



ENSEMBLE TRANSFER LEARNING MODEL FOR THE DETECTION OF DIABETIC RETINOPATHY DISEASE

Ishaq Aiche¹, Youcef Brik², Bilal Attallah³ and Oussama Bouguerra⁴

^{1, 2, 3, 4}LASS, Laboratory of Signal and Systems Analysis, Department of Electronics, Faculty of Technology, University of M'sila, Algeria.

¹<http://orcid.org/0009-0004-5133-4745>, ²<http://orcid.org/0000-0003-1565-0594>,

³<http://orcid.org/0000-0003-3661-957X>, ⁴<http://orcid.org/0009-0001-8564-5291>

Email: ishaq.aiche@univ-msila.dz, youcef.brik@univ-msila.dz, bilal.attallah@univ-msila.dz, oussama.bouguerra@univ-msila.dz

ARTICLE INFO

Article History

Received: January 25, 2026

Reviewed: March 26, 2026

Accepted: April 26, 2026

Published: May 29, 2026

Keywords:

Diabetic Retinopathy Detection,
Transfer Learning,
Deep Learning,
Model Fusion,
Ensemble Classifier Methods.

ABSTRACT

Diabetic retinopathy (DR) is a leading cause of vision loss, and early detection is essential for preventing blindness. This study investigates using transfer learning, deep learning and model fusion techniques to improve DR detection. Five pre-trained Convolutional Neural Networks (CNNs) ResNet50, VGG16, VGG19, DenseNet121 and EfficientNet-B7 were evaluated individually and in combination using the APTOS 2019 RGB retinal image dataset. Transfer learning was applied to fine-tune the models for the binary classification of DR and non-DR cases. Four fusion strategies were implemented to exploit the complementary strengths of the individual models. The experimental results demonstrate that all fusion approaches outperformed the standalone CNNs in terms of accuracy, sensitivity, specificity and Area Under the Curve (AUC). Of the individual models, DenseNet121 and EfficientNet-B7 performed best. Fuzzy-based fusion strategies further improved diagnostic accuracy. The best fusion model achieved an accuracy of 98.64% and a recall of 99.19%, demonstrating the robustness of the proposed approach. These findings highlight the potential of integrating transfer learning with advanced model fusion to create automated, reliable and scalable systems for the early screening of Diabetic Retinopathy (DR). Such systems could facilitate timely clinical intervention, enhance patient outcomes, and contribute to more efficient healthcare delivery.



Copyright ©2016 by authors and Galileo Institute of Technology and Education of the Amazon (ITEGAM). This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

I. INTRODUCTION

Diabetic retinopathy (DR) is a condition that causes vision problems and blindness. It is one of the main causes of vision loss among people with diabetes. It is very important to spot DR early so that the disease does not get worse. Nevertheless, it is hard to spot DR in its early stages. Manual retinal image analysis is time-consuming, subject to inter-observer variability, and requires expert interpretation. To address these limitations, automated detection systems powered by artificial intelligence (AI) and deep learning have emerged as promising tools to enhance the speed and accuracy of DR diagnosis [1]. Deep learning techniques, particularly convolutional neural networks (CNNs), have shown outstanding success in medical imaging tasks, including the classification and segmentation of retinal images. However, training these models from scratch demands large, well-labeled datasets, which are often difficult to obtain in the medical domain. Transfer learning offers a practical solution by enabling the adaptation of pre-trained models—originally trained on large-scale datasets like ImageNet—to domain-specific tasks. Architectures such as ResNet50, VGG16, VGG19, DenseNet121, and EfficientNet-B7 have proven highly effective when fine-tuned for DR detection [2], [3]. Although individual CNNs demonstrate strong predictive capabilities, combining multiple models through fusion strategies can further improve reliability and generalizability. Model fusion techniques harness the diverse strengths of various architectures, resulting more robust diagnosis systems. Prior studies indicate that ensemble methods often outperform individual models across multiple performance metrics [4]. This research extends current efforts by evaluating four fusion techniques: voting, averaging, fuzzy rank-based fusion, and the Sugeno fuzzy integral. These methods are applied to integrate the predictions of ResNet-50, VGG-16, VGG-19, DenseNet-121, and EfficientNet-B7 into a unified decision-making framework [5].

The study is based on the APTOS 2019 Blindness Detection dataset, which contains a wide range of retinal images representing various stages of diabetic retinopathy [6], [7]. The central research question addressed in this work is ‘can the fusion of multiple fine-tuned CNN models improve diabetic retinopathy detection performance compared to individual models?’ To answer this question, the study pursues the following objectives:

- Apply transfer learning and fine-tuning to adapt advanced CNN architectures for binary classification of diabetic retinopathy.
- Develop and implement a model fusion framework incorporating voting, averaging, fuzzy rank-based fusion, and Sugeno fuzzy integral strategies.
- Evaluate and compare the performance of individual and ensemble models using standard metrics such as accuracy, sensitivity, specificity, F1-score, and the area under the receiver operating characteristic curve (ROC-AUC).

The hypothesis is that ensemble methods combining the outputs of diverse CNNs will yield superior diagnostic performance and greater reliability than any single model. This research presents a scalable and accurate framework for automated diabetic retinopathy screening. It has the potential to significantly enhance early detection, improve clinical decision-making, and optimize the use of healthcare resources. Nonetheless, certain limitations are acknowledged, including reliance on a single dataset and the computational demands of ensemble modeling, which may affect deployment in resource-limited settings. The structure of the paper is as follows: Section II reviews work relating to the detection of diabetic retinopathy using deep learning and transfer learning methods. Section III details the proposed methodology, including data pre-processing, model architectures and ensemble fusion strategies. Section IV describes the experimental setup and evaluation metrics; it also presents and discusses the results. Finally, Section V is the conclusion of the study.

II. RELATED WORK

Various strategies have been explored to address the issue of diabetic retinopathy grading, including the development of distinctive features and the creation of software that analyzes retinal images. This strategy is intricate and requires much effort. The ophthalmologist must also apply his expertise in creating characteristics and annotating retinal images. This requirement is a big issue that must be addressed [8]. Many conventional methods need to ascertain the answers and comprehend the interconnections. The retention of extraneous image-specific attributes, such as rotation, brightness, and size, results in a negative impact on performance and accuracy, instead of facilitating the acquisition of relevant data [9]. The emergence of AI methodologies, including machine learning and deep learning, has enabled high-performance optical detection and classification to identify and segregate affected retinal regions [10]. The transfer learning strategy is based on its previous training on a large image database and can be adapted to the required dataset [11]. This technique is popular because there are many pre-trained models available, and there is a lack of labelled data for various real-world problems [12]. Quadratic weighted kappa was used in a study by Chilukoti to analyse the prediction of diabetic retinopathy [13]. The categorization using the pre-trained EfficientNet B3 model. Viji's work [14] used many models, including VGG, Xception, ResNet, Inception and EfficientNet, achieving an accuracy of 99.36% and a recall of 0.986. In [15], a CNN model was used to collect features from the input dataset, and a support vector machine model was used to perform classification. This method provided quicker execution than classifying with the CNN model. Medical imaging is essential for all basic health challenges and diverse medical conditions.

Diabetic retinopathy grading using fundus photographs has been extensively researched, employing various transfer learning methodologies utilizing publically accessible data sets. According to [16] utilized Inception [17], VGG16 [18], DenseNet [19], MobileNet [20], and ResNet models [21], with VGG16 achieving a 100% accuracy rate. In CNN models, using too many layers without spatial pooling results in a decrease in learnable parameters, which simplifies training [22]. Kothare and Malpe [23] developed a support vector machine and Naive Bayes model with a binary pattern methodology. The algorithm was utilized to choose relevant features. In contrast, the models were used to categorize patients as having the illness. The support vector machine is superior to the naive Bayes classifier in terms of accuracy, execution time, and memory utilization. According to [24] tried to create a simple way to predict and classify difficult exudates using something called artificial neural networks (ANN). The Speed-Up Robust Features approach was used to conduct feature extraction and detection. For classification purposes, a Feed-Forward Backpropagation Artificial Neural Network was utilized. The primary shortcoming of the work is its reliance on a dataset including merely 48 photographs. Ahmad et al. compare many architectures, including Inception-ResNetV2, ResNet50, NASNet, InceptionV3, VGG16, DenseNet121, Xception, and VGG19 in the publication [25]. The models performed classification and localization tasks, they were trained using a special set of data and then tested against the Messidor-2 dataset. The CNN and Contrast Limiting Adaptive Histogram Equalization (CLAHE) methods made the area under the curve better and improved the quality of fundus pictures.

The CLAHE approach has been implemented to magnify the retinal vessels, improving the preprocessing stage. Multiple machine learning classifiers have been integrated to improve diabetic retinopathy identification and decrease the error rate [26]. Fundus imaging can be used for precise diagnosis and assessment. An ophthalmologist requires extensive knowledge and commitment to manually analyze diabetic retinopathy using fundus images. Diabetic retinopathy was accurately classified using deep learning methodologies such as convolutional neural networks and transfer learning with models like ResNet, VGG, and GoogleNet [16], [27], [28]. In a prior article by [29], the authors utilized various SVM kernels in conjunction with the Ant Colony System (ACS) feature selection approach. It assesses a method for identifying diabetic retinopathy utilizing 250, 550, and 750 characteristics. The cubic SVM classifier surpassed all other kernels, with 92.6%, 91.8%, and 93% accuracy for 250, 550, and 750 attributes, respectively. According to [30] employed an Extreme Learning Machine (ELM) model on the APTOS-2019 dataset, achieving an accuracy of 97.27%. U-Net is a convolutional neural network design that is applied to image segmentation tasks. An asymmetric deep learning architecture for diabetic retinopathy screening was utilized in the study by [31] to perform segmentation. CLAHE is employed to assess and enhance photographs in the green channel. APTOS and MESSIDOR detected non-DRs with accuracy levels of 98.6% and 91.9% respectively, and PrDRs with accuracy levels of 96.8% and 95.76% respectively. Data augmentation is the predominant technique for mitigating imbalance difficulties in picture categorization [32]. It is a compilation of methodologies for generating additional data points from current data to augment data volume artificially.

