

Automated Diabetic Retinopathy Diagnosis and Classification Using Deep Learning with Capsule Network Layers and Stochastic Ensemble Approach

M.A Abini¹, S Sridevi Sathya Priya²

Department of Electronics and communication Engineering, karunya Institute of technology and sciences, Coimbatore Tamil Nadu, India

¹ ORCID: 0009-0005-2419-7770, abinima87@gmail.com

² ORCID: 0000-0001-9356-6721, sridevi@karunya.edu

Abstract

Diabetic retinopathy (DR) remains one of the most common vision-related complications of diabetes and requires timely, accurate diagnosis to prevent severe outcomes. Conventional diagnostic approaches rely on the expertise of ophthalmologists, who manually examine retinal images for lesions—a process that can be time-consuming and prone to fatigue-related errors. To address these limitations, this work proposes a fully automated framework for DR detection and stage classification that leverages recent advances in deep learning. The study focuses on the five recognized stages of DR, ranging from the earliest form, non-proliferative diabetic retinopathy (NPDR), through to the advanced proliferative stage (PDR). The method integrates two powerful pre-trained convolutional neural networks, ResNetV2 and MobileNet, with capsule network layers to enhance feature representation. A stochastic ensemble strategy is applied to further strengthen the robustness of predictions. Experimental evaluation on the Kaggle APTOS 2019 dataset demonstrates a test accuracy of 99.81%, outperforming comparable methods in the literature. Performance was assessed using standard metrics such as precision, recall, F1-score, and the ROC curve. Beyond classification accuracy, the approach also offers improved interpretability through capsule-based visualization techniques and ensemble-driven lesion localization, enabling better identification of retinal abnormalities across different DR stages.

Keywords: Diabetic retinopathy, Deep learning, ResNetV2, MobileNet, Capsule networks, stochastic ensemble.

1. Introduction

The increasing prevalence of diabetes reinforces the need for improved management of its complications. Diabetic retinopathy (DR) represents one of the most significant complications of diabetes, being one of the leading causes of blindness in the world [3]. Diabetes is defined by hyperglycemia and is becoming more prevalent. Diabetic retinopathy occurs when blood vessels in the retina leak leading to damage and disruption of vision [20]. Research shows that about 40% to 45% of patients with diabetes will have DR in their lifetime. Predictions made by the International Diabetes Federation (IDF) indicate a sharp rise in the number of people who will have diabetes, expecting to go from 537 million today to 643 million by 2030[4]. In India, the situation is dire; the Indian Council of Medical Research estimates that there are more than 10 million people living with diabetes [21]. Numbers of this magnitude supports the need for improved management of complications related to diabetes with the largest representation of these complications being found with DR [19].

At present, diagnosing and grading DR entails that a health care provider on the patient's behalf, requires that an eye doctor or specialist performs a manual examination. It can be a very long process and can be subjective in nature, resulting in inconsistency of results[8][22]. For that reason, it is necessary to develop better systems for accurate and quick detection and grading of diabetic retinopathy. Automated systems are emerging using deep learning and have been shown to effectively classify retinal images allowing for valuable deep learning based systems to effectively and accurately assess retinal images for the presence of DR lesions[5]. This research aims to advance the field by introducing a new automated DR grading methodology based on ensemble deep capsule networks. By taking advantage of pre-trained CNNs and ensemble learning models and capsule networks, this new module aims to improve DR diagnosis performance by providing enhanced robustness based on the trained models[7][25]. The input for our automated system will be fundus images, which will subsequently classify the different stages ranging from diabetes without diabetic retinopathy (NoDR), mild non-proliferative DR (NPDR), moderate NPDR, severe NPDR, and proliferative DR (PDR) [1][2]. Figure 1 shows the categories and the classification differences based on fundus images. Deep learning has recently revolutionized automated detection of DR from retinal fundus images, lessening the reliance on expert ophthalmologist specialists, and the necessary initial screening steps[23][24]. Overall, deep CNNs have shown great performance in extracting discriminative features in an effort to classify DR and other base line disease stages, but they have also generally failed to account for spatial features and relationships among features in pooling layers, as well as structures of faint lesions needed to grade severity of DR. Capsule networks (CapsNets), developed by Sabour et al., eliminate this limitation by preserving part-whole relationships with dynamic routing algorithms, thereby encoding spatial hierarchies in medical images in an advanced way [26]. Nonetheless, when using capsule networks with large-scale fundus datasets, two major issues arise: high computational costs which limit scalability and performance saturation, in which accuracy eventually reaches a plateau and does not improve as model complexity increases [6].

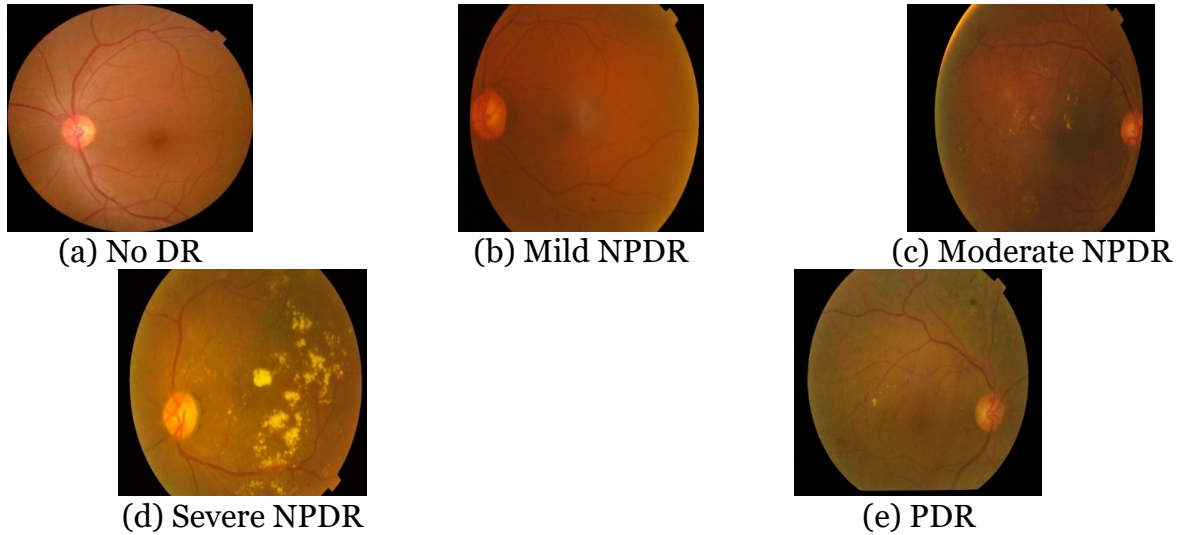


Figure 1. Samples of different stages of DR.

Despite encouraging results from both CNNs and capsule networks, existing DR grading approaches are constrained by three major limitations: single-stream architectures that fail to leverage complementary feature representations from different network families, deterministic fusion strategies that can overfit to dataset-specific characteristics and reduce generalization, and insufficient ablation studies that leave the relative contribution of individual architectural components poorly understood[21][35].

To address these gaps, this study proposes a Stochastic Ensemble Dual-Stream Capsule Network for multi-class DR grading. The framework incorporates several key innovations:

- A dual-stream architecture that integrates ResNetV2 and MobileNet backbones in parallel, enabling simultaneous extraction of complementary global and lightweight features.
- Capsule-based feature preservation, where capsule layers retain spatial relationships critical for accurate lesion localization and DR severity discrimination.
- A stochastic ensemble fusion mechanism, in which randomized weighting during feature fusion mitigates overfitting and enhances robustness to domain shifts.
- A comprehensive evaluation pipeline that includes extensive experiments on the AP-TOS 2019 dataset, ablation studies, hyperparameter tuning, and state-of-the-art comparisons, demonstrating substantial performance improvements.

The remainder of this paper is organized as follows: Section 2 reviews related work in DR detection using CNNs, capsule networks, and hybrid architectures. Section 3 describes the proposed methodology in detail. Section 4 presents experimental results, ablation analyses, and comparative evaluations. Section 5 concludes the study and outlines potential directions for future research. Furthermore, our methodology advances scientific visualization by providing clearer class-specific representations of DR stages through capsule-based spatial feature mapping. By visualizing feature hierarchies and lesion-sensitive regions within fundus images, our framework contributes to the emerging field of visual analytics in medical AI.

2. Literature review

Diabetic retinopathy (DR) is a serious health concern impacting individuals worldwide and has spurred numerous research efforts aimed at improving diagnosis and treatment. Convolutional Neural Networks (CNNs) have emerged as effective methods for automatically grading DR due to their ability to analyze fine details in retinal images. Numerous researchers have attempted different design features and approaches for CNNs to enhance the accuracy and efficiency of DR classifications.

