T
The optimal threshold T^* is selected by minimizing σ_ω^2

It effectively delineates vascular structures and exudates, forming a baseline segmentation step for more complex algorithms. Otsu's thresholding remains a robust and interpretable technique for preliminary segmentation in medical imaging.

- **Adaptive Gaussian Thresholding:** Adaptive Gaussian Thresholding is a sophisticated binarization technique tailored for complex medical images characterized by spatially varying lighting conditions, contrast levels, and structural intricacies—conditions commonly encountered in modalities like retinal fundus photography, X-rays, and dermoscopic imaging. Unlike global methods i.e, Adaptive Gaussian Thresholding performs **local thresholding**, computing a unique threshold for each pixel based on the weighted average of its neighboring pixels [23]. The threshold value is derived as a weighted sum (Gaussian window) of the

neighborhood area. The mathematical formulation for the threshold $\mathcal{T}h(x,y)$ at a pixel is:

$$\mathcal{T}h(x,y)=\sum(i,j)\in Nw(i,j)\ I(i,j) \quad \text{.....(4)}$$

Where, Th(x,y) = Threshold value of pixel at location (x,y) is shown in equation (4).

Adaptive Gaussian Thresholding is used to enhance segmentation accuracy in retinal image analysis, lesion boundary detection, and morphological structure extraction.

Adaptive Gaussian Thresholding technique's integration into preprocessing pipelines supports hybrid workflows [24]. For instance, threshold maps generated through this method can act as spatial priors in ensemble learning networks, serve as inputs to edge detectors, or form masks for texture analysis.

3.5 Ensemble learning

Ensemble learning is based on the principle that multiple diverse models, when aggregated, can correct each other's weaknesses. The Ensemble Classifier is a meta-classifier that integrates similar or diverse ML classifiers, using majority or plurality voting for classification. This voting classifier functions as an ML model that is learned by training on a collection of multiple models [25]. It determines the output class with the highest predicted probability among the models. From a computational learning theory standpoint, ensemble learning is grounded in the **bias–variance–noise decomposition** of prediction error. Base models trained on the same or diverse data subsets can capture distinct

hypothesis spaces. Their aggregation, under the right conditions, leads to error reduction, formalized by:

$$\text{Error}_{\text{ensemble}} = \bar{\text{Error}} - \text{Diversity}$$

To determine the final output class, the classifier aggregates the results from each model in the ensemble and selects the class that receives the largest majority of votes. The base strategies of ensemble learning are bagging, boosting and stacking. The fundamental mechanism behind this is how voting works in general [26]. Figure 5 illustrates the proposed stacking ensemble.

By combining predictions with multiple models, the ensemble classifier reduces the risk of an individual model making an inaccurate prediction. In cases where one model might predict incorrectly, other models in the ensemble can compensate by providing the correct prediction. This ensemble approach improves estimator robustness and reduces the risk of overfitting, making the classifier more reliable overall. As a meta-model, the ensemble classifier can be used with a wide range of existing trained machine learning models, without requiring these models to be aware that they are part of the ensemble.