Different methods can enhance the photos, including cropping, horizontal or vertical inversion, scaling, and rotation [33]. According to [34] demonstrates that employing a data augmentation technique improves the efficacy of transfer learning in the classification of diabetic retinopathy images.

III. MATERIALS AND METHODS

III.1 PROPOSED METHOD

In this study, deep learning algorithms are proposed for analyzing retinal fundus images to detect and categorize diabetic retinopathy (DR), which is classified into three main phases.

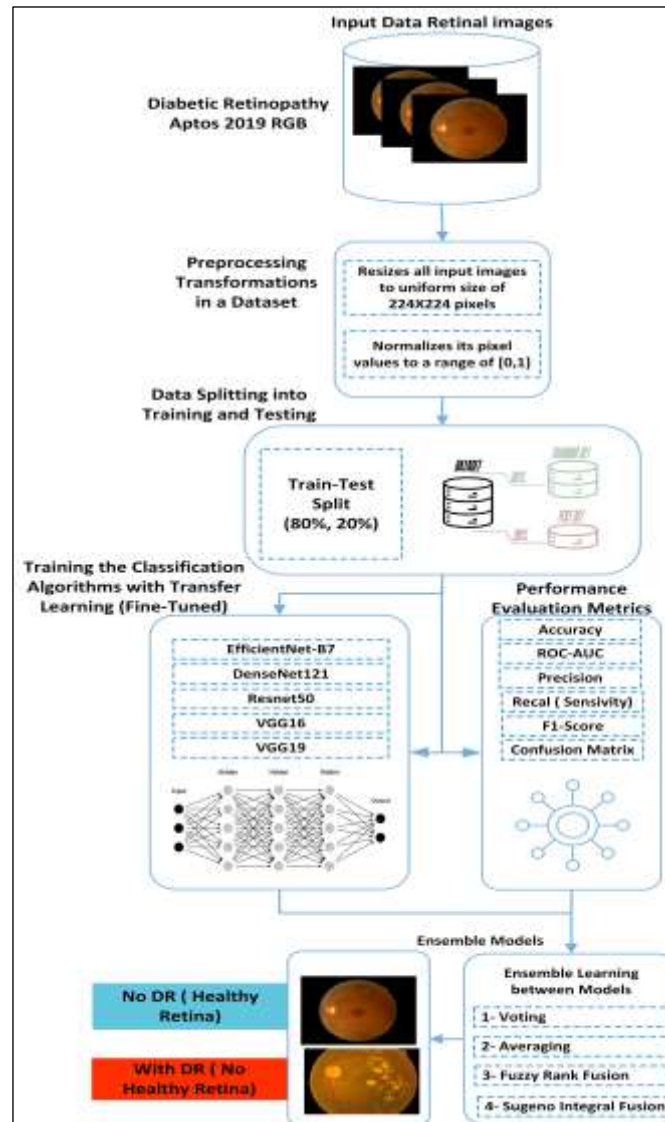


Figure 1: Proposed workflow of methodology for the Diabetic Retinopathy (DR) Prediction.

Source: Authors, (2026).

In the first phase, the fundus image dataset is pre-processed and enriched to reduce variability and address class imbalance. In the second phase, we fine-tune the last layers of the pre-trained convolutional neural networks - specifically ResNet50, VGG16, VGG19, DenseNet121 and EfficientNet-B7 - using the prepared dataset to benefit from their feature extraction capabilities. In the third and final phase, a fusion strategy is introduced to combine the outputs of the individual models in order to improve diagnostic robustness and accuracy. Four fusion methods are explored: majority voting, probability averaging, fuzzy rank-based fusion and the Sugeno fuzzy integral. Together, these steps provide a comprehensive and reliable framework for automated DR detection. The proposed methodology is illustrated in Figure 1.

III.2 DATASET

This research solely employs the Asia Pacific Tele-Ophthalmology Society (APTOS) dataset [35], concentrating on its standard and Diabetic Retinopathy (DR) categories. Images designated as DR are categorized as class 1, while standard retina images are categorized as class 0. The APTOS dataset was selected for its thorough representation of retinal anomalies, facilitating an in-depth examination of diabetic retinopathy diagnosis by deep learning methodologies. The research ensures consistent preparation, training, and assessment protocols by focusing on a solitary, high-quality dataset.

The severity of diabetic retinopathy can be assessed accurately and reliably using the APTOS DR Severity Scale. It has been widely used in clinical therapy and research throughout the Asia-Pacific region. It has been employed in developing and evaluating computer-aided diagnostic tools for diabetic retinopathy that autonomously detect and categorize the severity of the problem using retinal images. The dataset consists of 5990 retinal images, with 3662 assigned to the training set and 1928 allotted for testing. For this study, the 3,662 training data were obtained because the test phase data lacks specific labels. The number of images in class 0 is 1805 and the number of images in class 1 is 1857. Figure 2 illustrates a selection of these images, highlighting the diversity in the APTOS 2019 dataset.

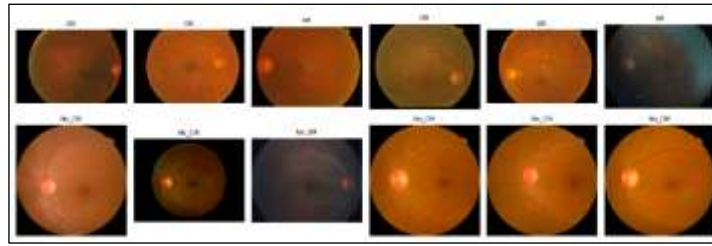


Figure 2: Visualizing the Diverse Classes of Diabetic Retinopathy Images.
Source: Authors, (2026).

III.3 DATA PRE-PROCESSING

This section looks at the dataset used and the ways it was prepared. The images are quite big, measuring 4288 by 2848. Figure 3 shows how images are spread across the training and validation sets. It also shows how many images from the 'No_DR' and 'With_DR' groups there are. The collection includes noisy images and labels, including blurred, overexposed or underexposed images. This variation is the result of data collected from many clinics using different cameras over a long period. This section explains the steps taken to get the dataset ready for the learning phase. The aim is to make training on the dataset faster. We changed the size of the original images to 224×224 . Then we used different batch sizes to train this collection of smaller images. An 80:20 ratio is used for selecting the model using the train-test-split method. Adjusting the model or finding patterns in the data can be achieved by using the training set. The test set checks how well the model works, especially with new and unknown data. There are 1444 training images for class 0 (no DR) and 1485 for class 1 (with DR). The numbers 361 and 372 are used to show the different classes. Class 0 is shown by the number 361 (No DR) and class 1 by the number 372 (With DR).

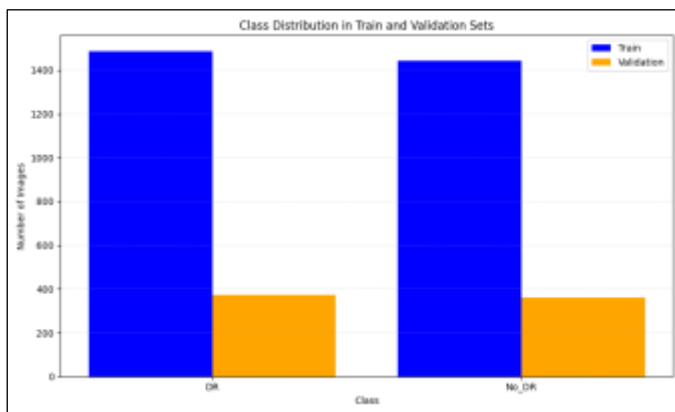
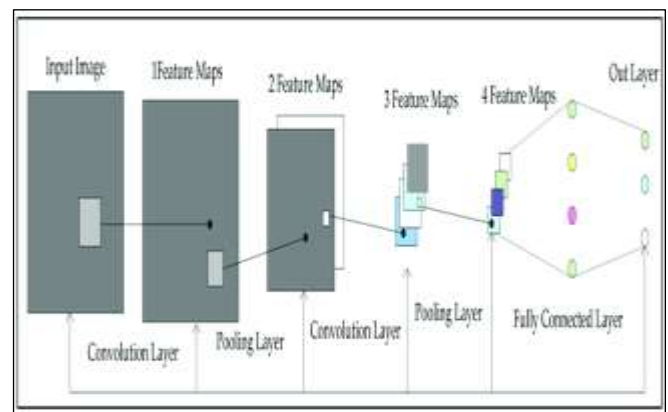


Figure 3: Distribution of Images in Training and Validation Datasets across 'No_DR' and 'With_DR' Categories.
Source: Authors, (2026).