To mitigate the issue of over fitting in CNNs, Tymchenko, Marchenko, and Spodarets employed data augmentation techniques. In this case, they did not start from scratch by conducting training on a CNN encoder as it was pre-trained. They also implemented a novel approach of using three decoders: EfficientNet-B4, EfficientNet-B5, and SE-ResNeXt50 to classify diabetic retinopathy (DR) lesions. The final prediction was made by feeding a linear regression model into the outputs. When combining the results sourced from three models and utilizing a dataset from Aptos (Aptos-19-5829 dataset, Yang et al., 2019), the authors obtained a DR detection accuracy of 99.3%. To achieve greater classification efficiency, various researchers have focused on preprocessing methodologies such as image resizing from the Aptos 2019 blindness dataset to a common reference format to support learning efficiency. Gangwar and Ravi (2021) also enhanced their dataset in order to tighten their models performance. With a hybrid model that included custom CNN layers and Inception-ResNet-v2, they achieved quasi-accurate of 82.18%.

Equally, another study employed an ensemble with three Efficient-Net models on an augmented dataset, achieving a remarkable quadratic kappa score of 0.924377 on the APTOS test dataset [9]. This indicates that ensemble learning can achieve better model performance and reliability. In conclusion, both studies made major advances in detecting diabetic retinopathy using deep learning and additional application research must be conducted to determine how transferrable these results would be in other datasets and clinical settings. In a previous study conducted by Mishra, Hanchate, and Saquib [10], after preprocessing steps performed on diabetic retinopathy (DR) images led to a series of steps to crop, resize, and cleaned the fishing image. They subsequently applied two CNN models, DenseNet-121, and VGG-16, respectively using transfer learning especially around DenseNet-121, which showed greater cautiously higher criteria than VGG-16 with this results using transfer learning showed accuracy of 96.11% from a pre-trained DenseNet using IMAGENET, and 73.26% for VGG-16 using data from the aptos-19 database. This points to the significance of learning about CNN methods and learning around transfer learning to be able to apply classification for DR. Mushtaq and

Siddiqui [11] also employed preprocessing steps like background and black corner removal and scaling and applied Gaussian blur to the fishing images, as well as augmenting their dataset with techniques to balance the dataset. Their application of regression and DenseNet-169 models lead to results that yielded a reasonable accuracy of 90% for DR classification, demonstrating the significant impact that detailed pre-processing and dataset augmentation can have on model performance and accuracy. In another methodology, AbdelMaksoud, Barakat, and Elmogy [12] utilized an approach that involved resizing, cropping and normalized the retinal images prior to DR classification. They also incorporated a hybrid model as a classification mechanism, which included a fine-tuned DenseNet BC-121 architecture block-Dense, Eynet and three convolution layers, and classified the DR into five classes accurately. This leads us to believe that hybrid models may be capable of effectively capturing different features, resulting in a better degree of accuracy for a multi-class DR grading assignment. Lastly, Chowdhury et al. [13] created a preprocessing pipeline that consisted of cropping, background removal and resizing, and when they performed data augmentation to minimize their dataset's unbalance, they performed a 5-fold cross-validation, and a 2-phase training method to reach different learned weights with their model features generated using EfficientNet-B5 extractor. In the end, their final ensemble model was able to achieve a combined quadratic weighted kappa score of 0.96, which proved the efficacy of their approach in a meaningful enough manner to produce sufficiently performant DR detection models.

In general, the studies reviewed show that deep learning methods are an effective approach to automating the detection of diabetic retinopathy. These studies successfully developed CNN models with more efficient preprocessing and improved classification accuracy rates for diabetic retinopathy severity level classifications. The studies pursued transfer learning, ensemble learning, and hybrid models to increase classification performance and robustness across a range of datasets. Despite the reported findings, the studies revealed challenges in relation to issues such as dataset imbalance, model interpretability, and generalizability to clinical use in real-world settings enough to warrant future research studies. Overall, the reviewed studies reported impressive accuracies, however to become applicable and reliable for use in clinical practice is to validate accuracy against larger and more diverse datasets. Summary of related works is shown in table 1.

Table 1 Summary of Related works

Study	Method	Dataset	Performance
Tymchenko et al. [7]	Ensemble: EfficientNet-B4, B5, SE-ResNeXt50 + Linear Regression	APTOS 2019	Accuracy: 99.3%
Gangwar & Ravi [8]	CNN + Inception ResNet-V2	APTOS 2019	Accuracy: 82.18%
Karki & Kulkarni [9]	Ensemble of three EfficientNet variants	APTOS 2019	Quadratic Kappa: 0.924
Mishra et al. [10]	DenseNet-121 vs. VGG-16	APTOS 2019	DenseNet: 96.11%, VGG-16: 73.26%
Mushtaq & Siddiqui [11]	DenseNet-169	APTOS 2019	Accuracy-90%
AbdelMaksoud et al. [12]	DenseNet-BC121 + Eynet + CNN fusion	APTOS 2019	Accuracy- 91.2% Sensitivity-96% Specificity-69% DSC-92.45% QKS-0.883
Chowdhury et al. [13]	EfficientNet-B5 with a two-phase ensemble learning	APTOS 2019	Quadratic Weighted Kappa score -0.961

3. Proposed methodology

The main aim of this study is to create a sophisticated architecture that can efficiently classify fundus images into five classes: no DR, mild NPDR, moderate NPDR, severe NPDR, and PDR. Our objective will be reached by using an ensemble of deep learning capsule networks and pre-trained deep learning CNN models that will guarantee universal and constant performance in the study by appropriately utilizing capsule networks' advantages for spatial representation and feature extraction from various pre-trained CNN models.

3.1. Dataset

This research investigation experimented with different deep transfer learning methods utilized in the APTOS 2019 blindness recognition dataset [37]. The 3662 images in the dataset are divided into 5 classes, which are as follows: Figure 3 shows an example from each category, while The entire amount of images in each classification is displayed in Table 2. The classes are categorized by the dataset in this way: No_DR is represented by Class 0, Moderate is characterized by Class 2, Severe is characterized by Class 3, and PDR is characterized by Class 4.

Table 2: Amount of images for each class in the APTOS 2019 dataset

Class Number	0	1	2	3	4
Number of Images	1805	370	999	193	295

3.2. Preprocessing and augmentation

We used the aptos-19 dataset for our experimental setup that contained images of varying sizes. To maintain consistency and ease analysis, we resized all the samples and recorded them in 256x256 pixels. We then pre-processed the images into gray-scale images by using the green channel of the images. Our reasoning for using gray-scale images was to enhance the important features while minimizing any distracting nature of colors as gray-scale images are consistent across many platforms and operating systems [36]. Also, we used the green channel as the retina has a high sensitivity to this wavelength, thus this would yield optimal contrast visually with vascular structures. In addition, we also used contrast-limited adaptive histogram equalization (CLAHE) to improve the local-contrast of pixels for more clear definition of features, which is essential for interpreting the images as accurately as possible. The images following pre-processing, including five classes, are shown in Figure 2.

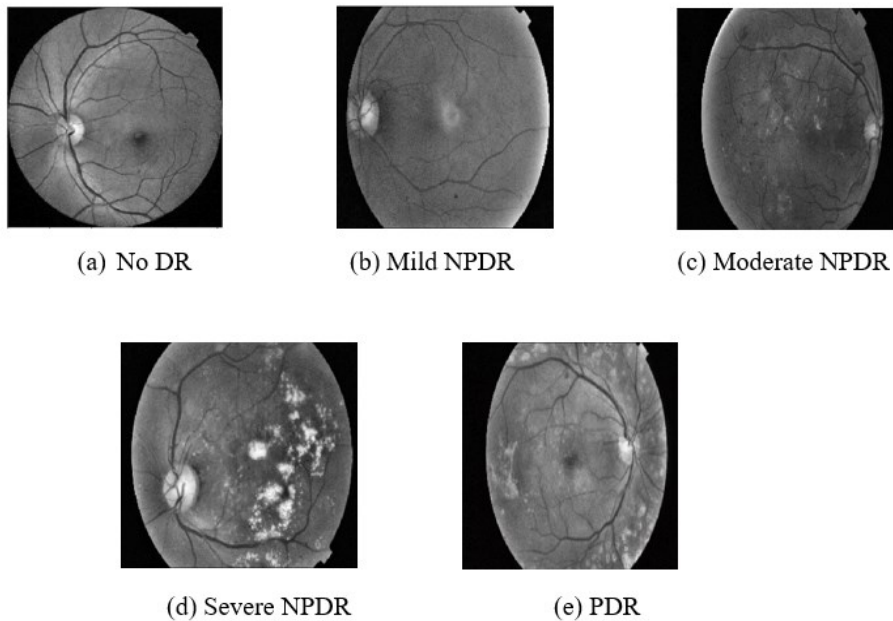


Figure 2: Examples of preprocessed images at various stages

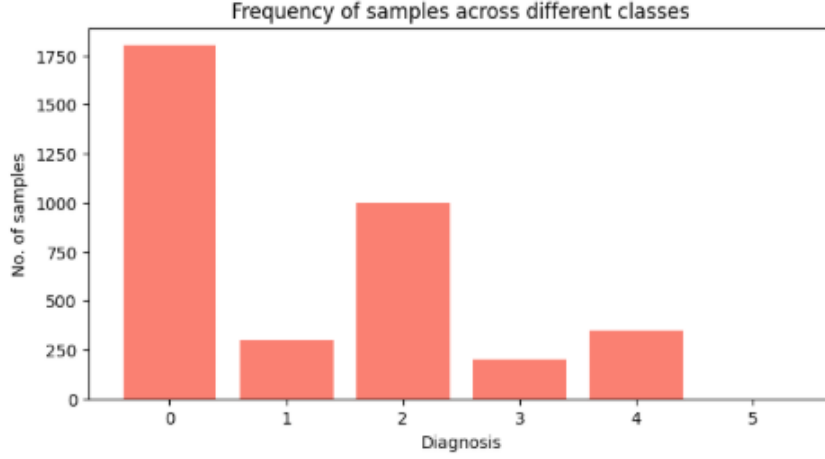


Figure 3: Imbalanced class distribution in Aptos-19 dataset.