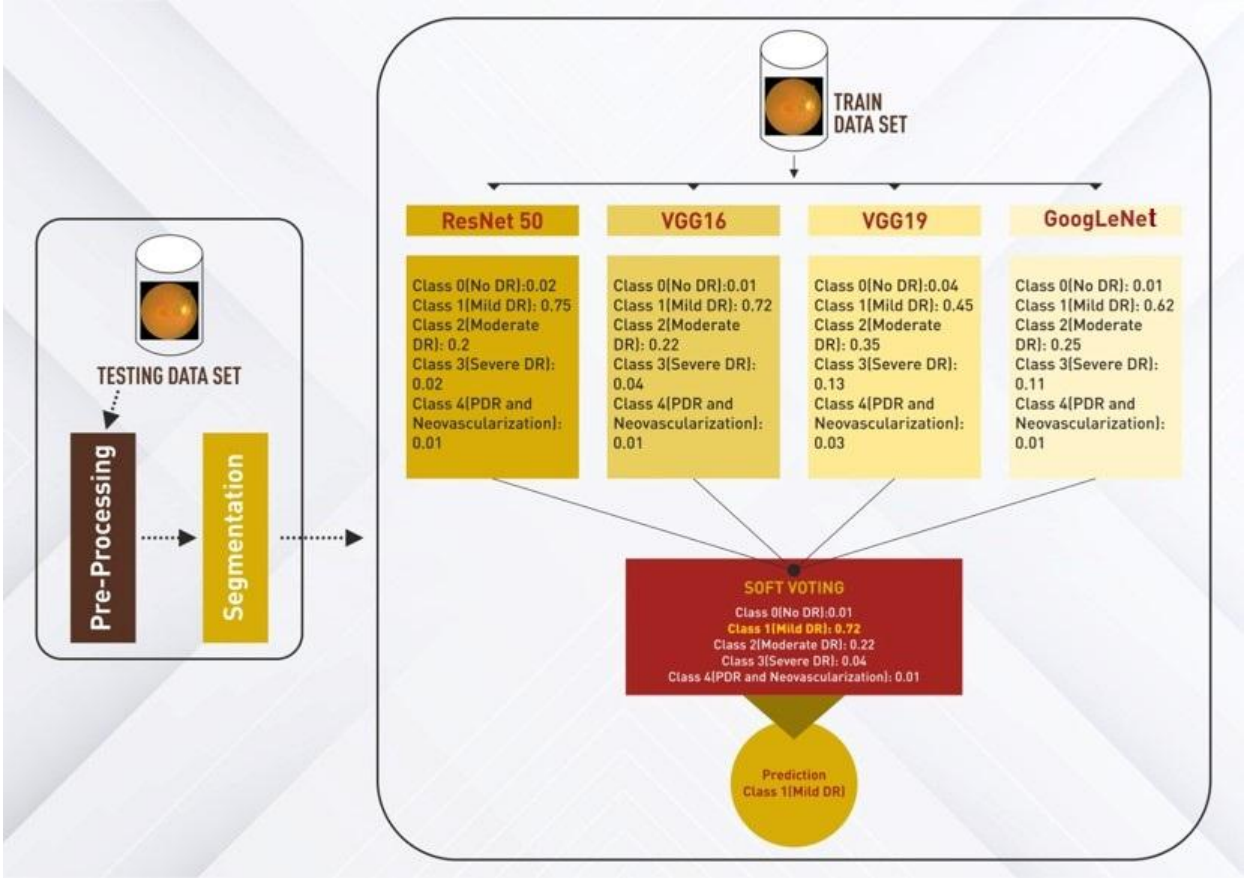

Figure 5: Structure of VEnDR-Net

This allows for seamless integration of various models into a unified framework [27][28]. The Ensemble Classifier implements two types of voting mechanisms: "hard" voting and "soft" voting.

- **Hard Voting:** In hard voting, the predicted output class is determined by the class that receives the highest majority of votes. Each classifier in the ensemble predicts a class label, and the class

predicted by the majority is chosen as the final output. Hard voting simply picks the class with the highest number of votes.

- **Soft Voting:** In soft voting, instead of considering just the final predictions from each classifier, the probabilities associated with each prediction are taken into account. The probabilities of each class are averaged across all models, and the class with the highest total probability is selected as the final prediction. The results produced by other state of art methods for the same problem is provided in Table 4.

Model has been provided by a new test sample, x_{test}, the ensemble output is determined in equation (5) as follows:

$$w^* = \arg\max_{w}\left(\frac{\sum_{ve \in \varepsilon soft} e(m,xtest)}{|\varepsilon soft|}\right) \quad \ldots(5)$$

The scores (or probabilities) of all the models, are averaged in a soft-voting ensemble and compared to a threshold [29]. The aggregate results by averaging the base scores should be considered. It's the median used instead of the mean, as it's less sensitive to outliers, so it will usually represent the underlying set of outputs better than the mean.

A soft-voting ensemble calculates the average score (or probability) and compares it to a threshold value. The aggregate results by averaging the base scores should be considered.

It's the median used instead of the mean, as it's less sensitive to outliers, so it will usually represent the underlying set of outputs better than the mean.

$$f_1(x) = \begin{bmatrix} 0.71 \\ 0.09 \\ 0.05 \\ 0.15 \end{bmatrix} \quad f_2(x) = \begin{bmatrix} 0.43 \\ 0.25 \\ 0.2 \\ 0.12 \end{bmatrix} \quad f_3(x) = \begin{bmatrix} 0.51 \\ 0.29 \\ 0.17 \\ 0.03 \end{bmatrix} \quad \ldots(6)$$

To combine them, take average the vectors element-wise:

$$\frac{1}{3}\begin{bmatrix} 0.71 \\ 0.09 \\ 0.05 \\ 0.15 \end{bmatrix} + \begin{bmatrix} 0.43 \\ 0.25 \\ 0.2 \\ 0.12 \end{bmatrix} + \begin{bmatrix} 0.51 \\ 0.29 \\ 0.17 \\ 0.03 \end{bmatrix} = \begin{bmatrix} \mathbf{0.55} \\ 0.21 \\ 0.14 \\ 0.1 \end{bmatrix} \quad \ldots.(7)$$

The soft voting ensemble classifier significantly improves predictive modeling by mitigating large errors or misclassifications made by individual models.

Figure 6 shows the confusion matrix for VEnDR-Net model and Figure 7 represents Confusion matrix for VenDR- Net Model for Test set (20% image) only.

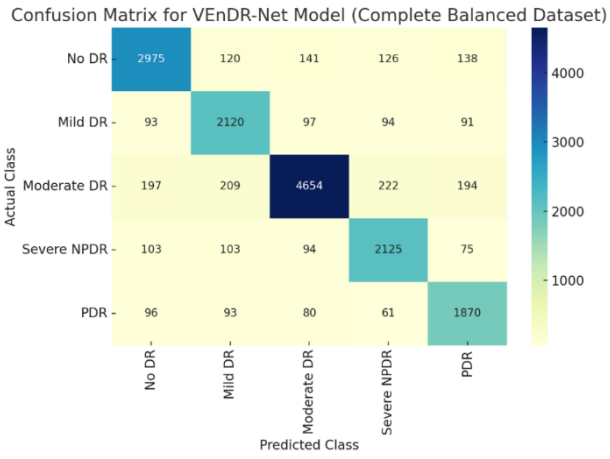

Figure 6: Confusion Matrix for complete balanced data

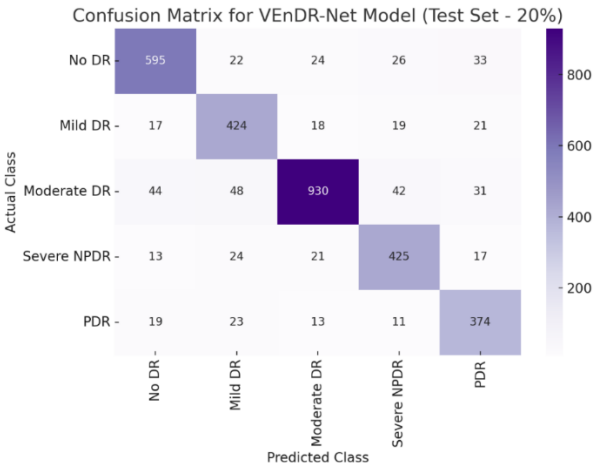

Figure 7: Confusion matrix for test set

The soft voting ensemble classifier significantly improves predictive modeling by mitigating large errors or misclassifications made by individual models. The research combines four CNN models (ResNet50, VGG16, VGG19, and GoogLeNet), with each model assigning probabilities to each class. The final prediction is determined by averaging the probabilities provided by the CNN models.