III.4 DATA PRE-PROCESSING

Many researchers, particularly those specialising in medical image processing, prefer to use transfer learning (TL) to develop effective models, rather than training deep learning (DL) models such as convolutional neural networks (CNNs) from scratch [36]. In this study, we have implemented a method for detecting diabetic retinopathy (DR) using a multistage learning process that incorporates feature extraction, fine-tuning approaches, and, ultimately, the fusion of all models. Transfer learning is a common technique used to train neural networks to achieve optimal results. It involves training a model from start to finish on a dataset with sufficient data from a related domain, and then using the limited data from the target dataset to refine it. Rather than starting from scratch, this approach builds a robust foundation for the model [37]. One of the main limitations of CNNs is that they require a large number of sample training images. The number and depth of parameters determine how much data is required for a network. Networks with more layers and parameters require more data [38], [39]. Otherwise, overfitting would negatively affect performance. However, one of the main problems in creating DL models for the medical industry is that there is not enough data. Therefore, using the idea of transfer learning for medical projects would be a good idea [8], [40]. The Visual Geometry Group (VGG) at the University of Oxford created a model called VGG16 in 2014 (see Figure 4). The main framework, VGGNet, has led to several modifications, including VGG1, VGG13, VGG16 and VGG19. Out of all these models, people use the last two versions, VGG16 and VGG19, the most. The VGG16 model is widely used in medical applications to recognise and categorise different types of clinical images [27]. The learnable parameter layers of the model are represented by its weighted layers, which total 16. Despite its modest kernel size, it is very time consuming to train [15].



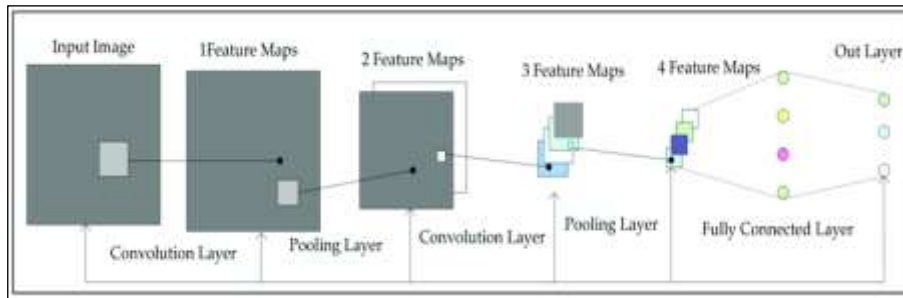


Figure 4: Architecture of the VGGNet Model Showing Convolutional, Pooling, and Fully Connected Layers.
Source: [41].

The DenseNet121 model (Figure 5) has several advantages, including strong gradient flow, efficient computation, smooth error signal propagation, and reduced feature complexity [15]. The main aim in creating this network was to solve the problem of the gradient vanishing as the number of layers increases. If you lose or delete information, the network gets deeper. DenseNet121 solves this problem by connecting the layers of the network [42].

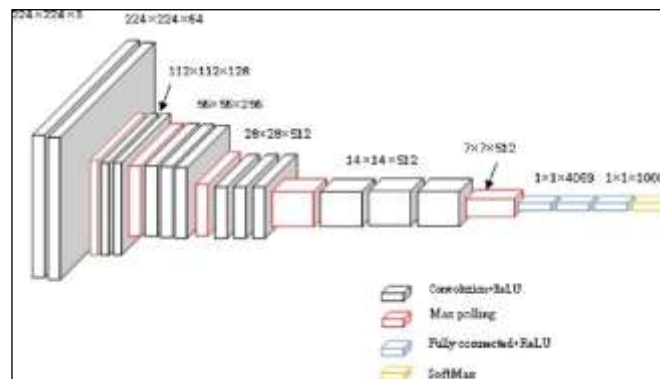


Figure 5: Architecture of the DenseNet121 Model.
Source: [43].

ResNet (Residual Network), introduced by [44] in 2015, is a special type of deep learning network that solves the problem of vanishing gradients by using skip connections (residual connections). These connections help the network to understand how different things are connected, which makes it easier to train very complex networks.

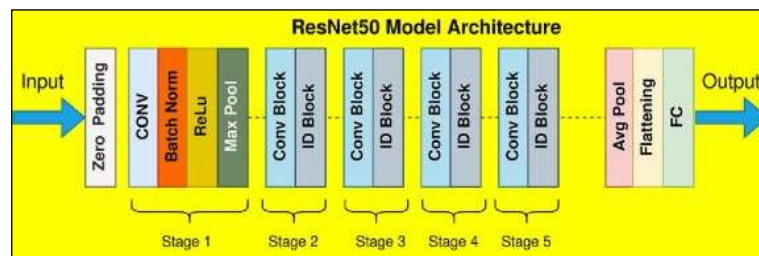


Figure 6: Architecture of the ResNet50 Model.
Source: [43].

EfficientNet is a family of convolutional neural networks developed by Tan and Le (2019) at Google AI (Figure 7). It is designed to optimise both accuracy and efficiency by systematically scaling the depth, width, and resolution of the network using a compound scaling method [45]. There are eight variations, ranging from B0 to B7. Each subsequent model number refers to a variant with more parameters and higher accuracy.

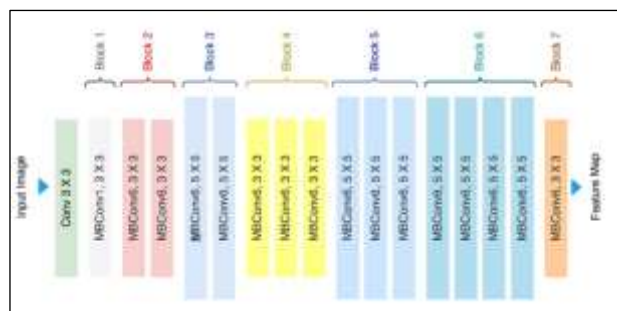


Figure 7: Architecture of the Efficient-Net Model.
Source: [45].

III.5 ENSEMBLE CLASSIFIER METHODS

In this study, the authors present a novel ensemble-based method for combining predictions from multiple deep learning models. The proposed approach exploits the principles of ensemble learning by aggregating the prediction probabilities of individual models to generate a unified, fused score. This strategy increases the robustness and accuracy of the final decision. Figure 7 shows the basic ideas of ensemble learning as used in the suggested system.

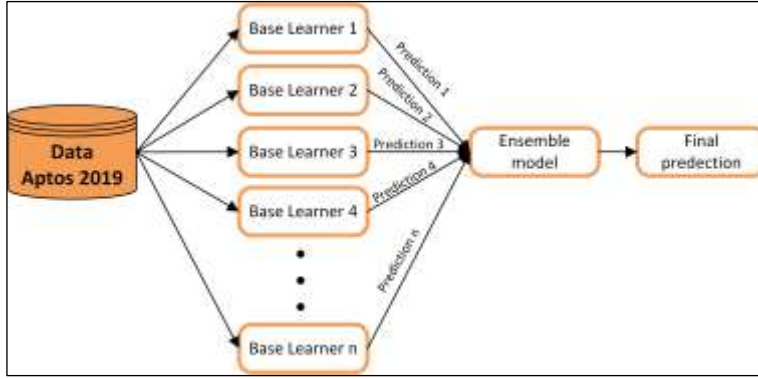


Figure 8: Ensemble Learning Strategy for Combining Predictions from Multiple Models.

Source: Authors, (2026).

Majority voting: The soft voting approach determines the predicted class of diabetic retinopathy by selecting the class label $i=1$ to n that has the highest average probability across all models $j=1$ to m [46]. Equation 1.

$$Fp = \text{label}(\max(p') \parallel p'_i = \sum_{j=1}^m p_{ij}) \quad (1)$$

Fp : Final prediction

Where $i=1$ to n .

Average (arithmetic mean): The simplest method of combining several scores is to take the average Equation (2). Add together all the scores and divide by the total number of scores to find the mean [47].

$$Fp = \frac{1}{n} \times \sum_{i=1}^n S_i \quad (2)$$

Fuzzy rank: According to the fuzzy rank, a rank is given to each model's prediction score. These ranks are then combined to provide a final prediction [48] (see Algorithm 1 for more information). Two non-linear functions are used for each base classifier to obtain the fuzzy ranks. These amounts are then distributed among the classifiers once the ranks have been added together. Finally, a prediction is made for the class with the lowest total score.

Algorithm 1: Fuzzy Rank Algorithm.

Input:

C : Number of classes

BC : Number of base classifiers

i : Base classifier index (1 to BC)

$P[i][k]$: Confidence scores predicted by base classifier i for each class k

Output:

$\text{class}(I)$: Predicted class for input I

Algorithm:

for each base classifier i **do**

for each class k **do**

 Calculate fuzzy ranks

$$R_1[i][k] = 1 - e^{-\left(\frac{P_{ik}-1}{2}\right)^2}$$

$$k = 1, 2, 3, \dots, C; i = 1, 2, 3, \dots, BC$$

$$R_2[i][k] = 1 - \tanh\left(-\left(\frac{P_{ik}-1}{2}\right)^2\right)$$

$$k = 1, 2, 3, \dots, C; i = 1, 2, 3, \dots, BC$$

 Fuse ranks

$$RS[i][k] = R_1[i][k] \times R_2[i][k]$$

end for

end for

for each class k **do**

 Sum fuse scores across base classifiers

$$FS[k] = \sum_{i=1}^{BC} RS[i][k]$$

end for

predicted class

$$\text{class}(I) = \arg \min(FS[k])$$

Source: Authors, (2026).

Sugeno integral: The Sugeno fuzzy integral lets you add together the outputs of multiple classifiers according to fuzzy measures that represent how relevant the sources are. I have used it to make predictions about diabetic retinopathy using information from CNNs that have already been trained [49]. The Sugeno integral method is defined by considering μ as a fuzzy measure on a set $A = \{a_1, a_2, \dots, a_n\}$. The ranges of these elements are represented by, where is a function mapping A to the interval $[0, 1]$ such that $(a1) \leq (a2) \leq \dots \leq (an)$. The Sugeno integral can be expressed by Equation (3) as:

$$\int_{i=1}^n r(a) d\mu = \max_{1 \leq x \leq n} \left(\min(r(a_x), \mu(A_x)) \right) \quad (3)$$

The fuzzy measures of $\{a_i\}$, denoted as $(\{a_i\})$, are primarily determined by domain experts for a specific statistical problem. These fuzzy measures help to compute $(A \times x)$ using the Sugeno λ fuzzy measures $\lambda \in (-1, \infty)$ defined by Eq. (4):

$$1 + \lambda = \prod_{i=1}^n (\mu(\{a_i\})^\lambda + 1) \quad (4)$$

After obtaining the values of λ from Eq. (4), the value of $\mu(A \times x)$ is calculated using Eq. (5):

$$\mu(\{a_1, a_0\}) = \mu(\{a_1\}) + \mu(\{a_0\}) + \lambda \mu(\{a_1\}) \mu(\{a_0\}) \quad (5)$$

Where $1 \leq l, o \leq n$. Algorithm 2 provides a comprehensive explanation of the Sugeno integral method.