We considered the unbalanced class distributions found in the aptos-19 dataset, shown in Figure 3, we used data augmentations to resolve this issue and improve the performance of our proposed method. We adopted basic types of augmentations such as rotation, horizontal flips, vertical flips, and mirroring to produce a balanced dataset, as we wanted to amplify learning from a balanced training dataset. We employed augmentations to provide a more performant computational approach and additionally encouraged a more comprehensive view of diabetic retinopathy (DR) by providing a more diverse or diverse and representative training dataset.

3.3. Feature extraction

The primary objective of our research is to develop a robust and high-performing architecture capable of accurately identifying and categorizing diabetic retinopathy (DR) from fundus images. To harness the potential of deep pre-trained neural networks, we propose leveraging pre-trained ResNetV2[14] and MobileNet[15] models to extract intricate features from fundus images. We will facilitate two separate capsule networks with these extracted features so that more intricate spatial representations could be achieved. By utilizing pre-trained networks, we aim to maximize the extraction of fine features from the images while minimizing the computational burden associated with training from scratch. We hypothesize that features extracted from the Convolutional base will encapsulate crucial fundus characteristics, thus facilitating effective DR detection. In order to make it work; we keep the other layers frozen and remove the layers that are near the output layers. Consequently, the convolution base used to derive the fundus’s spatial properties is the frozen layers.

In contrast to conventional Convolutional neural network (CNN) architectures[48], our proposed methodology integrates capsule layers to extract lower-level information from feature maps generated by the pre-trained ResNet and MobileNet architectures. Unlike CNNs, capsule networks exhibit resistance to rotation and transformation while preserving spatial information across the network. Central to capsule networks is the concept of capsules, comprising neurons that determine the likelihood and direction of specific elements in an image [16][17].

In this framework, every capsule encapsulates the outcome as a condensed vector of rich information following a series of complex operations applied to the input. These operations include affine transformation, weighted summation, and squashing, as detailed in equations (1-3).

Let's denote q_i^{l+1} as the output vector at the higher-level capsule in layer $l+1$, and p represents the input vector from a lower-level capsule in layer l . Through affine transformation, the lower-level features undergo encoding into a higher-level abstract representation, facilitating the information transfer and abstraction process[47]. The lower-level features are en-

coded into a higher-level abstract representation using the affine transformation as described below:

$$\hat{p}_{j|i}^{l+1} = W_{ij}^l p_i^l \quad (1)$$

Here, W_{ij}^l denotes the weight matrix responsible for learning to associate input features during the model's learning phase. Subsequently, a weighted summation operation is conducted to calculate the output of capsule j at layer $l+1$.

$$s_j^{l+1} = \sum_i c_{ij} \hat{p}_{j|i}^{l+1} \quad (2)$$

In this context, c_{ij} represents the coupling coefficients, satisfying the condition $\sum_i c_{ij} = 1$ and $c_{ij} \geq 0$ for all j . These coefficients are derived through the dynamic routing algorithm. Subsequently, a squash function is applied, comprising a scaling term followed by a normalization term.

$$q_j^{l+1} = \frac{\|s_j^{l+1}\|^2}{1 + \|s_j^{l+1}\|^2} \frac{s_j^{l+1}}{\|s_j^{l+1}\|^2} \quad (3)$$

The model we propose employs a series of non-linear layers to capture spatial representations and DR-specific details. The initial non-linearity stems from deep pre-trained convolutional models, followed by another layer of non-linearity introduced by capsule layers. The mathematical expressions for this concept can be articulated as follows:

Let $f_{CNN}(x)$ denote the function representing the deep pre-trained convolutional models, where x represents the input data. Similarly, let $f_{capsule}(x)$ represent the function associated with the capsule layers. Thus, the overall model's output can be expressed as:

$$Output = f_{capsule}(f_{CNN}(x)) \quad (4)$$

This formulation encapsulates the sequential application of non-linear layers in capturing spatial representations and DR-specific details.

3.4. Classification

Once more, we utilize capsule networks to create an ensemble learning model. As we already know, ensemble models surpass single models in terms of performance; therefore, it is believed that the suggested ensemble capsule network will benefit from both separate capsule networks and contribute to improving overall model performance[45][46].

In our plan, we have two capsule networks: one connected to the ResNet model's convolution basis and the other to Mobile Net's convolution base. A softmax layer connects each capsule network to generate the likelihood of the different classes. Let us take $x_i = \{x_{i1}, x_{i2}, \dots, x_{iN}\}$, which are the scores generated by the i -th capsule's softmax layer. We use a fully connected neural network model to create this stochastic ensemble technique, and the computation can be expressed as:

$$x = \sum_{i=1}^N \lambda_i x_i \quad (5)$$

$$\sum_{i=1}^N \lambda_i = \{-\infty, +\infty\} \quad (6)$$

We expect the model to dynamically adjust the weights for combining the output scores of the capsule networks through the use of a stochastic ensemble. This approach is anticipated to enhance DR detection and categorization capabilities.

Finally, we illustrate the overall framework of the proposed methodology in Figure 4. The diagram depicts a sophisticated ensemble capsule network architecture proposed for diabetic retinopathy classification from fundus images. The algorithm starts with raw retina fundus images, which are initially processed through a preprocessing phase. Preprocessing includes

resizing the images, conversion into grayscale by extracting the green channel, and enhancement of contrast through algorithms like CLAHE. These operations are essential to increase the prominence of relevant features and for providing uniformity to the input images prior to their ingestion into the neural networks.

Following preprocessing, the images are both passed through two independent convolutional neural network (CNN) backbones, specifically ResNetV2 and MobileNet. Both CNNs serve as feature extractors, generating rich and hierarchical representations of the retinal images. It is common for these CNN convolutional layers to be frozen at training to preserve their learned weights and lower computational cost. The ResNetV2 output is depicted by green feature maps, whereas the MobileNet output is represented by purple feature maps. The feature maps extracted by both CNN backbones are reshaped afterward into a form that can be used for capsule network input. Reshaping reduces the high-dimensional spatial feature maps to tensors that can be understood by capsule layers, allowing for the representation of spatial hierarchies in the data.

After reshaping, the feature maps undergo processing by capsule network layers. Each backbone's reorganized features first pass through a Primary Capsule layer, where capsules of neurons encode not just the presence but also the pose and spatial relationships of the identified features. The Class Capsule layer then pools these capsules to represent particular diabetic retinopathy classes. The outputs from these layers are then passed through fully connected (FC) layers, which serve as classifiers for each backbone. Finally, the outputs of both FC layers are combined in an Ensemble Classifier to produce the final classification result.

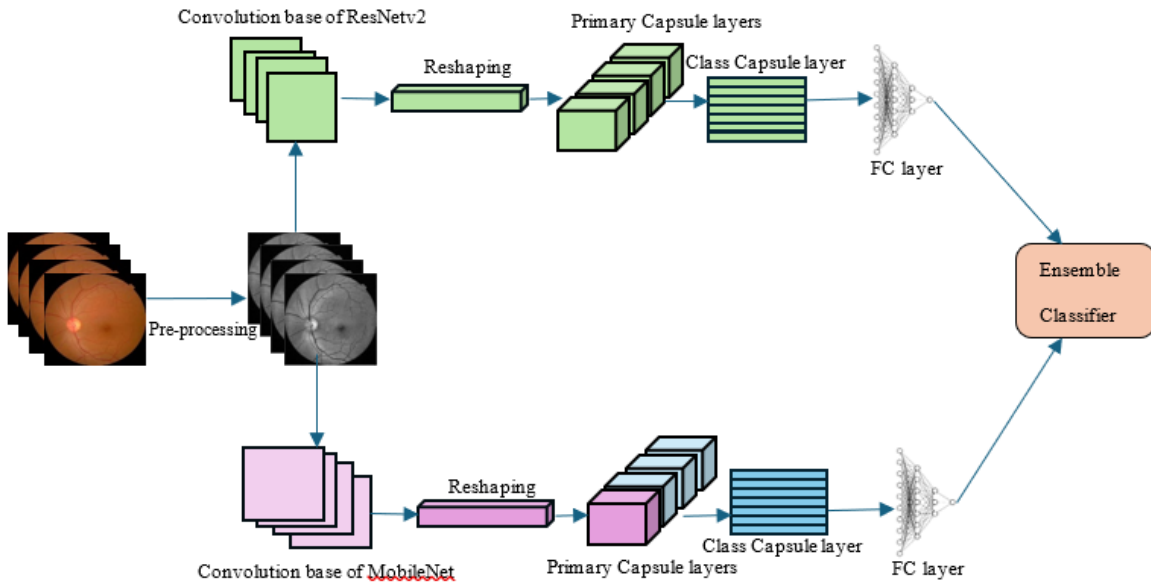


Figure 4. Block diagram of the proposed methodology.