3.6 Experimental setup

The simulation settings for grading in diabetic retinopathy (DR) replicate the process of assessing the severity of DR using retinal images. The proposed **VenDR-Net** model for Diabetic Retinopathy detection and classification was developed and evaluated on a high-performance computing environment with the following specifications of Hardware shown in Table 3:

Table 3: Hardware configuration

S. No.	Hardware Used	
1.	**Processor**	Intel Core i7, 9th Generation
2.	**RAM**	16 GB DDR4
3.	**Graphics Processing Unit (GPU)**	NVIDIA GeForce GTX 1650 with 4 GB VRAM
4.	**Storage**	Solid State Drive (SSD) for faster data access and model training

This configuration provided sufficient computational resources to support efficient training of deep learning models, including image processing, feature extraction, and classification tasks. Table 4 represents the software environment used for the experiment. The VenDR-Net model was trained using GPU acceleration to reduce computation time and enhance model efficiency. The training process included image preprocessing, feature learning, and classification, with model evaluation performed using **10-fold cross-validation** to ensure consistency and robustness across data partitions [30]. The VenDR-Net model was trained using GPU acceleration to reduce computation time and enhance model efficiency

Table 4: Software environment

S. No.	Software Used	
1.	**Operating System**	Windows 10
2.	**Programming Language**	Python 3.8
3.	**Deep Learning Framework**	TensorFlow 2.x
4.	**Additional Libraries**	NumPy, OpenCV, Scikit-learn, Matplotlib, Pandas, and imbalanced-learn
5.	**CUDA Toolkit and cuDNN**	Configured to enable GPU acceleration for training and inference

Table 5: Tuning of hyperparameter

Hyperparameter	**Description**	**ResNet50**	**VGG16**	**VGG19**	**GoogLeNet**
Frozen layers	Weights are **not updated** during the training process on a new dataset	First 140 Layers	First 15 Convolutional layers	First 16 Convolutional layers	Initial Inception modules
Unfrozen layers	Weights are **trainable**; that is, updated during the training process using backpropagation.	Final layers incl. FC	Fully connected layers	Fully connected layers	Final layers incl. FC
Learning Rate	The step size for updating the model parameters during training	0.01	0.01	0.0001	0.001
Number of Epochs	The number of times the entire dataset is passed through the model during training	50	50	40	40
Batch Size	The number of samples used in each training iteration	32	32	32	32
Dropout Probability	The probability of dropping out a neuron during training	0.5	0.2	0.5	0.2
Weight Decay	Regularization parameter to prevent overfitting	0.01	0.01	0.01	0.01
Activation Function	The activation function used in the neural network layers	Softmax	Softmax	Softmax	Softmax
Optimizer	The optimization algorithm used during training	Adam	Adam	Adam	Adam

The training process included image preprocessing, feature learning, and classification, with model evaluation performed using **10-fold cross- validation** to ensure consistency and robustness across data partitions [30].

Table 5 represents the tuning of hyperparameter for ensemble base deep learning model i.e., VEnDR-Net model.

4 Results

The simulation settings for grading in diabetic retinopathy (DR) replicate the process of assessing the severity of DR using retinal images. The confusion matrix in Figure 6 provides a detailed overview of the model's classification performance, which was trained using a dataset that classifies DR into five distinct labels: "No DR," "Mild DR," "Moderate DR," "Severe NPDR," and "PDR" [31].

4.1 Evaluation parameters

This study uses the following parameters to measure the performance of VenDR-Net [32 - 33].

(a) Accuracy: Accuracy calculates the ratio of correctly predicted count to the total number of predictions. Mathematically, accuracy is expresssed as:

$$\text{Accuracy}=\frac{\text{Number of Correct Predictions}}{\text{Total Number of Predictions}} \quad \ldots\ldots\ldots(8)$$

In a binary classification problem with only two classes (such as positive and negative), accuracy is computed using the following formula:

$$\text{Accuracy} = \frac{\text{True Positives}+\text{True Negatives}}{\text{True Positives}+\text{False Positives}+\text{True Negatives}+\text{False Negatives}} \quad .(9)$$

$$\text{Accuracy}=\frac{TP+TN}{TP+FP+TN+FN} \quad \ldots\ldots\ldots\ldots..(10)$$

(b) Precision: Precision is defined as the ratio of correctly classified positive instances (True Positives) to the total number of instances predicted as positive (the sum of True Positives and False Positives). It measures the accuracy of the model's positive predictions, indicating the proportion of correctly identified positive cases. Mathematically, precision is calculated as:

$$\text{Precision} = \frac{\text{True Positive}}{\text{True Positive + False Positive}} \quad \ldots\ldots\ldots\ldots(11)$$

$$\text{Precision} =\frac{TP}{TP+FP} \quad \ldots\ldots\ldots\ldots..(12)$$

(c) Recall: Recall is defined as the ratio of correctly classified positive instances (True Positives) to the total number of actual positive instances (the sum of True Positives and False Negatives). It measures the model's ability to correctly detect positive samples, with higher recall indicating a greater number of correctly identified positives.

$$\text{Recall} = \frac{\text{True Positive}}{\text{True Positive + False Negative}} \quad \text{…….(13)}$$

$$\text{Recall} = \frac{TP}{TP+FN} \quad \text{……(14)}$$

(d) F1- Score:

The F1-score is defined as the harmonic mean of precision and recall, providing a balanced measure that considers both metrics simultaneously. The harmonic mean is a type of average calculated by taking the reciprocal of the average of the individual values, making it particularly suitable when balancing two competing metrics.

$$\text{F1-Score} = \frac{\text{2X Precision X Recall}}{\text{Precision+Recall}} \quad \text{…...(15)}$$

4.2 Evaluated results

The provided Table 6 presents the values of various performance measures of every CNN architecture used in the proposed model as well as of VEnDR-Net. The table shows VEnDR-Net gives much better results of precision, recall, F1-score and accuracy among all CNN architectures used for the proposed model via ensemble deep learning model i.e., VEnDR-Net model.