Algorithm 2: Sugeno integral:

```

procedure SugenoIntegral( $\mathcal{C}, \mathcal{BC}, P, F$ )
   $\lambda \leftarrow$  initialize by solving Equation ()
   $predictions \leftarrow$  initialize empty list of final predictions
  for class_index  $i \in \{1, 2, 3, \dots, C\}$  do
     $P_\pi \leftarrow$  Sorted array  $P$  in descending order
     $F_\pi \leftarrow$  Permutation of  $F$  corresponding to  $P_\pi$ 
     $f\_prev \leftarrow F_\pi[0]$ 
     $pred \leftarrow \min(P_\pi[0], F_\pi[0])$ 
    for  $n \in \{1, 2, 3, \dots, C\}$  do
       $f\_current \leftarrow f\_prev + F_\pi[n] + \lambda \times F_\pi[n] \times f\_prev$ 
       $pred \leftarrow \max(pred, \min(P_\pi[n], f\_current))$ 
       $f\_prev \leftarrow f\_current$ 
    end for
     $predictions[i] \leftarrow pred$ 
  end for
   $class(l) \leftarrow \arg \max(predictions[i])$ 
end procedure

```

Source: Authors, (2026).

IV. RESULTS AND DISCUSSIONS

This section gives a full review of our suggested CNN technique, using some common ways to judge it on the test data samples. In addition, a comparison was made between the proposed method and several popular deep learning techniques in order to demonstrate the relative performance of the proposed method in comparison to other techniques. The dataset under consideration was divided into two sections: a training set and a testing set. The training part has 80% of the data (1485 with_DR and 1444 No_DR), and the testing part has 20% of the data (372 with_DR and 361 No_DR). This distribution is illustrated in Figure 3.

IV.1 EXPERIMENTAL PROTOCOL

The experimental analysis for diabetic retinopathy classification was performed on the Kaggle platform, which provides a cloud-based environment equipped with high-performance GPUs and extended memory resources ideal for deep learning tasks. Experiments were implemented using Python 3.7 within Kaggle's integrated Jupyter Notebook interface, allowing seamless coding, visualisation and model training. The setup included essential libraries such as TensorFlow, PyTorch, NumPy and Matplotlib, supporting an efficient workflow for deep learning model development and evaluation.

IV.2 EVALUATION METRICS

For the classification of diabetic retinopathy, the selection of appropriate evaluation metrics is critical to accurately assess model performance. Commonly used metrics include recall, precision, accuracy, F1-score and true negative rate (specificity), each providing different insights into different aspects of the model's predictive ability. The mathematical formulations of these metrics are presented in Eqs. (6 -10), which form the basis for performance evaluation in this study.

Accuracy: is measured by the percentage of instances that the correct classification has been given out of the total number of cases.

$$\text{Accuracy (Acc)} = \frac{(TP + TN)}{(TP + TN + FP + FN)} * 100\% \quad (6)$$

Recall: This function calculates how well the classifier can find all the positive instances. This is also known as sensitivity.

$$\text{Recall (Sensitivity)} = \frac{TP}{(TP + FN)} * 100\% \quad (7)$$

True Negative Rate (TNR): is a metric that quantifies the precision of a system's negative identification, calculated as the ratio of true negative instances that are correctly identified.

$$\text{TNR (Specificity)} = \frac{TN}{(TN + FP)} * 100\% \quad (8)$$

Precision: The ability of the classifier to only return relevant instances is measured by precision:

$$\text{Precision (Pre)} = \frac{TP}{(TP + FP)} * 100\% \quad (9)$$

F1_score: The F1 score is the arithmetic average of precision and recall. False positives and false negatives are both considered.

$$F1_{\text{score}} = \frac{2 * \text{Precision} * \text{Recall}}{(\text{Precision} + \text{Recall})} * 100\% \quad (10)$$

IV.3 RESULTS OF CLASSIFICATION MODELS

In this section, we present a comparative analysis of the performance of five deep learning models trained on the original APTOS 2019 dataset. As summarised in Table 1, all models demonstrated high predictive performance, with accuracies converging around 98.23%. This consistency highlights the robustness of the dataset and the effectiveness of fine-tuned hyperparameters, such as the use of the Adam optimiser with a learning rate of 0.0001 and an epoch size of 50 for most models. Among the models evaluated, EfficientNet-B7 achieved the highest AUC of 99.78%, indicating an exceptional ability to discriminate between DR and non-DR cases. Notably, despite identical accuracy to other models, EfficientNet-B7 stood out for its superior recall (98.92%) and low misclassification rates, as confirmed by its confusion matrix [[352,9],[4,368]] [[352,9],[4,368]]. This suggests its effectiveness in minimising false negatives, which is critical in medical diagnosis. The ResNet50 model, trained with a significantly larger batch size of 128 and a higher learning rate of 0.001, also achieved excellent results, with balanced precision (97.87%) and recall (98.66%), and a strong AUC of 99.51%.

This suggests that larger batch sizes combined with Adam optimisation can still produce competitive results. DenseNet121, trained with a smaller batch size of 8, showed a favorable trade-off between false positives and false negatives, reflected in a confusion matrix of [[355.6],[7.365]] [[355.6],[7.365]]. The value of its AUC is 99.33%. Both VGG16 and VGG19 produced almost identical accuracies (98.23%), but differed slightly in precision and recall, with VGG16 achieving higher precision (98.38%) and VGG19 demonstrating better recall (98.92%). These differences highlight the subtle trade-offs between the models in correctly identifying positive and negative samples, as demonstrated by their confusion matrix. The ensemble of results indicate that all five models can detect diabetic retinopathy with high accuracy when optimized properly. The confusion matrices show consistently low false positive and false negative rates for all models, underscoring their potential clinical applicability. Figures (9 to 13) present the evaluation results of all models employed in the designed system using the dataset.

Table 1: Performance Metrics of Classification Models on the APTOS 2019 Dataset Including hyper-parameter settings and Confusion Matrices.

Models	Hyper-parameter				Evaluation Metrics				
	Batch size	Epochs	Optimizers	learning rate	Accuracy	Precision	Recall	F1-Score	AUC
VGG16	16	50	Adam	10^{-4}	0.9823	0.9838	0.9812	0.9825	0.9949
VGG19	16	50	Adam	10^{-4}	0.9823	0.9761	0.9892	0.9826	0.9931
EfficientNet-B7	8	50	AdamW	10^{-4}	0.9823	0.9761	0.9892	0.9826	0.9978
DenseNet-121	128	50	Adam	10^{-4}	0.9823	0.9838	0.9812	0.9825	0.9933
Res Net-50	128	50	Adam	10^{-3}	0.9823	0.9787	0.9866	0.9826	0.9951

Source: Authors, (2026).

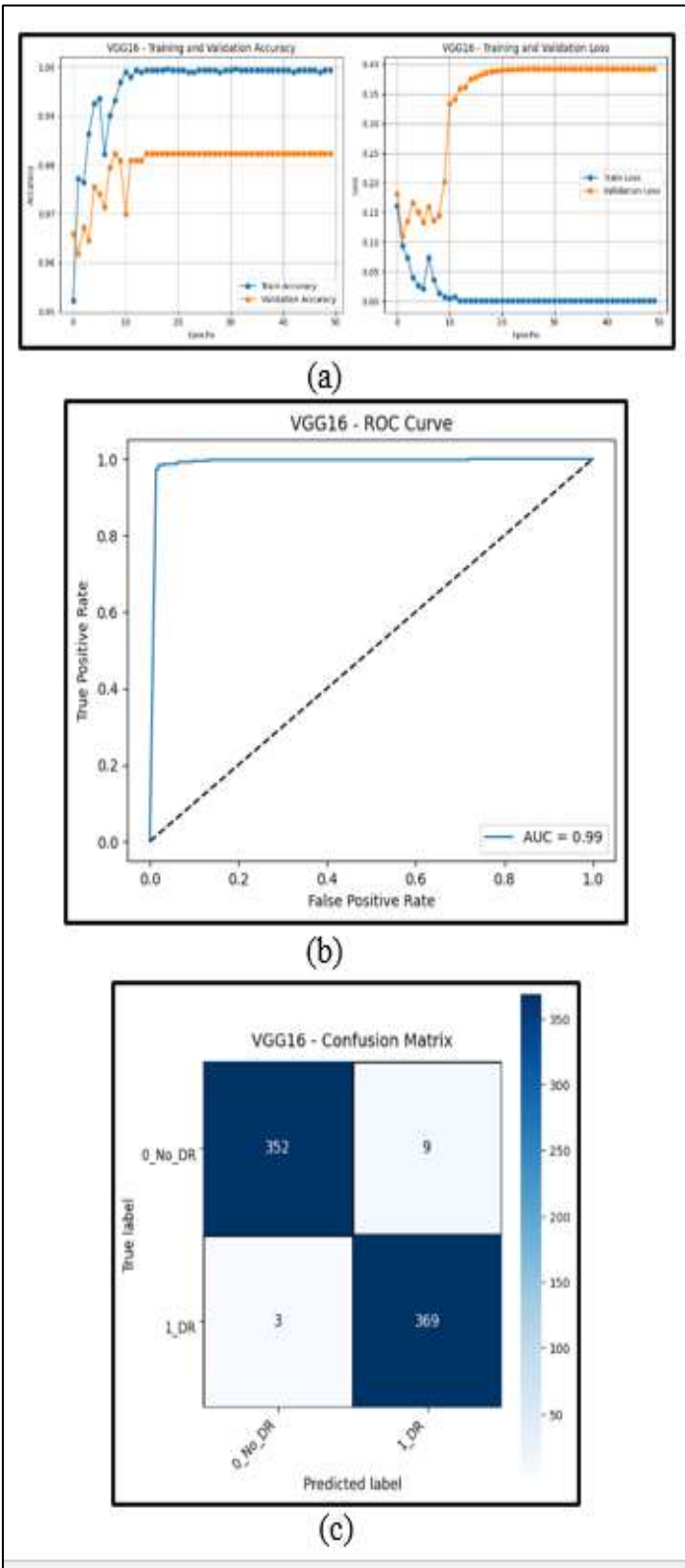


Figure 9: Evaluation of the VGG16 classifier: (a) training and validation accuracy and loss, (b) ROC curve analysis, and (c) confusion matrix evaluation.
Source: Authors, (2026).