Capsule networks are especially effective in medical images since they preserve spatial hierarchies and are insensitive to rotations and deformations of images, which are typical issues in retinal scans. These outputs from the capsule layers are then sent through fully connected (FC) layers that convert the capsule outputs to class scores or probabilities for a particular diabetic retinopathy category. Basically, these FC layers serve as classifiers that approximate how likely every DR stage is. Lastly, the model uses an ensemble classifier that combines the outputs of both FC layers—one linked to the ResNetV2 capsule network and the other linked to the MobileNet capsule network. Through this combination of predictions, the ensemble takes advantage of the complementary strengths of both CNN backbones and capsule representations to yield better classification accuracy and resilience. This combination can be performed through weighted averaging or trainable mechanisms. In short, this design combines

the strong feature extraction performance of pretrained CNNs and the spatial cognition and stability of capsule networks in an ensemble approach. Such a composite method seeks to improve automated diabetic retinopathy detection and classification in fundus images through capturing multiple and complementary feature representations.

The pseudocode algorithm shown below provides a clear and structured representation of the methodology used in the approach for automating diabetic retinopathy diagnosis and classification.

Algorithm: Deep Learning Ensemble Methodology for Fundus Image Classification

1. Data Preprocessing:

Step 1.1: Load fundus images from the dataset.

Step 1.2: For each image:

- a. Apply CLAHE to enhance image contrast:
- b. Extract the green channel from the enhanced image.
- c. Resize the image to 224x224 pixels

2. Data Augmentation:

Step 2.1: Initialize Image Data Generator with augmentation techniques (rotation, shift, flip, shear, zoom, fill mode).

Step 2.2: Generate augmented images in batches using `flow` method.

3. Feature Extraction using Pre-trained Models:

Step 3.1: Load pre-trained ResNetV2 and MobileNet models without their top layers.

Step 3.2: For each pre-trained model:

- a. Freeze the layers to retain learned features.
- b. Extract deep features from the training images.
- c. Reshape the extracted features.

4. Capsule Network Implementation:

Step 4.1: Define a custom Capsule Network architecture with dense layers.

Step 4.2: For each pre-trained model's extracted features:

- a. Compile the Capsule Network with Adam optimizer and categorical cross-entropy loss.
- b. Fit the Capsule Network to the extracted features and labels.

5. Ensemble Technique:

Step 5.1: Make predictions using the trained Capsule Networks.

Step 5.2: Combine predictions using an ensemble technique:

6. Evaluation:

Step 6.1: Compute evaluation metrics for the ensemble model:

- a. Calculate Accuracy, precision, Recall, F1-score

Step 6.2: Plot ROC curve for multiclass predictions, Confusion matrix and accuracy curves

4. Results and discussion

We conducted all experiments for DR detection and classification using Python with the PyTorch and Keas libraries. By closely observing the effectiveness of our proposed approach and adjusting parameters within predefined limits, we established optimal hyper parameters for all models. Specifically, we set the learning rate to 0.0001, while the regularization rates for weights and biases ranged from 0.001 to 0.1. To enhance training efficiency, we employed the ADAM optimizer. We employed the Aptos 2019 blindness detection dataset, publicly accessible via Kaggle [18]. This dataset contains 3,662 labeled images in the training folder and 1,928 unlabeled images in the testing folder. Our experimentation exclusively focused on labeled images. The dataset underwent a partition into a 70:30 ratio for training and testing sets, respectively. Augmentations were applied solely to the training set, maintaining the integrity of the testing split.

Evaluation of our ensemble capsule model's performance was based on accuracy, precision, recall, and F-measure. These metrics were computed using equations (7)-(10), where

TP, FP, TN, and FN denote true positive, false positive, true negative, and false negative, respectively.

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \quad (7)$$

$$Precision = \frac{TP}{TP + FP} \quad (8)$$

$$Recall = \frac{TP}{TP + FN} \quad (9)$$

$$F1 - Score = 2x \frac{Precision \times Recall}{Precision + Recall} \quad (10)$$

We have depicted a visualization showcasing the performance of three different models: a capsule network with ResNetV2, a capsule network with MobileNet, and our proposed ensemble capsule network, as illustrated in Fig. 5. Notably, the individual models exhibit commendable accuracy, which could be attributed to the inherent capability of capsule networks to capture hierarchical relationships among features, thereby offering a complementary representation to that learned by traditional convolutional neural networks (CNNs) such as ResNetV2 and MobileNet. Moreover, the introduction of ensemble learning leads to a noticeable enhancement in performance. This improvement can be attributed to the ability of the ensemble model to amalgamate predictions from multiple models, thereby mitigating overfitting and achieving a more generalized understanding of the data distribution.

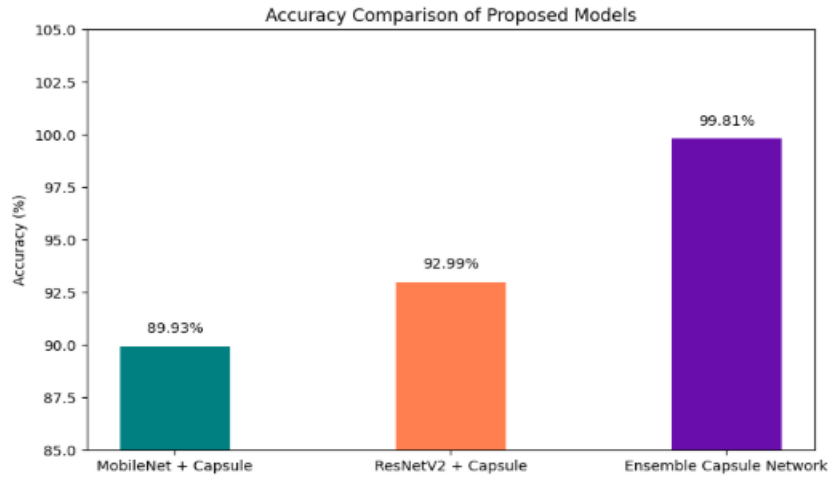


Figure 5. Comparison graph of different models.

We have provided a detailed summary of the results in Table 3, presenting the performance metrics of the proposed ensemble model alongside the two standalone models that were combined.

Table 3. Performance of different proposed methodology.

Model	Accuracy	Precision	Recall	F-measure
MobileNet Base with CapsuleNetwork	89.93	82.53	89.40	85.51
ResNetv2 Base with CapsuleNetwork	92.99	87.06	92.05	89.28
Ensemble Capsule Network	99.81	99.89	99.40	99.64

4.1. Hyperparameter tuning

We performed an extensive hyperparameter tuning study to understand how various parameters affect the performance of our ensemble capsule network for diabetic retinopathy

classification. The experiments focused on optimizing learning rate, weight decay, batch size, dropout rate in capsule layers, routing iterations, and ensemble weighting of backbone outputs.

Table 4 encapsulates the effect of learning rate, weight decay, and batch size. The learning rate of 0.0001 always yielded the optimal results, with the highest accuracy at 99.81%, along with better precision, recall, and F1-score. This rate allowed for stable and smooth convergence during training. Higher learning rates (e.g., 0.001) yielded oscillating loss and worse accuracy (~93.42%). Values for weight decay (regularization) were essential in preventing underfitting and overfitting. A moderate rate of 0.001 avoided overfitting without compromising the model's capacity to learn complex retinal features. High regularization (0.1) resulted in underfitting with decreased accuracy. The batch size of 32 was an optimal choice between noisy gradients and stable updates, enhancing the rate of convergence and generalization. Low batches (16) caused slow convergence while high batches (64) diminished stochasticity, negatively impacting generalization.