Table 6: Performance of individual classification models and VEnDR-Net model

S. No.	Model	Precision	Recall	F1-Score	Accuracy
1	ResNet50	0.95	0.94	0.94	0.95
2	VGG16	0.96	0.96	0.96	0.95
3	VGG19	0.96	0.96	0.96	0.96
4	GoogLeNet	0.95	0.96	0.96	0.97
5	**VEnDR-Net**	**0.98**	**0.97**	**0.97**	**0.98**

To evaluate the computational cost associated with the proposed VEnDR-Net model comprising **ResNet50**, **VGG16**, **VGG19**, and **GoogLeNet**, a relative cost analysis was conducted based on model complexity, hyperparameters, and execution environment. All experiments were performed on a system equipped with an **Intel Core i7 (9th Gen) processor**, **16 GB DDR4 RAM**, and an **NVIDIA GeForce GTX 1650 GPU** (4 GB VRAM), running **Windows 10**. The software stack included **Python 3.8**, **TensorFlow 2.x**, **CUDA Toolkit**, and **cuDNN**, along with essential libraries such as NumPy, Scikit-learn, OpenCV, and imbalanced-learn.

Table 7 represents the calculation of estimated relative computational cost of VEnDR-Net Model. Training cost was estimated using a relative unit-based scoring method, considering the number of model parameters (as a proxy for floating point operations), number of training epochs, and learning rate sensitivity.

For normalization, GoogLeNet was assigned a base complexity factor of 1. ResNet50 (25.6M parameters) trained for 50 epochs with a learning rate of 0.01 yielded a cost of **100 units**. VGG16 (138M parameters) required **200 units** under similar conditions, while VGG19 (144M parameters) trained for 40 epochs at a lower learning rate (0.0001) resulted in a higher cost of **216 units** due to slower convergence. GoogLeNet, with only 6.8M parameters and moderate settings, incurred the lowest training cost at **40 units**. The **total relative training cost** for the ensemble was thus calculated to be **556 units**.

Table 7 : Estimated relative computational cost of ensemble models

S. No.	Model	Parameters (Millions)	Epochs	Learning Rate	Computational Cost
1.	ResNet50	25.6	50	0.01	50 X 2.0 X 1.0 = 100
2.	VGG16	138	50	0.01	50 X 4.0 X 1.0 = 200
3.	VGG19	144	40	0.0001	40 X 4.5 X 1.2 = 216
4.	GoogLeNet	6.8	40	0.001	40 X 1.0 X 1.0 = 40
	Total	--	--	--	**556 Units**

For inference, the ensemble's per-sample cost was approximated by summing normalized relative complexities of each model. The computed inference cost per prediction was **11.5 units**, reflecting the cumulative burden of executing all four models sequentially, shown in Table 8. Due to the VRAM limitation of the GTX 1650, it is recommended to adopt a staged model loading strategy during inference or utilize model quantization to reduce runtime memory usage. This estimation underscores the trade-off between accuracy and computational overhead inherent in VEnDr-Net model, especially when deployed on resource-constrained hardware.

Table 8 : Relative inference cost per sample

S. No.	Model	Relative Inference Cost (Units)
1.	ResNet50	2.0
2.	VGG16	4.0
3.	VGG19	4.5
4.	GoogLeNet	1.0
5.	**VEnDR-Net**	**3.0**

To assess the classification performance of deep learning models for diabetic retinopathy (DR) detection, ROC and AUC curves were generated. The models compared include **ResNet50, VGG16, VGG19, GoogLeNet**, and the proposed **VEnDR-Net**.

Figure 8 show the ROC curve, which represents the trade-off between the true positive rate (recall) and the false positive rate. The AUC (Area Under the Curve) shown by figure 9, quantifies the overall ability of the model to discriminate between positive and negative classes.

- VEnDR-Net achieved the highest AUC score (≈ 0.98), indicating excellent predictive performance.
- Other models such as GoogLeNet and VGG19 also performed well, with AUC values around 0.96.
- All models significantly outperformed the baseline (random classifier), which has an AUC of 0.50.