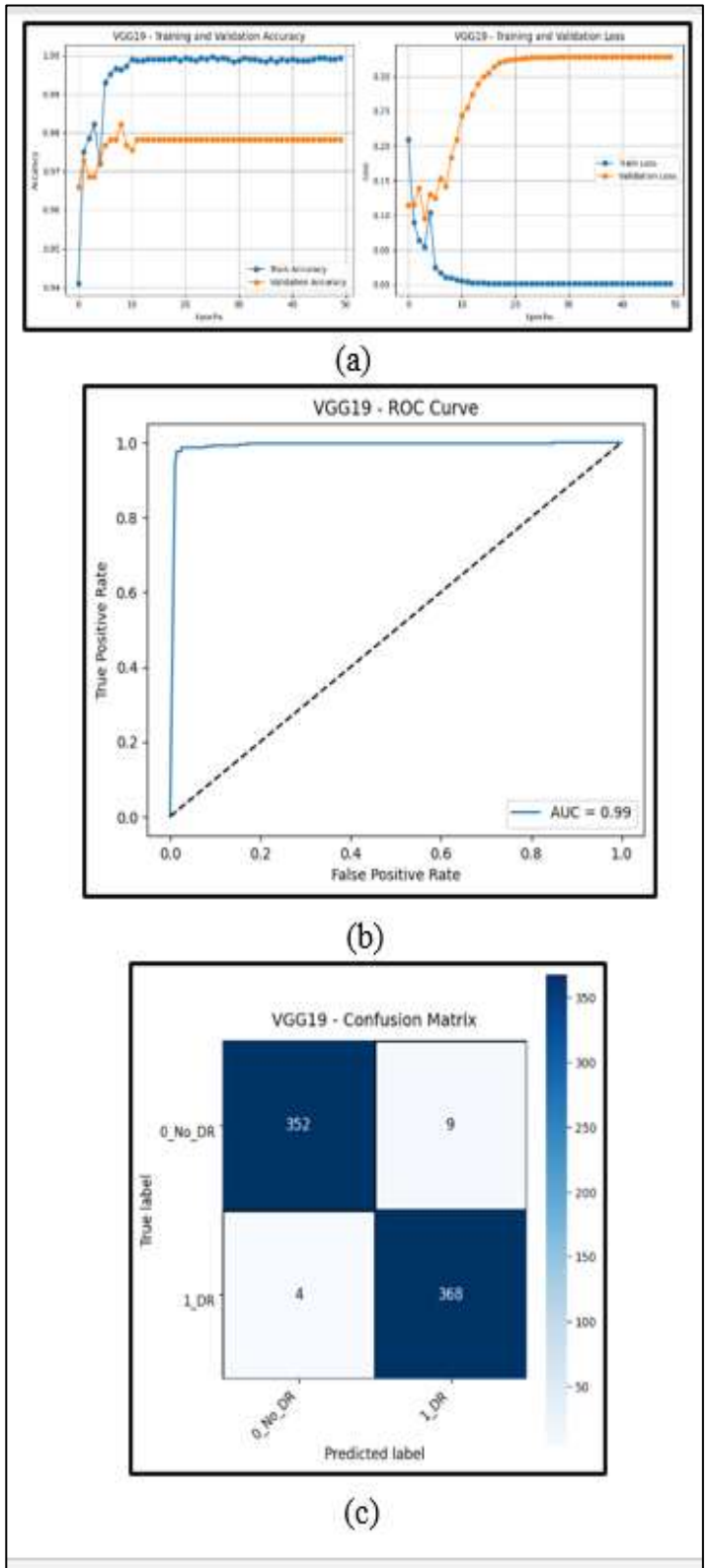
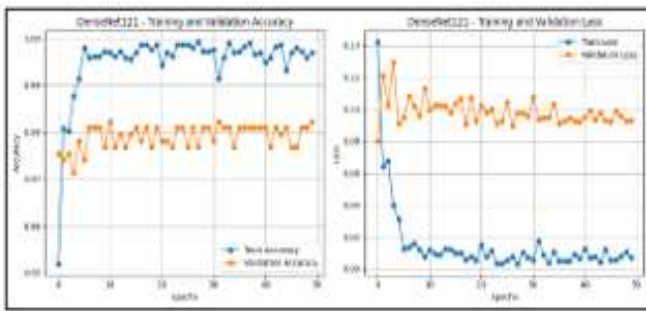
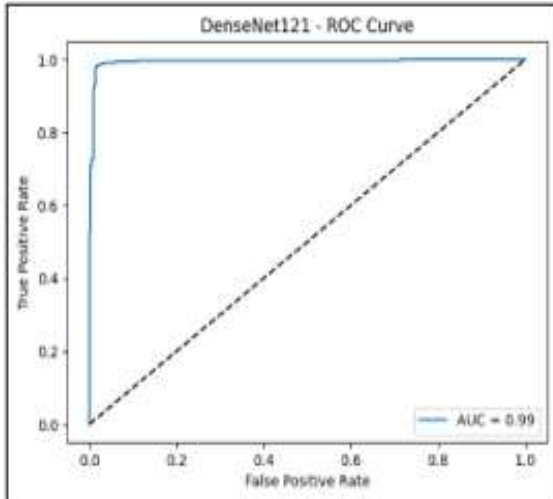


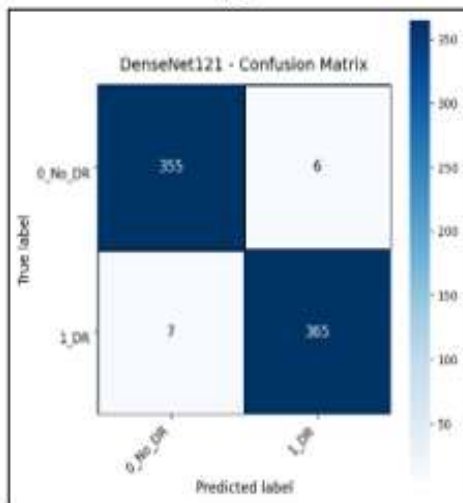
Figure 10: Evaluation of the VGG19 classifier: (a) training and validation accuracy and loss, (b) ROC curve analysis, and (c) confusion matrix evaluation.
Source: Authors, (2026).



(a)

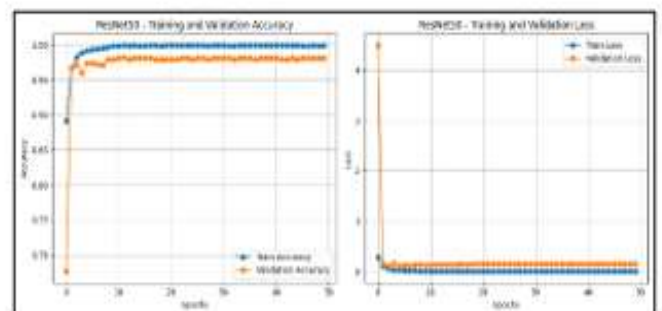


(b)

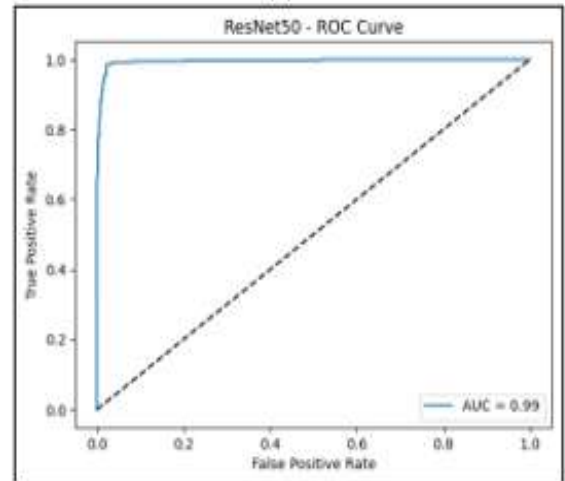


(c)

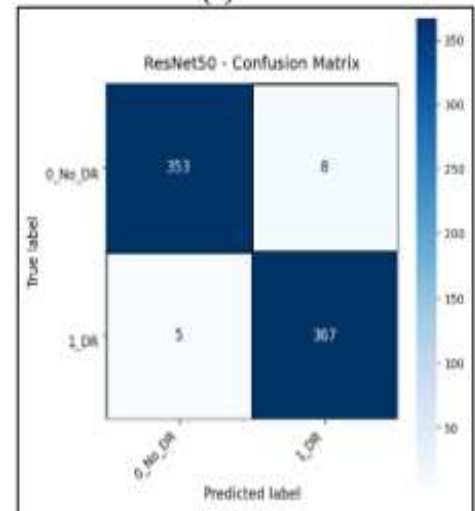
Figure 11: Evaluation of the DenseNet121 classifier: (a) training and validation accuracy and loss, (b) ROC curve analysis, and (c) confusion matrix evaluation.
Source: Authors, (2026).