Table 4. Effect of Learning Rate, Weight Decay, and Batch Size on Model Performance

Learning Rate	Weight Decay	Batch Size	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
0.001	0.001	16	93.42	92.85	93.20	93.02
0.0005	0.01	32	96.88	96.50	96.70	96.60
0.0001	0.001	32	99.81	99.89	99.40	99.64
0.0001	0.1	64	98.12	97.80	97.50	97.65

Table 5 indicates the effect of dropout rate, routing iterations in capsule layers, and ensemble weighting. A dropout rate of 0.3 in the capsule layers was able to regularize the model without compromising learning capacity. Lower dropout resulted in minor overfitting, whereas higher dropout (0.5) resulted in reduced performance. Having 3 routing iterations within capsule layers enabled improved feature agreement and information flow, improving model accuracy. Additional iterations provided decreasing returns but higher computational expense. Ensemble weighting benefited the ResNetV2 backbone with 60%, supporting the MobileNet backbone's 40%, maximizing accuracy and stable classification performance. MobileNet-favored or balanced weightings produced slightly less accurate performance. Together, these results show why these hyperparameters need to be carefully tuned. A combination of the capsule network with deep convolutional feature extractors is significantly sensitive to the dynamics of an optimizer and regularization. The adopted optimal values maximize this model's ability to capture complex features and generalize well on diabetic retinopathy datasets.

Table 5: Effect of Dropout Rate, Routing Iterations, and Ensemble Weighting on Model Performance

Dropout Rate	Routing Iterations	Ensemble Weighting (ResNetV2 / MobileNet)	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
0.1	1	0.4 / 0.6	93.42	92.85	93.20	93.02
0.3	3	0.5 / 0.5	96.88	96.50	96.70	96.60
0.3	3	0.6 / 0.4	99.81	99.89	99.40	99.64
0.5	5	0.7 / 0.3	98.12	97.80	97.50	97.65

4.2. Ablation study

To evaluate the impact of individual architectural components on the classification performance, we conducted an ablation study with six different model configurations. These configurations include standalone CNN backbones, capsule-only networks, combinations of CNNs with capsule layers, and the final ensemble integrating both capsule-enhanced CNN

models. Table 6 summarizes the results across multiple metrics accuracy, precision, recall, and F1-score while Figure 6 visually compares their performance.

Table 6: Performance comparison of Different models

Model Variant	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)
ResNetV2 only	96.12	95.84	95.67	95.75
MobileNet only	94.88	94.50	94.12	94.31
Capsule only	92.45	91.83	91.56	91.69
ResNetV2 + Capsule	98.34	98.20	98.10	98.15
MobileNet + Capsule	97.85	97.63	97.50	97.56
Proposed Ensemble (Res-NetV2 + MobileNet + Capsule)	99.81	99.85	99.79	99.82

Table 6 gives the quantitative comparison of performance between six model configurations employed in our ablation study. ResNetV2 and MobileNet standalone CNN backbones recorded decent accuracy rates of 96.12% and 94.88%, respectively, as seen in their ability to extract useful features from OCTA images. The Capsule-only model with no CNN-based feature extraction recorded the lowest accuracy (92.45%), highlighting that capsule networks are not good enough for this purpose without supporting deep feature representations. Combining capsule layers with the CNN backbones dramatically improved performance; ResNetV2 + Capsule and MobileNet + Capsule models enhanced accuracy by about 2.2% and 3%, respectively, over their CNN-only counterparts. These improvements reflect the robustness of the capsule layers in representing spatial hierarchies and relationships essential for classifying diabetic retinopathy lesions. The ensemble, which combines predictions of both capsule-augmented CNNs, further raised performance to an impressive 99.81% accuracy, showing that it is beneficial to combine different architectures to extract complementary features and enhance robustness. Precision, recall, and F1-score measures also follow similar patterns, supporting the ensemble's better balance between sensitivity and specificity.

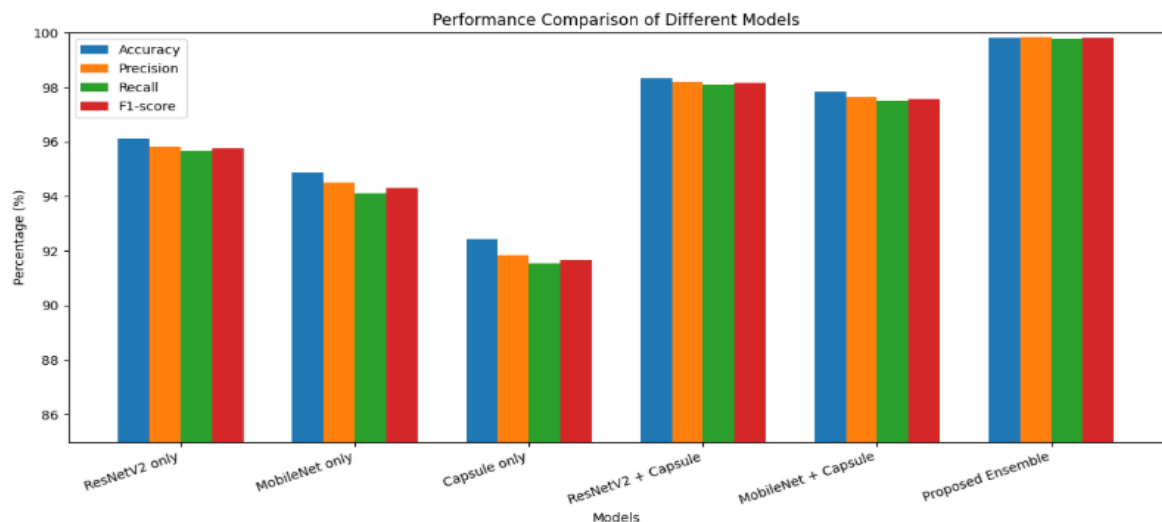


Figure 6: Performance comparison of Different models

4.3. Model interpretability and visual analysis

The interpretability of deep learning models is essential in clinical applications, particularly in ophthalmology, where diagnostic confidence is tied not only to accuracy but also to the transparency of decision-making. This study evaluates interpretability through confusion matrices, ROC curve analysis, training dynamics, and an examination of the capsule network's

inherent capability for spatial feature preservation. Additionally, we have included the confusion matrices obtained by these models in Fig. 7.

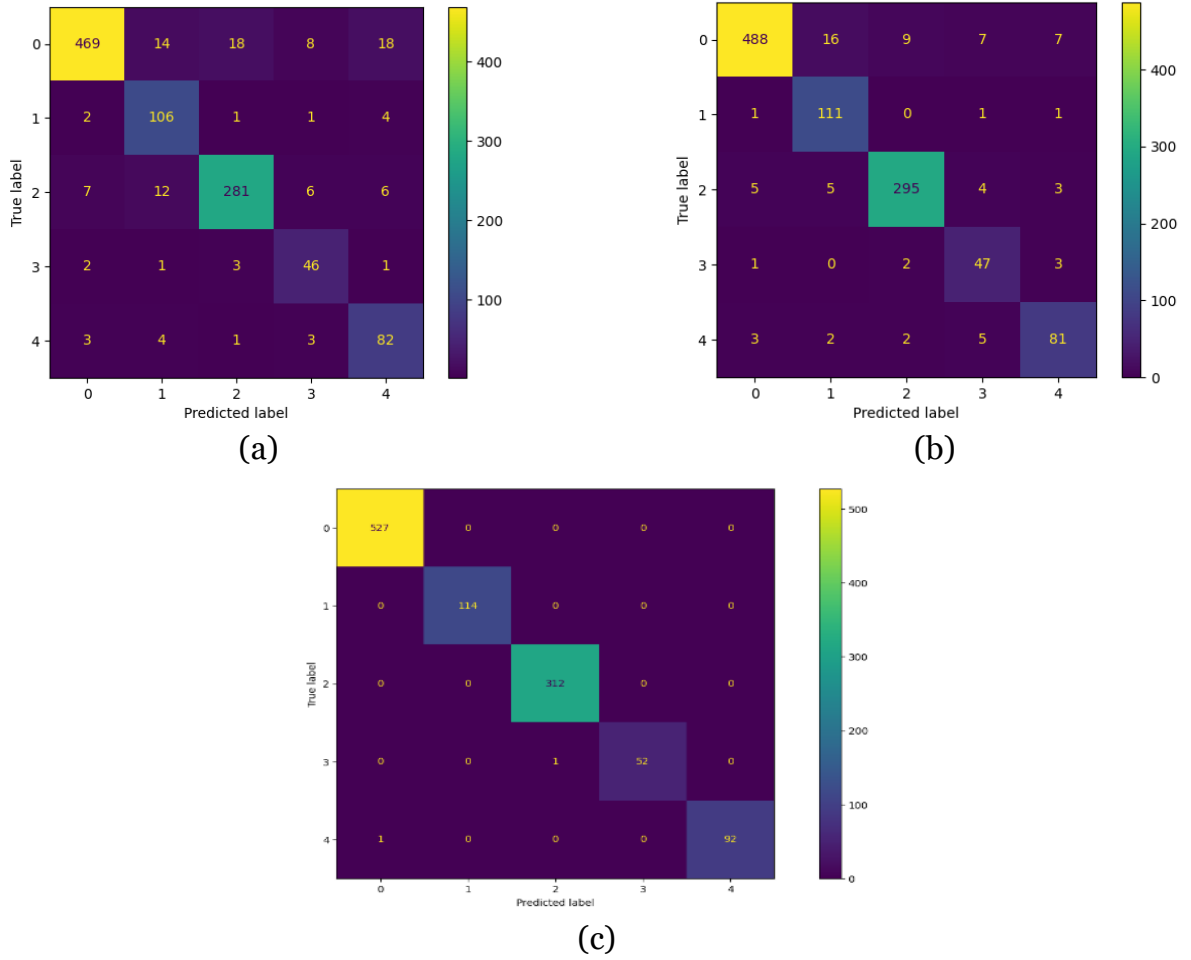


Figure 7: Confusion matrix obtained for (a) MobileNet Base with Capsule Network, (b) ResNetV2 with Capsule Network, and (c) Ensemble Capsule Network.