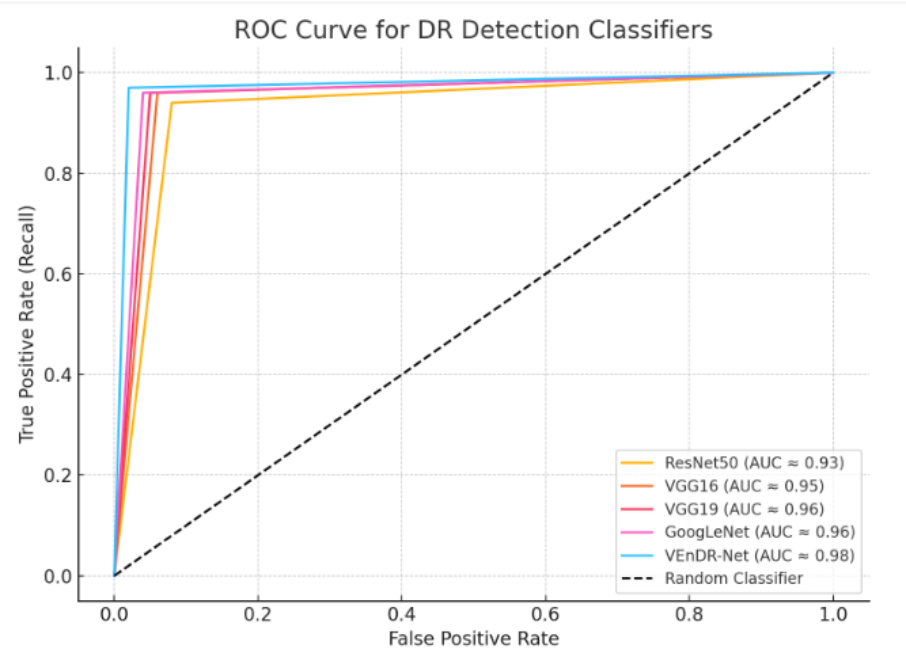

Fig 8: ROC Curve for DR detection

These results demonstrate that VEnDR-Net provides superior sensitivity and specificity compared to traditional CNN architectures, making it a promising approach for clinical DR screening.

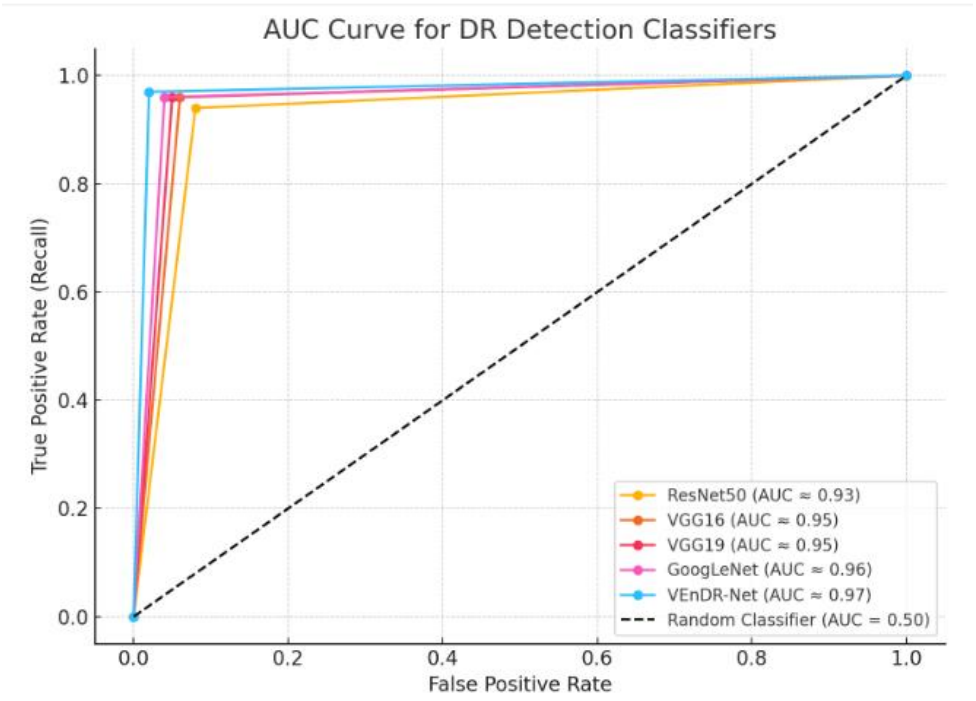

Fig 9: AUC Curve for DR detection

Table 9 represents the precision, recall and accuracy obtained by VEnDR-Net for each class. The precision, recall and accuracy achieved by VEnDR-Net are 98.46, 96.91 and 98.464 respectively.

Table 9: Performance measure of VEnDR-Net for each class

S. No.	Stages of DR	Precision	Recall	Accuracy
1	No DR	98.12	98.83	98.98
2	Mild DR	97.23	97.78	98.54
3	Moderate DR	98.56	95.69	98.58
4	Severe NPDR	98.34	96.94	98.26
5	PDR	97.58	95.31	97.96
Average		97.96	96.91	98.46

A comparative analysis of several methods or models based on different performance metrics is presented in Table 10. VEnDR-Net outperforms the other existing models in precision and accuracy as shown in Figure 10.

This proposed model VEnDR-Net enhance the accuracy and achieve the improvement of 1.49% from the existing models.

Table 10: Performance comparison of VEnDR-Net with existing models

Paper	**Precision**	**Recall**	**F1-Score**	**Accuracy**
Zhuang Ai et al. (2021) [35]	0.93	0.92	0.92	0.92
ZUBAIR KHAN et al. (2021) [36]	0.67	0.56	0.61	0.83
Muhammad Kashif Jabbar et al. (2022) [10]	0.98	0.95	0.97	0.96
Ling Dai et al. (2021) [38]	0.95	0.93	0.94	0.95
Abràmoff et al. (2020) [2]	0.96	0.96	0.97	0.97
Zhao et al. (2020) [4]	0.96	0.96	0.96	0.96
Silva et al. (2021) [14]	0.97	0.97	0.97	0.97
VEnDR-Net (Proposed Model)	**0.98**	**0.97**	**0.97**	**0.98**