(a)



(b)



(c)

Figure 12: Evaluation of the ResNet50 classifier: (a) training and validation accuracy and loss, (b) ROC curve analysis, and (c) confusion matrix evaluation.
Source: Authors, (2026).

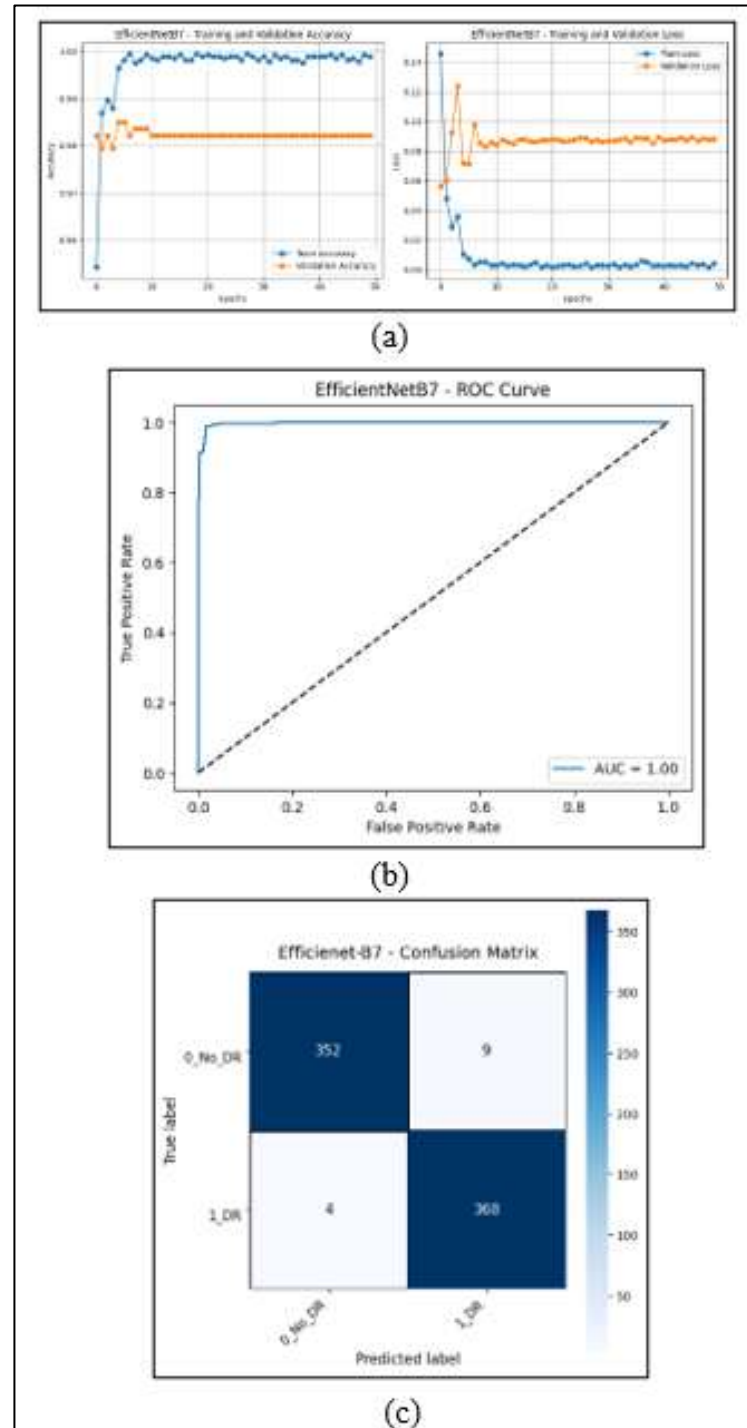


Figure 13: Evaluation of the EfficientNet-B7 classifier: (a) training and validation accuracy and loss, (b) ROC curve analysis, and (c) confusion matrix evaluation.

Source: Authors, (2026).

IV.4 MODELS RESULTS OF ENSEMBLE LEARNING

Once the best performing individual models (VGG16, VGG19, EfficientNet-B7, DenseNet121 and ResNet50) were identified based on classification metrics, ensemble learning was used to combine their strengths and further improve performance. Four ensemble strategies were evaluated: Majority Voting, Average, Fuzzy Rank, and Sugeno Integral, respectively applied to the top three, four, and five models ranked by accuracy. The detailed performance results of these fusion strategies are presented in Table 2, while the corresponding confusion matrices for each method are shown in Fig. 14, providing a visual insight into the classification results.

Top three classifiers (VGG16, VGG19, EfficientNet-B7)

All ensemble methods (Majority Voting, Average, Fuzzy Rank and Sugeno Integral) achieved the highest accuracy of 0.9864 with identical evaluation metrics: Precision (0.9814), Recall (0.9919), F1 score (0.9866), Confusion matrix: [[354, 7], [3, 369]]. This result confirms the robustness of the ensemble techniques, especially when combining the three best performing models. The confusion matrix shows minimal false positives and false negatives, indicating excellent generalization and clinical reliability.

Top four classifiers (adding DenseNet121)

Ensemble methods using four models showed a slight drop in accuracy to 0.9823 (Majority Voting and Average), while Fuzzy Rank maintained 0.9864: Precision (0.9838), Recall (0.9812), F1 score (0.9825), Confusion matrix: [[355, 6], [7, 365]]. These results show that the inclusion of an additional model (DenseNet121) slightly increased the false negative rate and reduced recall and F1-score. However, the fuzzy rank fusion proved to be more stable, maintaining the best scores with minimal loss of performance.

Top five classifiers (adding ResNet50)

When all five models were combined:

Accuracy: 0.9836 (Majority Voting, Average, Sugeno Integral)

Precision (0.9813), Recall (0.9866), F1 score (0.9839), Confusion matrix: [[354, 7], [5, 367]].

This configuration achieved a good balance between precision and recall, improving performance over the top-4 configuration. However, it did not surpass the performance of the top-3 combination, suggesting that adding more models does not improve the overall performance. The best performance (accuracy of 0.9864) was consistently achieved by all fusion strategies when using the top three classifiers. This highlights the importance of selecting high quality, complementary models for ensemble learning. The Fuzzy Rank and Sugeno Integral methods showed superior stability, achieving the highest scores even when more models were added. The confusion matrices show that most misclassifications occur in the negative class, with very few false negatives - a critical factor in medical diagnostic applications where missed cases are riskier.

V. CONCLUSIONS

In this study, an advanced deep learning-based system for automated detection of diabetic retinopathy (DR) was proposed using the APTOS 2019 dataset of RGB fundus images. Five pre-trained convolutional neural networks: VGG16, VGG19, EfficientNet-B7, DenseNet121 and ResNet50, were fine-tuned using transfer learning and individually evaluated for their classification performance. All models achieved high accuracy (around 98.23%), with EfficientNet-B7 outperforming the others in ROC AUC (99.78%) and recall (98.92%), indicating strong discriminative power and minimal false negatives - an essential feature in medical diagnostics. To further improve performance, ensemble-learning strategies including Majority Voting, Average, Fuzzy Rank and Sugeno Integral were applied to combinations of the best performing models. The best results were consistently achieved by fusing the top three models (VGG16, VGG19 and EfficientNet-B7), which gave an accuracy of 98.64%, precision of 98.14%, recall of 99.19% and F1 score of 98.66%.

These metrics were supported by confusion matrices showing very low false positives and negatives, reinforcing the clinical reliability of the system. While the addition of more models (up to five) resulted in slight performance trade-offs, particularly in recall, fusion strategies such as Fuzzy Rank and Sugeno Integral showed strong robustness and stability across configurations. The results underscore that careful selection and combination of complementary, high-performing models yields superior results compared to standalone classifiers or excessive model aggregation. Overall, this work validates the effectiveness of integrating transfer learning with ensemble learning for DR detection and provides a scalable, accurate and clinically applicable solution for early detection, potentially improving patient outcomes and supporting efficient healthcare delivery.

Table 2: Evaluation results of the ensemble method based on score level fusion of all models on the APTOS 2019 dataset.

Order of Models	Methods	Models	Evaluation Metrics			
			Accuracy	Precision	Recall	F1-Score
1- VGG16	Majority voting	Top 3 Models	0.9864	0.9814	0.9919	0.9866
		Top 4 Models	0.9823	0.9838	0.9812	0.9825
		Top 5 Models	0.9836	0.9813	0.9866	0.9839
2- VGG19	Average	Top 3 Models	0.9864	0.9814	0.9919	0.9866
		Top 4 Models	0.9823	0.9838	0.9812	0.9825
		Top 5 Models	0.9836	0.9813	0.9866	0.9839
3- EfficientNet-B7	Fuzzy rank	Top 3 Models	0.9864	0.9814	0.9919	0.9866
		Top 4 Models	0.9864	0.9814	0.9919	0.9866
		Top 5 Models	0.9864	0.9814	0.9919	0.9866
4- DenseNet121	Sugeno integral	Top 3 Models	0.9864	0.9814	0.9919	0.9866
		Top 4 Models	0.9836	0.9813	0.9866	0.9839
		Top 5 Models	0.9836	0.9813	0.9866	0.9839

Source: Authors, (2026).