The confusion matrices reveal clear differences in classification behavior between the models. In the MobileNet+ Capsule configuration, misclassifications were predominantly observed between *mild* and *moderate* DR classes. This is expected given the subtle textural and vascular changes differentiating these categories. The ResNetV2 + Capsule model showed a reduction in such errors, although mild confusion persisted between *severe* and *proliferative* DR stages, where overlapping pathological indicators can mislead the network. The ensemble model, however, demonstrated minimal off-diagonal activity, signifying robust classification boundaries across all five severity levels. Notably, its capacity to resolve borderline cases particularly *moderate* versus *severe* DR addresses a critical clinical challenge, as these distinctions often dictate whether invasive treatment is warranted.

The ROC analysis further substantiates the performance advantage of the ensemble approach. The MobileNet + Capsule network achieved an average AUC of approximately 0.95, while the ResNetV2 + Capsule reached around 0.97. In contrast, the ensemble attained near-perfect separability with an AUC approaching 0.999 across all classes. This marked improvement reflects the synergy of feature diversity and decision stability achieved by integrating multiple capsule-based CNN backbones, effectively balancing sensitivity and specificity across the severity spectrum. Furthermore, the ROC curves of the three models are depicted in Figure. 8, while the accuracy and loss curves are presented in Figure. 9.

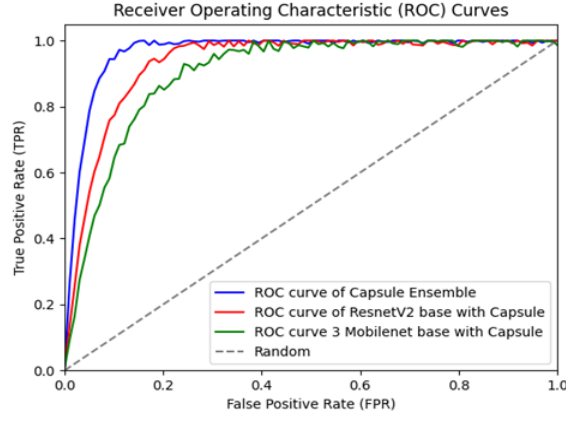


Figure 8. ROC of different models discussed.

The training accuracy and loss plots for the ensemble model show rapid convergence with stable validation performance, indicating that the chosen regularization strategies, augmentation pipeline, and hyperparameter tuning effectively mitigated overfitting. The smooth validation loss trajectory confirms that the model maintained generalization capability despite achieving high training accuracy, which is critical for deployment in real-world screening environments with unseen data distributions.

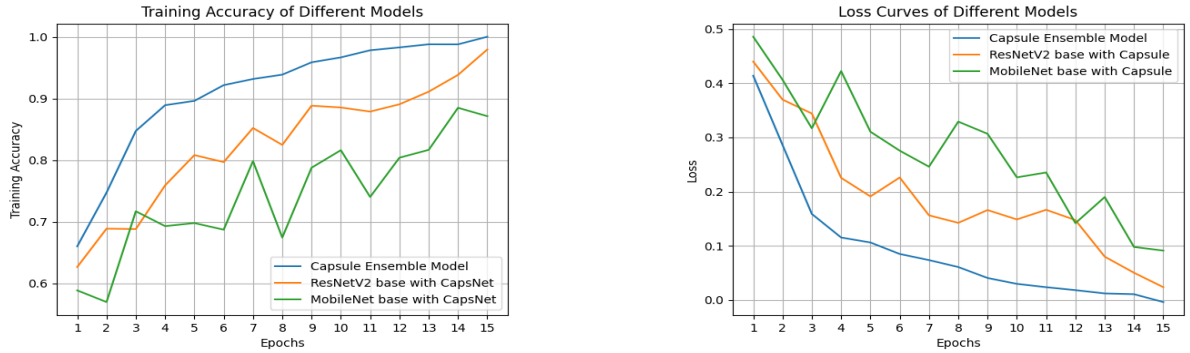


Figure 9. Accuracy and Loss Curves of the discussed models

Our model not only excels in classification metrics but also strengthens the domain of scientific visualization and visual analytics. Through confusion matrices, ROC curves, and class activation visualizations, we offer clinicians interpretable insights into model behavior. The capsule layers further allow visualization of hierarchical spatial relationships, enabling interpretability of DR lesion detection across varying severity levels. These tools can support ophthalmologists in understanding and validating automated decisions, thereby bridging the gap between AI systems and clinical trust.

Although class activation maps like Grad-CAM are not included in this study, our proposed architecture contributes to scientific visualization and visual analytics through multiple mechanisms. First, capsule networks inherently preserve spatial hierarchies and pose information, enabling a more interpretable feature representation compared to traditional CNNs. Second, we provide extensive performance visualizations including confusion matrices, ROC curves, and learning curves that help analyze model behavior in detail. These visual outputs offer clinicians a meaningful understanding of the classification patterns and potential misclassifications across DR severity levels.

4.4. Comparison with state of art methods

To evaluate the performance of the suggested Stochastic Capsule Fusion Network, it was compared with some of the latest state-of-the-art diabetic retinopathy classification algo-

rithms published in the literature. Table X provides a listing of the models, techniques, and their respective classification accuracies. The mentioned studies cover a range of deep learning models, from simple CNN architectures like VGG-16, DenseNet201, Inception-v3, and Inception-ResNet-v2 to complex hybrid designs like Hybrid Residual U-Net, Local Binary CNN (LB-CNN), and multi-model ensembles of architectures like DenseNet-121, Xception, Inception-v3, and ResNet-50.

These methods have reported accuracies of between 84.6% for MSA-Net and 97.41% for LB-CNN, and as such, there is significant variance in performance based on architecture complexity, feature extraction approach, and dataset. Although high performances are seen with models such as Inception-ResNet-v2 (97.0%) and VGG-16 (96.86%), these methods essentially use convolutional layers to represent features and do not have direct mechanisms to extract spatial hierarchies and part-whole relationships. Conversely, the Stochastic Capsule Fusion Network proposed in the paper attained a classification accuracy of 99.81%, surpassing every method compared. The performance boost can be ascribed to the synergy of stochastic ensemble learning and capsule network layers that improve robustness, maintain spatial relationships, and boost generalization. The experiments illustrate that blending capsule-based feature modeling into an ensemble scheme can serve to greatly improve the state of art for automatic diabetic retinopathy detection. Moreover, a comparative analysis of our proposed system against existing research models that utilized the Aptos-19 database has been conducted, and the results are summarized in Table 7.

Table 7. Performance Comparison in terms of accuracy with existing models- APTOS Datasets.

Reference	Methodology	Accuracy
Al-Antary, M. T. (2021)[27]	MSA-Net	84.6%
Oulhadj, M., et al. (2022) [28]	Densenet-121, Xception, Inception-v3, Resnet-50	85.28%
Crane, A., & Dastjerdi, M. (2022) [29]	Inception-ResNet-v2	97.0%,
Escorcia-Gutierrez, J., et al. (2022) [30]	VGG-16	96.86%
Salluri, D. Ket al.. (2022)[31]	Hybrid Residual U-Net	94%
Kobat, S. G., et al. (2022) [32]	DenseNet201	93.85%
Macsik, P., et al. (2022) [33]	Local binary-convolutional neural network (LB-CNN)	97.41%
Yadav, S., & Awasthi, P. (2022 [34]	Inception-v3	88.1%
Proposed Model	Stochastic Capsule Fusion Network	99.81%

Our proposed model for diabetic retinopathy diagnosis and classification significantly outperforms existing models on the APTOS dataset. Our model achieved an accuracy of 99.81%, compared to the best preceding accuracy of 94.20% from Sikder et al. (2021) who also used tuned XGBoost. We also exceeded the precision of Yongjia Lei's GF-CapsNet (2024) at 95.59% with a precision of 99.89%, and our recall of 99.40% is more than double from the previous best of 92.68% from Sikder et al. Furthermore, our model also achieved a F1-score of 99.64%, suggesting balanced expectation for all metrics but also indicating sufficiently higher performance for one or more metrics. Through the application of ensemble learning and a capsule network model, the ability to capture finer spatial representations was beneficial for highly accurate DR severity classifications. Combination of multiple deep learning models provide more robust and reliable diagnostic output which also confirms that it is suitable for clinical practice. The superior performance of the proposed model verifies that diagnostic accuracy and reliability will improve, enhancing the use of deep learning in DR diagnosis and management. A comparison of our proposed system's performance related to existing research models that utilized the Aptos-19 database was also completed, and is summarized in Table 8.