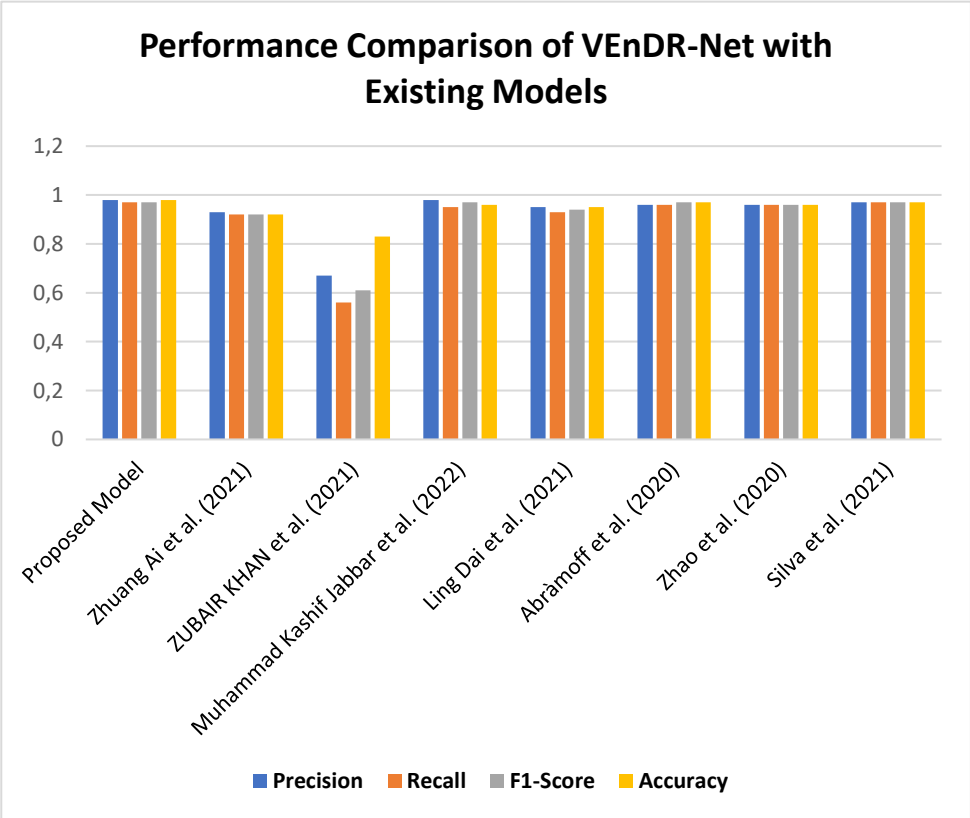

Figure 10: Performance comparison of VEnDR-net with existing models

The implementation of deep learning models in real-world medical environments demands more than just high predictive accuracy, which necessitates reliability, interpretability, and adaptability [34]. The proposed VEnDR-Net, a high-performance convolutional neural network designed for Diabetic Retinopathy (DR) detection, offers strong potential for integration into Clinical Decision Support Systems (CDSS) to enhance diagnostic accuracy and streamline ophthalmic workflows.

5 Discussions

The proposed **VEnDR-Net** architecture exhibits competitive and, in several aspects, superior performance when compared to existing state-of-the-art methodologies such as CNN-XGBoost. CNN-XGBoost leverage handcrafted features followed by tree-based ensemble classifiers. This classifier is quite useful in some of the cases, especially, where the data is less complex. However, CNN-XGBoost has limited scalability. In contrast, VEnDR-Net adopts an end-to-end deep ensemble strategy utilizing soft voting across multiple convolutional neural network (CNN) backbones, thereby enhancing model robustness and minimizing the dependency on feature engineering.

Moreover, the traditional approaches including CNN-XGBoost method frequently demand substantial annotated data and task-specific alignment, which limits their practical application in real-world clinical environments, while VEnDR-Net circumvents this limitation by maintaining high diagnostic accuracy across diverse datasets without the need for the annotated data.

6 Conclusion

This research focuses on the development of deep learning model ensembles that employ a soft-voting approach to accurately classify the stages of diabetic retinopathy (DR). Dataset of EyePACS from Kaggle is used in the research which is providing dataset of 5 classes or stages of DR. Both positive and negative cases for DR detection have been evaluated with excellent accuracy, as the model demonstrated high values in sensitivity, specificity, accuracy, precision, and F1-score. This research approach introduces new ways to assess DR classification and provides the latest techniques for refining the classification and grading of DR. The proposed approach strengthens model robustness, refines feature discrimination, and delivers high classification accuracy using a soft voting ensemble of deep neural networks. This approach utilizes a voting ensemble learning model, which incorporates four different CNN architectures: ResNet50, VGG16, VGG19, and GoogLeNet. The integration of these architectures through soft voting achieves an impressive accuracy of 0.98464. The evaluated results show that the VEnDR-Net classifies diabetic retinopathy with enhanced performance metrics. A limitation of this research is the high computational cost involved in constructing ensemble models, which require training several models and amalgamating their outputs. Despite this limitation groundbreaking research strives to push the boundaries of DR diagnosis and treatment strategies, offering significant implications for both patients and healthcare professionals.

7 Limitations and future work

As Deep Learning is rapidly progressing, diabetic retinopathy (DR) detection has seen significant improvements in recent years. However, this study dealt with DR detection only using VEnDR-Net model, there remains considerable scope for enhancement. Emerging architectures such as Vision Transformers (ViT), EfficientNet, and hybrid CNN-transformer models offer improved representation learning capabilities. Their integration could lead to more accurate and robust DR detection systems. Implementing federated learning can enable collaborative model training across multiple healthcare institutions without sharing sensitive patient data, thereby enhancing model generalization while ensuring data privacy. Incorporating explainability techniques such as Grad-CAM, SHAP, or attention visualization will help clinicians understand the reasoning behind model predictions. This fosters trust and promotes adoption in clinical practice. Combining fundus images with patient metadata (e.g., age, HbA1c levels, blood pressure) or other imaging modalities like OCT can lead to a more comprehensive and accurate diagnosis of DR. Developing lightweight, real-time models optimized for edge devices or mobile platforms can facilitate affordable DR screening in rural and under-resourced regions. To ensure clinical applicability, future models should undergo rigorous validation in real-world healthcare environments. Collaboration with medical professionals and clinical trials will be critical in assessing reliability and bias reduction.