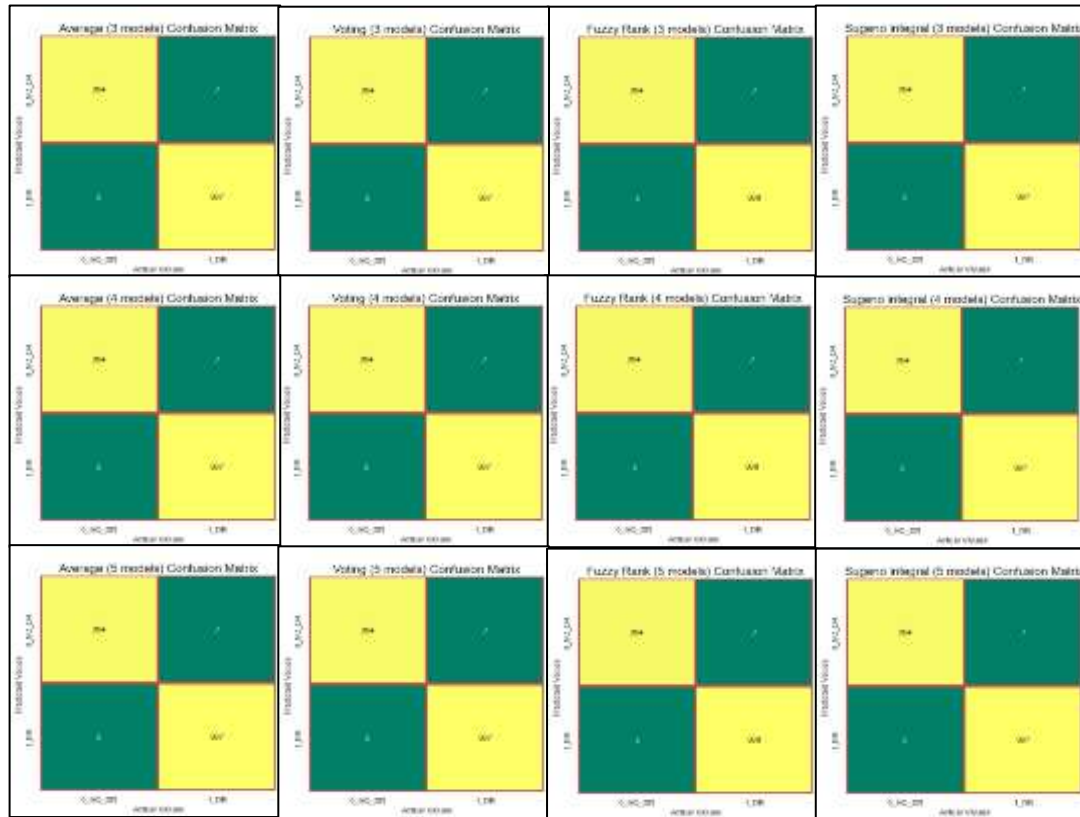


Figure 14: Confusion matrices of all ensemble learning algorithms using Top 5 classifiers trained on the APTOS 2019 dataset (RGB images).

Source: Authors, (2026).

Table 3: Summary of Literature Review and Performance Comparison with the Proposed Model.

References	Datasets	Models/Technique	Metrics	Results / Performance
[30]	EyePACS, APTOS	Extreme learningmachine	Accuracy	91.78% (EyePACS), 97.27% (APTOS)
[50]	APTOS19, DIARETDB1	Ensemble model-DenseNet101, ResNext	Accuracy	96.98% (binary), 86.08% (multi)
[31]	APTOS, MESSIDOR	CNN and SVM	Accuracy	97.25% (APTOS), 94.8% (MESSIDOR)
[29]	APTOS	AlexNet, ResNet-101with ACS	Accuracy	93%
[51]	EyePACS, APTOS	Source-Free TransferLearning	Accuracy	91.2%
[52]	EyePACS (training), APTOS (testing)	ResNet, EfficientNet,andSwin Transformer	Accuracy	76.35%
[42]	APTOS	ResNet, Inception V3, InceptionResNet-V2, DenseNet-169, Xception, EfficientNet-B4	quadratic weighted kappa	0.824
[53]	APTOS	AlexNet, Res-Net18,SqueezeNet, GoogleNet,VGG16, and VGG19	Accuracy	97.9%
[54]	Kaggle data	CNN with grid search	Accuracy	89%
[55]	EyePACS	Inception-V3	Accuracy	90.9%
Our	APTOS	Ensemble learning methods	Accuracy	Acc = 98.64%, recall = 99.19%

Source: Authors, (2026).

VI. AUTHOR'S CONTRIBUTION

Conceptualization: Ishaq Aiche, Youcef Brik, Bilal Attallah and Oussama Bougeurra.

Methodology: Ishaq Aiche, Youcef Brik, Bilal Attallah and Oussama Bougeurra.

Investigation: Ishaq Aiche, Youcef Brik, Bilal Attallah and Oussama Bougeurra.

Discussion of results: Ishaq Aiche, Youcef Brik, Bilal Attallah and Oussama Bougeurra.

Writing – Original Draft: Ishaq Aiche.

Writing – Review and Editing: Ishaq Aiche, Youcef Brik and Bilal Attallah.

Resources: Ishaq Aiche, Youcef Brik, Bilal Attallah and Oussama Bougeurra.

Supervision: Youcef Brik, Bilal Attallah.

Approval of the final text: Ishaq Aiche, Youcef Brik, Bilal Attallah and Oussama Bougeurra.

VII. REFERENCES

- [1] M. Kropp, O. Golubnitschaja, A. Mazurakova, L. Koklesova, N. Sargheini, T. T. K. S. Vo, et al., “Diabetic retinopathy as the leading cause of blindness and early predictor of cascading complications risks and mitigation,” *EPMA Journal*, vol. 14, no. 1, pp. 21-42, 2023.
- [2] G. Latif, J. Alghazo, M. A. Khan, G. B. Brahim, K. Fawagreh, and N. Mohammad, “Deep convolutional neural network (CNN) model optimization techniques—Review for medical imaging,” *AIMS Mathematics*, vol. 9, no. 8, pp. 20539-20571, 2024.
- [3] K. A. Alavee, M. Hasan, A. H. Zillanee, M. Mostakim, J. Uddin, E. S. Alvarado, et al., “Enhancing early detection of diabetic retinopathy through the integration of deep learning models and explainable artificial intelligence,” *IEEE Access*, vol. 12, pp. 1-12, 2024.
- [4] C. Nadir, B. Attallah, and Y. Brik, “A sequential combination of convolution neural network and machine learning for finger vein recognition system,” *Signal Image Video Process.*, vol. 18, no. 11, pp. 8267-8278, 2024.
- [5] G. Rajarajeshwari and G. C. Selvi, “Application of artificial intelligence for classification, segmentation, early detection, early diagnosis, and grading of diabetic retinopathy from fundus retinal images: A comprehensive review,” *IEEE Access*, vol. 12, pp. 1-15, 2024.
- [6] F. Mohsen, U. Shah, A. Islam, Z. Shah, and S. B. Belhaouari, “Multimodal deep learning for diabetic retinopathy grading: Integrating linear-Radon sinograms and retinal fundus images,” in *Proc. 2024 IEEE 26th Int. Workshop Multimedia Signal Process. (MMSP)*, West Lafayette, IN, USA, 2024, pp. 1-5.
- [7] A. Mumuni and F. Mumuni, “Data augmentation: A comprehensive survey of modern approaches,” *Array*, vol. 16, p. 100258, 2022.
- [8] N. Asiri, M. Hussain, F. Al Adel, and N. Alzaiddi, “Deep learning based computer-aided diagnosis systems for diabetic retinopathy: A survey,” *Artif. Intell. Med.*, vol. 99, p. 101701, 2019.
- [9] S. Qummar, F. G. Khan, S. Shah, A. Khan, S. Shamshirband, Z. U. Rehman, et al., “A deep learning ensemble approach for diabetic retinopathy detection,” *IEEE Access*, vol. 7, pp. 150530-150539, 2019.
- [10] A. Sebastian, O. Elharrouss, S. Al-Maadeed, and N. Almaadeed, “A survey on deep-learning-based diabetic retinopathy classification,” *Diagnostics*, vol. 13, no. 3, p. 345, 2023.
- [11] D. Galatro, M. Machavolu, and G. Navas, “Transfer learning strategies for neural networks: A case study in amine gas treating units,” *Results Eng.*, vol. 24, p. 103027, 2024.
- [12] L. Torrey and J. Shavlik, “Transfer learning,” in *Handbook of Research on Machine Learning Applications and Trends: Algorithms, Methods, and Techniques*, IGI Global, 2010, pp. 242-264.
- [13] S. V. Chilukoti, L. Shan, A. S. Maida, and X. Hei, “A reliable diabetic retinopathy grading via transfer learning with quadratic weighted kappa metric,” *Res. Square*, 2023.
- [14] R. Vij and S. Arora, “A novel deep transfer learning based computerized diagnostic systems for multi-class imbalanced diabetic retinopathy severity classification,” *Multimed. Tools Appl.*, vol. 82, no. 22, pp. 34847-34884, 2023.
- [15] D. U. N. Qomariah, H. Tjandrasa, and C. Fatichah, “Classification of diabetic retinopathy and normal retinal images using CNN and SVM,” in *Proc. 2019 12th Int. Conf. Inf. Commun. Technol. Syst. (ICTS)*, Surabaya, Indonesia, 2019, pp. 152-157.
- [16] D. Rahhal, R. Alhamouri, I. Albataineh, and R. Duwairi, “Detection and classification of diabetic retinopathy using artificial intelligence algorithms,” in *Proc. 2022 13th Int. Conf. Inf. Commun. Syst. (ICICS)*, Valencia, Spain, 2022, pp. 15-21.
- [17] S. Ioffe and C. Szegedy, “Batch normalization: Accelerating deep network training by reducing internal covariate shift,” *arXiv preprint*, 2015.
- [18] K. Simonyan and A. Zisserman, “Very deep convolutional networks for large-scale image recognition,” *arXiv preprint*, 2014.
- [19] G. Huang, Z. Liu, L. van der Maaten, and K. Q. Weinberger, “Densely connected convolutional networks,” in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2017, pp. 4700-4708.
- [20] A. G. Howard et al., “MobileNets: Efficient convolutional neural networks for mobile vision applications,” *arXiv preprint*, 2017.
- [21] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2016, pp. 770-778.
- [22] D. Doshi, A. Shenoy, D. Sidhpura, and P. Gharpure, “Diabetic retinopathy detection using deep convolutional neural networks,” in *Proc. 2016 Int. Conf. Comput. Analytics and Secur. Trends (CAST)*, Pune, India, 2016, pp. 261-266.
- [23] K. S. Kothare and K. Malpe, “Design and implementation of inspection model for knowledge patterns classification in diabetic retinal images,” in *Proc. 2019 3rd Int. Conf. Comput. Methodologies and Commun. (ICCMC)*, Erode, India, 2019, pp. 1220-1223.
- [24] M. A. Hussain, S. O. B. Islam, M. I. Tiwana, and W. S. Qureshi, “Detection and classification of hard exudates with fundus images complements and neural networks,” in *Proc. 2019 5th Int. Conf. Control, Autom Robot (ICCAR)*, Beijing, China, 2019, pp. 206-211.
- [25] M. Ahmad, N. Kasukurthi, and H. Pande, “Deep learning for weak supervision of diabetic retinopathy abnormalities,” in *Proc. 2019 IEEE 16th Int. Symp. Biomed. Imaging (ISBI 2019)*, Venice, Italy, 2019, pp. 573-577.
- [26] A. Bilal, G. Sun, Y. Li, S. Mazhar, and A. Q. Khan, “Diabetic retinopathy detection and classification using mixed models for a disease grading database,” *IEEE Access*, vol. 9, pp. 23544-23553, 2021.
- [27] D. A. Da Rocha, F. M. F. Ferreira, and Z. M. A. Peixoto, “Diabetic retinopathy classification using VGG16 neural network,” *Res. Biomed. Eng.*, vol. 38, no. 2, pp. 761-772, 2022.