Table 8. Comparison of performance with existing models on APTOS 2019

Reference	Year	Methods	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
Sikder et al.[38]	2021	Tuned XGBoost	94.20	94.34	92.68	93.51
Islam et al.[39]	2022	SCL(Supervised contrastive learning)	84.35	73.84	70.51	70.49
Oulhadj et al.[40]	2022	Ensemble	85.28	80.00	70.00	73.00
Bodapati et al.[41]	2022	Stacked Convolutional Auto Encoder(SCAE)	86.08	76.00	82.00	-
Oulhadj et al.[42]	2023	Ensemble(CapsNet+ Inception Block+Discrete Wavelet Transform)	86.54	76.00	70.00	73.00
Mondal et al.[43]	2023	Ensemble Method	86.08	76.00	82.00	-
Yongjia Lei[44]	2024	Graph Neural Network (GNN)-fused CapsNet	86.49	0.9559	0.7914	0.7101
Proposed Model	2024	Ensemble Capsule Network	99.81	99.89	99.40	99.64

5. Conclusion and future scope

The increasing incidence of diabetes-related complications, especially diabetic retinopathy (DR), is a serious problem worldwide and diabetic retinopathy is the most important cause of vision impairment across the globe. The grading and detection of DR lesions by ophthalmologists manually are an arduous and time-consuming process. This study aimed to tackle this issue by utilizing advances in deep learning methods to create an automated DR diagnosis and classification system. By combining a number of pre-trained deep learning networks layers with capsule network layers for feature extraction and employing a stochastic ensemble process for classification with fundus images, our proposed solution provides an opportunity to assist ophthalmologists with accurate diagnoses and grading of DR. Evaluation of the proposed model using the Aptos-19 dataset showed stellar results, with an overall testing accuracy of 99.81%. The accuracy highlights the potential of the method for improving both the efficiency and effectiveness of DR diagnosis and grading. Future work could explore if integration with an interactive visual analytics dashboards, enabling dynamic searching and exploration of DR diagnosis, lesion heatmaps, and class-specific prediction trajectories, would provide transparency and decision support and encourage clinical implementation. Further research and validation on diverse datasets and real-world clinical settings will be essential to confirm the generalizability and robustness of the proposed methodology. Additionally, efforts to streamline the integration of automated DR diagnosis systems into clinical practice and healthcare workflows are warranted to ensure widespread adoption and impact.

Conflict of interest

The authors declare that we have no conflict of interest. On behalf of all authors, the corresponding author states that there is no conflict of interest.

Acknowledgements

APTOS 2019 datasets are freely available to the public and were contributed by Kaggle, which the authors thank for this

References

1. Kropp, M., Golubnitschaja, O., Mazurakova, A., Koklesova, L., Sargheini, N., Vo, T. K. S., de Clerck, E., Polivka, J., Jr, Potuznik, P., Polivka, J., Stetkarova, I., Kubatka, P., & Thumann, G. (2023). Diabetic retinopathy as the leading cause of blindness and early predictor of cascading complications—Risks and mitigation. *The EPMA Journal*, 14(1), 21–42. <https://doi.org/10.1007/s13167-023-00314-8>
2. Shin, E. S., Sorenson, C. M., & Sheibani, N. (2014). Diabetes and retinal vascular dysfunction. *Journal of Ophthalmic and Vision Research*, 9(3), 362–373. <https://doi.org/10.4103/2008-322X.143378>
3. Gargeya R., Leng T. Automated Identification of Diabetic Retinopathy Using Deep Learning. *Ophthalmology*. 2017;124:962–969. doi: 10.1016/j.ophtha.2017.02.008.
4. International Diabetes Federation. (2022, November 24). *Facts & figures*. <https://idf.org/about-diabetes/diabetes-facts-figures/>
5. Press Information Bureau. (2025, July 20). *Update on treatment of diabetes*. <https://pib.gov.in/PressReleasePage.aspx?PRID=1944600>
6. Alyoubi, W. L., Shalash, W. M., & Abulkhair, M. F. (2020). Diabetic retinopathy detection through deep learning techniques: A review. *Informatics in Medicine Unlocked*, 20, 100377. <https://doi.org/10.1016/j.imu.2020.100377>
7. Tymchenko, B., Marchenko, P., & Spodarets, D. (2020). *Deep learning approach to diabetic retinopathy detection* [Preprint]. arXiv. <https://arxiv.org/abs/2003.11544>
8. Gangwar, A. K., & Ravi, V. (2021). Diabetic retinopathy detection using transfer learning and deep learning. In V. Bhateja, S. L. Peng, S. C. Satapathy, & Y. D. Zhang (Eds.), *Evolution in computational intelligence* (Vol. 1176). Springer. https://doi.org/10.1007/978-981-15-5788-0_64
9. Karki, S. S., & Kulkarni, P. (2021). Diabetic retinopathy classification using a combination of EfficientNets. In *2021 International Conference on Emerging Smart Computing and Informatics (ESCI)* (pp. 68–72). IEEE. <https://doi.org/10.1109/ESCI50559.2021.9397035>
10. Mishra, S., Hanchate, S., & Saquib, Z. (2020). Diabetic retinopathy detection using deep learning. In *2020 International Conference on Smart Technologies in Computing, Electrical and Electronics (ICSTCEE)* (pp. 515–520). IEEE. <https://doi.org/10.1109/ICSTCEE49637.2020.9277506>
11. A., A. M., & Priya, S. S. S. (2025). A novel deep learning approach for diabetic retinopathy classification using optical coherence tomography angiography. *Multimedia Tools and Applications*. Advance online publication. <https://doi.org/10.1007/s11042-025-20708-2>
12. Abini, M. A., & Sridevi Sathya Priya, S. (2025). Detection and classification of diabetic retinopathy using modified Inception V3. *International Journal Bioautomation*, 29(1), 77–92. <https://doi.org/10.7546/ijba.2025.29.1.001004>
13. A., M. A., & Sathya Priya, S. S. (2025). A survey on computer aided systems for diabetic retinopathy detection and classification using deep learning. In *2025 International Conference on Electronics and Renewable Systems (ICEARS)* (pp. 1475–1480). IEEE. <https://doi.org/10.1109/ICEARS64219.2025.10940665>
14. A., M. A., & Sathya Priya, S. S. (2023). A deep learning framework for detection and classification of diabetic retinopathy in fundus images using residual neural networks. In *2023 9th International Conference on Smart Computing and Communications (ICSCC)* (pp. 55–60). IEEE. <https://doi.org/10.1109/ICSCC59169.2023.10335079>
15. Mushtaq, G., & Siddiqui, F. (2021). Detection of diabetic retinopathy using deep learning methodology. *IOP Conference Series: Materials Science and Engineering*, 1070(1), 012049. <https://doi.org/10.1088/1757-899X/1070/1/012049>