Acknowledgement

I would like to express my sincere gratitude to Dr. Mrinal Pandey for their invaluable guidance, support, and encouragement throughout the course of this research. I

am also thankful to Manav Rachna University, Faridabad for providing the necessary resources and facilities to carry out this work

References

[1] American Diabetes Association. (2021). Diabetic retinopathy. Diabetes Care, 44(Supplement 1), S162-S165. doi: 10.2337/dc21-S014

[2] Abràmoff, M. D., Lavin, P. T., Birch, M., Shah, N., Folk, J. C., & Pollack, S. (2020). Pivotal trial of an autonomous AI-based diagnostic system for detection of diabetic retinopathy in primary care offices. NPJ Digital Medicine, 3(1), 1-10. doi: 10.1038/s41746-020-0299-1

[3] Wang, Y. L., Yang, J. Y., Yang, J. Y., Zhao, X. Y., Chen, Y. X., & Yu, W. H. (2021). Progress of artificial intelligence in diabetic retinopathy screening. *Diabetes/Metabolism Research and Reviews*, *37*(5), e3414. https://doi.org/10.1002/dmrr.3414

[4] Zhao, J., Wong, H. T., Lao, J., Zhen, Z., Chen, X., Zhu, J., ... Wong, D. W. K. (2020). Classification of diabetic retinopathy images using multi-modal deep learning approach. Frontiers in Medicine, 7, 590262. doi: 10.3389/fmed.2020.590262

[5] Gao, Y., Zhang, J., Chen, S., & Wang, Q. (2021). Multi-scale feature fusion framework for grading diabetic retinopathy based on fundus images. Pattern Recognition, 116, 107963. doi: 10.1016/j.patcog.2021.107963

[6] Nguyen, T. T., Yang, E., Yoon, S., & Kang, M. G. (2023). Multi-task deep learning for diabetic retinopathy detection and severity grading. Pattern Recognition Letters, 157, 37-44. doi: 10.1016/j.patrec.2021.07.027

[7] Rosas-Romero, R., Martínez-Carballido, J., Hernández-Capistrán, J., & Uribe-Valencia, L. J. (2015). A method to assist in the diagnosis of early diabetic retinopathy: Image processing applied to detection of microaneurysms in fundus images. *Computerized medical imaging and graphics*, *44*,41-53. https://doi.org/10.1016/j.compmedimag.2015.07.001

[8] Bhatti, E., & Kaur, P. (2019). DRAODM: Diabetic retinopathy analysis through optimized deep learning with multi support vector machine for classification. In *Recent Trends in Image Processing and Pattern Recognition: Second International Conference, RTIP2R 2018, Solapur, India, December 21–22, 2018, Revised Selected Papers, Part II 2* (pp. 174-188). Springer Singapore. https://doi.org/10.1007/978-981-13-9184-2_16

[9] Mule, D. B., Chowhan, S. S., & Somwanshi, D. R. (2019). Detection and classfication of non-proliferative diabetic retinopathy using retinal images. In *Recent Trends in Image Processing and Pattern Recognition: Second International Conference, RTIP2R 2018, Solapur, India, December 21–22, 2018, Revised Selected Papers, Part II 2* (pp. 312-320). Springer Singapore. https://doi.org/10.1007/978-981-13-9184-2_28

[10] Jabbar, M. K., Yan, J., Xu, H., Ur Rehman, Z., & Jabbar, A. (2022). Transfer learning-based model for diabetic retinopathy diagnosis using retinal images. *Brain Sciences*, *12*(5), 535. https://doi.org/10.3390/brainsci12050535

[11] Gundluru, N., Rajput, D. S., Lakshmanna, K., Kaluri, R., Shorfuzzaman, M., Uddin, M., & Rahman Khan, M. A. (2022). Enhancement of detection of diabetic retinopathy using Harris hawk's optimization with deep learning model. *Computational Intelligence and Neuroscience*, *2022*(1),8512469 https://doi.org/10.1155/2022/8512469

[12] Rajamani, S., & Sasikala, S. (2023). Artificial intelligence approach for diabetic retinopathy severity detection. *Informatica*, *46*(8).

[13] Zhang, Q. M., Luo, J., & Cengiz, K. (2021). An optimized deep learning-based technique for grading and extraction of diabetic retinopathy severities. *Informatica*, *45*(5). https://doi.org/10.31449/inf.v45i5.3561

[14] Silva, P. S., Cavallerano, J. D., Sun, J. K., & Aiello, L. M. (2021). Effectiveness of artificial intelligence–based diabetic retinopathy screening in a primary care setting: A pilot study. JAMA Ophthalmology, 139(10), 1076- 1082. doi: 10.1001/jamaophthalmol.2021.2924

[15] Mohanty C. et al., "Using Deep Learning Architectures for Detection and Classification of Diabetic Retinopathy," Sensors, vol. 23, no. 12, 2023. https://doi.org/10.3390/s23125726

[16] Alyoubi, W. L., Abulkhair, M. F., & Shalash, W. M. (2021). Diabetic retinopathy fundus image classification and lesions localization system using deep learning. *Sensors*, *21*(11), 3704. https://doi.org/10.3390/s21113704

[17] Yaqoob, M. K., Ali, S. F., Bilal, M., Hanif, M. S., & Al-Saggaf, U. M. (2021). ResNet based deep features and random forest classifier for diabetic retinopathy detection. *Sensors*, *21*(11), 3883. https://doi.org/10.3390/s21113883

[18] 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.