- [28] M. W. Nadeem, H. G. Goh, M. Hussain, S. Y. Liew, D. Andresen, P. Liatsis, et al., "Deep learning for diabetic retinopathy analysis: a review, research challenges, and future directions," *Sensors*, vol. 22, no. 18, p. 6780, 2022.
- [29] A. M. Fayyaz, M. I. Sharif, S. Azam, A. Karim, and J. El-Den, "Analysis of diabetic retinopathy (DR) based on deep learning," *Information*, vol. 14, no. 1, p. 30, 2023.
- [30] M. Nahiduzzaman, M. R. Islam, M. O. F. Goni, M. S. Anower, M. Ahsan, J. Haider, et al., "Diabetic retinopathy identification using parallel convolutional neural network based feature extractor and ELM classifier," *Expert Syst. Appl.*, vol. 217, p. 119557, 2023.
- [31] P. K. Jena, B. Khuntia, C. Palai, M. Nayak, T. K. Mishra, and S. N. Mohanty, "A novel approach for diabetic retinopathy screening using asymmetric deep learning features," *Big Data Cogn. Comput.*, vol. 7, no. 1, p. 25, 2023.
- [32] D. A. Van Dyk and X. L. Meng, "The art of data augmentation," *J. Comput. Graph. Stat.*, vol. 10, no. 1, pp. 1–50, 2001.
- [33] C. Shorten and T. M. Khoshgoftaar, "A survey on image data augmentation for deep learning," *J. Big Data*, vol. 6, no. 1, pp. 1–48, 2019.
- [34] Z. Mungloo-Dilmohamud, M. Heenaye-Mamode Khan, K. Jhumka, B. Beedassy, N. Mungloo, and S. Panchoo, "Balancing data through data augmentation improves the generality of transfer learning for diabetic retinopathy classification," *Appl. Sci.*, vol. 12, no. 11, p. 5363, 2022.
- [35] A. M. Mutawa, S. Alnajdi, and S. Sruthi, "Transfer learning for diabetic retinopathy detection: a study of dataset combination and model performance," *Appl. Sci.*, vol. 13, no. 9, p. 5685, 2023.
- [36] S. Guefrachi, A. Echtioui, and H. Hamam, "Diabetic retinopathy detection using deep learning multistage training method," *Arab J. Sci. Eng.*, vol. 50, no. 2, pp. 1079–1096, 2024.
- [37] Z. Zhao, L. Alzubaidi, J. Zhang, Y. Duan, and Y. Gu, "A comparison review of transfer learning and self-supervised learning: Definitions, applications, advantages and limitations," *Expert Syst. Appl.*, vol. 242, p. 122807, 2024.
- [38] M. Elsharkawy, A. Sharafeldeen, A. Soliman, F. Khalifa, M. Ghazal, F. Taher, et al., "The role of different retinal imaging modalities in predicting progression of diabetic retinopathy: A survey," *Sensors*, vol. 22, no. 9, p. 3490, 2022.
- [39] F. Saeed, M. Hussain, and H. A. Aboalsamh, "Automatic diabetic retinopathy diagnosis using adaptive fine-tuned convolutional neural network," *IEEE Access*, vol. 9, pp. 41344–41359, 2021.
- [40] I. Kandel and M. Castelli, "Transfer learning with convolutional neural networks for diabetic retinopathy image classification: A review," *Appl. Sci.*, vol. 10, no. 6, p. 2021, 2020.
- [41] N. N. Prakash, V. Rajesh, D. L. Namakhwa, S. D. Pande, and S. H. Ahammad, "A DenseNet CNN-based liver lesion prediction and classification for future medical diagnosis," *Sci. Afr.*, vol. 20, p. e01629, 2023.
- [42] M. Al-Smadi, M. Hammad, Q. B. Baker, and A. Sa'ad, "A transfer learning with deep neural network approach for diabetic retinopathy classification," *Int. J. Electr. Comput. Eng.*, vol. 11, no. 4, pp. 3492–3500, 2021.
- [43] T. Singh, "Architecture of DenseNet121," *OpenGenus IQ*, 2021. [Online]. Available: <https://iq.opengenus.org/architecture-of-densenet121/>.
- [44] A. K. Sharma, A. Nandal, A. Dhaka, L. Zhou, A. Alhudhaif, F. Alenezi, and K. Polat, "Brain tumor classification using the modified ResNet50 model based on transfer learning," *Biomed. Signal Process. Control*, vol. 86, p. 105299, 2023.
- [45] M. Tan and Q. Le, "EfficientNet: Rethinking model scaling for convolutional neural networks," in *Proc. 36th Int. Conf. Mach. Learn. (ICML)*, Long Beach, CA, USA, 2019, pp. 6105–6114.
- [46] M. Sharafudeen and S. V. C. S., "Detecting skin lesions fusing handcrafted features in image network ensembles," *Multimed. Tools Appl.*, vol. 82, no. 2, pp. 3155–3175, 2023.
- [47] J. M. Tanskanen, X. Z. Gao, J. Wang, P. Guo, J. A. Hyttinen, and V. S. Dimitrov, "Experimental comparison of geometric, arithmetic and harmonic means for EEG event related potential detection," in *Proc. 8th Int. Conf. Comput. Intell. Secur.*, 2012, pp. 112–116.
- [48] A. Manna, R. Kundu, D. Kaplun, A. Sinitca, and R. Sarkar, "A fuzzy rank-based ensemble of CNN models for classification of cervical cytology," *Sci. Rep.*, vol. 11, no. 1, p. 14538, 2021.
- [49] S. Dey, R. Roychoudhury, S. Malakar, and R. Sarkar, "An optimized fuzzy ensemble of convolutional neural networks for detecting tuberculosis from chest X-ray images," *Appl. Soft Comput.*, vol. 114, p. 108094, 2022.
- [50] S. S. Mondal, N. Mandal, K. K. Singh, A. Singh, and I. Izonin, "Edldr: An ensemble deep learning technique for detection and classification of diabetic retinopathy," *Diagnostics*, vol. 13, no. 1, p. 124, 2022.
- [51] C. Zhang, T. Lei, and P. Chen, "Diabetic retinopathy grading by a source-free transfer learning approach," *Biomed. Signal Process. Control*, vol. 73, p. 103423, 2022.
- [52] S. Xu, Z. Huang, and Y. Zhang, "Diabetic retinopathy progression recognition using deep learning method," *arXiv preprint*, 2020. [Online]. Available: [arXiv:2009.11544](https://arxiv.org/abs/2009.11544).
- [53] N. E. M. Khalifa, M. Loey, M. H. N. Taha, and H. N. E. T. Mohamed, "Deep transfer learning models for medical diabetic retinopathy detection," *Acta Inform. Med.*, vol. 27, no. 5, pp. 327–331, 2019.
- [54] S. Rajamani and S. Sasikala, "Artificial intelligence approach for diabetic retinopathy severity detection," *Informatica*, vol. 46, no. 8, 2023.
- [55] M. T. Hagos and S. Kant, "Transfer learning based detection of diabetic retinopathy from small dataset," *arXiv preprint*, 2019. [Online]. Available: [arXiv:1905.07203](https://arxiv.org/abs/1905.07203).