16. AbdelMaksoud, E., Barakat, S., & Elmogy, M. (2022). A computer-aided diagnosis system for detecting various diabetic retinopathy grades based on a hybrid deep learning technique. *Medical & Biological Engineering & Computing*, 60, 2015–2038. <https://doi.org/10.1007/s11517-022-02564-6>
17. Chowdhury, P., Islam, M. R., Based, M. A., & Chowdhury, P. (2021). Transfer learning approach for diabetic retinopathy detection using efficient network with 2 phase training. In 2021 6th International Conference for Convergence in Technology (I2CT) (pp. 1–6). IEEE. <https://doi.org/10.1109/I2CT51068.2021.9418111>
18. A., M. A., & Sathya Priya, S. S. (2023). Detection and classification of diabetic retinopathy using pretrained deep neural networks. In 2023 International Conference on Innovations in Engineering and Technology (ICIET) (pp. 1–7). IEEE. <https://doi.org/10.1109/ICIET57285.2023.10220715>
19. Abini, M. A., & Priya, S. S. S. (2025). Advanced capsule networks for accurate detection and classification of diabetic retinopathy from fundus images. In S. Manoharan, A. Tugui, & I. Perikos (Eds.), *Proceedings of 5th International Conference on Artificial Intelligence and Smart Energy (ICAIS 2025) (Information Systems Engineering and Management, vol. 41)*. Springer. https://doi.org/10.1007/978-3-031-90478-3_37
20. Abini, M. A., & Priya, S. S. S. (2025). Automatic detection and classification of diabetic retinopathy from optical coherence tomography angiography images using deep learning – A review. *AIUB Journal of Science and Engineering*, 23(3), 277–297.
21. A., M. A., Vinod, A., Rafeeqe, S., Manju, T. S., & Noushad, M. S. (2024). MobileNet-enhanced skin cancer detection and classification using dermatoscopic images. In 2024 IEEE International Conference on Smart Power Control and Renewable Energy (ICSPCRE) (pp. 1–6). IEEE. <https://doi.org/10.1109/ICSPCRE62303.2024.10674858>
22. Khan, A., Chefranov, A., & Demirel, H. (2023). Building discriminative features of scene recognition using multi-stages of inception-ResNet-v2. *Applied Intelligence*, 53(15), 18431–18449. <https://doi.org/10.1007/s10489-023-04460-4>
23. Howard, A. G., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T., Andreetto, M., & Adam, H. (2017). MobileNets: Efficient convolutional neural networks for mobile vision applications. arXiv:1704.04861. <https://doi.org/10.48550/arXiv.1704.04861>
24. Rangel, G., Cuevas-Tello, J. C., Rivera, M., & Renteria, O. (2023). A Deep Learning Model Based on Capsule Networks for COVID Diagnostics through X-ray Images. *Diagnostics*, 13(17), 2858. <https://doi.org/10.3390/diagnostics13172858>
25. M. A., A., & Sridevi Sathya Priya, S. (2025). Deep learning-based diabetic retinopathy classification of augmented fundus images using convolutional neural networks. *International Journal of Image and Graphics*, 25(03), 2750040. <https://doi.org/10.1142/S0219467827500409>
26. A. M., A., & Priya, S. S. S. (2025). Ensemble models for diabetic retinopathy detection and classification using vision transformers and capsule networks with advanced feature extraction techniques. *Australian Journal of Electrical and Electronics Engineering*, 1–22. <https://doi.org/10.1080/1448837X.2025.2539557>
27. Al-Antary, M. T., & Arafa, Y. (2021). Multi-scale attention network for diabetic retinopathy classification. *IEEE Access*, 9, 54190–54200. <https://doi.org/10.1109/ACCESS.2021.3070685>
28. Oulhadj, M., Riffi, J., Chaimae, K., El Hassani, A., & Merbouha, A. (2022). Diabetic retinopathy prediction based on deep learning and deformable registration. *Multimedia Tools and Applications*, 81, 28709–28727. <https://doi.org/10.1007/s11042-022-12968-z>
29. Crane, A. B., Choudhry, H. S., & Dastjerdi, M. H. (2024). Effect of simulated cataract on the accuracy of artificial intelligence in detecting diabetic retinopathy in color fundus photos. *Indian Journal of Ophthalmology*, 72(Suppl 1), S42–S45. https://doi.org/10.4103/IJO.IJO_1163_23
30. Escorcia-Gutierrez, J., et al. (2022). Analysis of pre-trained convolutional neural network models in diabetic retinopathy detection through retinal fundus images. In: Saeed, K.,

Dvorský, J. (eds.) Computer Information Systems and Industrial Management. CISIM 2022. Lecture Notes in Computer Science, vol. 13293, pp. 167–178. Springer, Cham. https://doi.org/10.1007/978-3-031-10539-5_15

31. Salluri, D. K., Sistla, V., & Kolli, V. K. K. (2023). HRUNET: Hybrid residual U-Net for automatic severity prediction of diabetic retinopathy. *Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization*, 11(3), 530–541. <https://doi.org/10.1080/21681163.2022.208302>

32. Kobat, S. G., Baygin, N., Yusufoglu, E., Baygin, M., Barua, P. D., Dogan, S., Yaman, O., Celiker, U., Yildirim, H., Tan, R. S., Tuncer, T., Islam, N., & Acharya, U. R. (2022). Automated diabetic retinopathy detection using horizontal and vertical patch division-based pre-trained DenseNet with digital fundus images. *Diagnostics*, 12(8), 1975. <https://doi.org/10.3390/diagnostics12081975>

33. Mácsik, P., Pavlovičová, J., Goga, J., & Kajan, S. (2022). Local binary CNN for diabetic retinopathy classification on fundus images. *Acta Polytechnica Hungarica*, 19(7), 27–45. <https://doi.org/10.12700/APH.19.7.2022.7.2>

34. Yadav, S., & Awasthi, P. (2022). Diabetic retinopathy detection using deep learning and Inception-V3 model. *International Research Journal of Modern Engineering and Technology Science*, 4, 1731–1735.

35. Canayaz, M. (2022). Classification of diabetic retinopathy with feature selection over deep features using nature-inspired wrapper methods. *Applied Soft Computing*, 128, 109462. <https://doi.org/10.1016/j.asoc.2022.109462>

36. Bodapati, J. D., Rohith, V. N., & Dondeti, V. (2022). Ensemble of deep capsule neural networks: An application to pediatric pneumonia prediction. *Physical and Engineering Sciences in Medicine*, 45(3), 949–959. <https://doi.org/10.1007/s13246-022-01169-5>

37. “APTOS 2019 blindness detection,” Kaggle.com. [Online]. Available: <https://www.kaggle.com/c/aptos2019-blindness-detection/data>. [Accessed: 2-July -2025].

38. Sikder, N., Masud, M., Bairagi, A. K., Arif, A. S. M., Nahid, A.-A., & Alhumyani, 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>

39. Islam, M. R., Abdulrazak, L. F., Nahiduzzaman, M., Goni, M. O. F., Anower, M. S., Ahsan, M., Haider, J., & Kowalski, M. (2022). Applying supervised contrastive learning for the detection of diabetic retinopathy and its severity levels from fundus images. *Computers in Biology and Medicine*, 146, 105602. <https://doi.org/10.1016/j.compbimed.2022.105602>

40. Oulhadj, M., Riffi, J., & Chaimae, K. (2022). Diabetic retinopathy prediction based on deep learning and deformable registration. *Multimedia Tools and Applications*, 81, 28709–28727. <https://doi.org/10.1007/s11042-022-12659-8>

41. Bodapati, J. D. (2022). Stacked convolutional auto-encoder representations with spatial attention for efficient diabetic retinopathy diagnosis. *Multimedia Tools and Applications*, 81, 32033–32056. <https://doi.org/10.1007/s11042-022-12942-8>

42. Oulhadj, M., Riffi, J., Khodriss, C., Mahraz, A. M., Bennis, A., Yahyaouy, A., Chraibi, F., Abdellaoui, M., Andaloussi, I. B., & Tairi, H. (2023). Diabetic retinopathy prediction based on wavelet decomposition and modified capsule network. *Journal of Digital Imaging*, 36(4), 1739–1751. <https://doi.org/10.1007/s10278-023-00872-x>

43. Mondal, S. S., Mandal, N., Singh, K. K., Singh, A., & Izonin, I. (2023). EDLDR: An ensemble deep learning technique for detection and classification of diabetic retinopathy. *Diagnostics*, 13(1), 124. <https://doi.org/10.3390/diagnostics13010124>

44. Lei, Y., Lin, S., Li, Z., Zhang, Y., & Lai, T. (2024). GNN-fused CapsNet with multi-head prediction for diabetic retinopathy grading. *Engineering Applications of Artificial Intelligence*, 133, 107994. <https://doi.org/10.1016/j.engappai.2023.107994>

45. Abini, M. A., & Sridevi Sathya Priya, S. (2025). Multistage classification of diabetic retinopathy in fundus images using hybrid capsule networks. *Journal of Multiscale Modelling*. <https://doi.org/10.1142/S1756973725500064>

46. Abini, M.A., Sridevi Sathya Priya, S. (2025). DRDResNet—A Deep Learning Model for Diabetic Retinopathy Detection and Classification. In: Kanhe, A., Balanethiram, S., Hsiung, PA., Jayakody, D.N.K. (eds) *Advances in VLSI, Signal Processing and Wireless Communication*. IEdTC 2023. *Lecture Notes in Electrical Engineering*, vol 1323. Springer, Singapore. https://doi.org/10.1007/978-981-96-1587-2_12
47. Abini, M.A., Priya, S.S.S. (2025). An Augmented Efficient Net Based Deep Learning Model for Diabetic Retinopathy Detection and Classification. In: Soni, B., Saini, P., Verma, G.K., Gupta, B.B. (eds) *Beyond Artificial Intelligence*. AICTA 2023. *Lecture Notes in Networks and Systems*, vol 1326. Springer, Singapore. https://doi.org/10.1007/978-981-96-4170-3_26
48. Naseer, R. K., Afzal, A., Arshaq, M., R. K. M., & A. M. A. (2025). RiViT – Rice leaf disease identification and classification using vision transformer (ViT). 2025 4th International Conference on Advances in Computing, Communication, Embedded and Secure Systems (ACCESS), 245–250. IEEE. <https://doi.org/10.1109/ACCESS65134.2025.11135614>