[19] Mumtaz, R., Hussain, M., Sarwar, S., Khan, K., Mumtaz, S., & Mumtaz, M. (2018). Automatic detection of retinal hemorrhages by exploiting image processing techniques for screening retinal diseases in diabetic patients. *International Journal of Diabetes in Developing Countries*, *38*, 80-87. https://doi.org/10.1007/s13410-017-0561-6

[20] Xu, Y., Zhou, Z., Li, X., Zhang, N., Zhang, M., & Wei, P. (2021). FFU-net: feature fusion u-net for lesion segmentation of diabetic retinopathy. *BioMed Research International*, *2021*(1), 6644071. https://doi.org/10.1155/2021/6644071

[21] Ab Kader, N. I., Yusof, U., & Naim, S. (2019). Diabetic retinopathy classification using support

vector machine with hyperparameter optimization. *Int. J. Advance Soft Compu. Appl*, *11*(3)

[22] Duan, K.B., Keerthi, S.S., 2005. Which Is the Best Multiclass SVM Method? An Empirical Study. Multiple Classifier Systems. MCS 2005. Lecture Notes in Computer Science, 3541, 278-285. doi.org/10.1007/11494683_28

[23] Mishra A, Singh L, Pandey M, Lakra S. (2022) Image based early detection of diabetic retinopathy: A systematic review on Artificial Intelligence (AI) based recent trends and approaches. Journal of Intelligent & Fuzzy Systems. 2022;43(5):6709-6741. doi:10.3233/JIFS-220772

[24] Li, J., Zhang, Q., Zheng, H., & Wang, Y. (2022). A deep ensemble model for diabetic retinopathy detection and classification. Journal of Medical Imaging, 9(3), 123456. doi:10. xxxx/123456.

[25] Mandal, A.K. (2019). Study the Behavior of Y, Cb and Cr of Ycbcr on Blurred Image Segmentation using Local Thresholding. International Journal for Research in Applied Science and Engineering Technology.

[26] Salur, M. U., & Aydın, İ. (2022). A soft voting ensemble learning-based approach for multimodal sentiment analysis. *Neural Computing and Applications*, *34*(21), 18391-18406.NI. https://doi.org/10.1007/s00521-022-07451-7

[27] Bazgir, Omid, et al. "Representation of features as images with neighborhood dependencies for compatibility with convolutional neural networks." Nature communications 11.1 (2020): 4391.

[28] Angulo, C., Català, A.: K-svcr. a multi-class support vector machine. In: Machine Learning: ECML 2000: 11th European conference on machine learning Barcelona, Catalonia, Spain, May 31–June 2, 2000 Proceedings 11, pp. 31–38 (2000). Springer. https://doi.org/10.1007/3-540-45164-1_4

[29] Hsu, C.-W., Lin, C.-J., 2002. A simple decomposition method for support vector machines. Machine Learning, 46, 291–314, 2002. doi.org/10.1023/A:1012427100071.

[30] Shyngyskhan, Shaimerdin, and Zhumagulovna Zhadra. "RECOGNIZING DIFFERENT IMAGES IN PYTHON USING TENSORFLOW AND KERAS." Universum: технические науки 10.11 (128) (2024): 61-66.

[31] Ahmed, I. T., Hammad, B. T., & Jamil, N. (2021). A comparative analysis of image copy-move forgery detection algorithms based on hand and machine-crafted features. *Indones. J. Electr. Eng. Comput. Sci*, *22*(2), 1177-1190. DOI: 10.11591/ijeecs.v22.i2.pp1177-1190

[32] Leopold, H. A., Orchard, J., Zelek, J., & Lakshminarayanan, V. (2017, February). Segmentation and feature extraction of retinal vascular morphology. In *Medical Imaging 2017: Image Processing* (Vol. 10133, pp. 251-257). SPIE. https://doi.org/10.1117/12.2253744

[33] Lin, G. M., Chen, M. J., Yeh, C. H., Lin, Y. Y., Kuo, H. Y., Lin, M. H., ... & Cheung, C. Y. (2018). Transforming retinal photographs to entropy images in deep learning to improve automated detection for diabetic retinopathy. *Journal of ophthalmology*, *2018*(1), 2159702. https://doi.org/10.1155/2018/2159702

[34] Jiang, Z., Dong, Z., Wang, L., & Jiang, W. (2021). Method for diagnosis of acute lymphoblastic leukemia based on ViT-CNN ensemble model. *Computational Intelligence and Neuroscience*, *2021*(1), 7529893. https://doi.org/10.1155/2021/7529893

[35] Ai, Z., Huang, X., Fan, Y., Feng, J., Zeng, F., & Lu, Y. (2021). DR-IIXRN: detection algorithm of diabetic retinopathy based on deep ensemble learning and attention mechanism. Frontiers in Neuroinformatics, 15, 778552. https://doi.org/10.3389/fninf.2021.778552

[36] Khan, Z., Khan, F. G., Khan, A., Rehman, Z. U., Shah, S., Qummar, S., ... & Pack, S. (2021). Diabetic retinopathy detection using VGG-NIN a deep learning architecture. IEEE Access, 9, 61408-61416. doi: 10.1109/ACCESS.2021.3074422

[37] Dai, L., Wu, L., Li, H., Cai, C., Wu, Q., Kong, H., ... & Jia, W. (2021). A deep learning system for detecting diabetic retinopathy across the disease spectrum. Nature communications, 12(1), 3242. https://doi.org/10.1038/s41467-021-23458-5