Journal of Information Hiding and Multimedia Signal Processing ©2025 ISSN 2073-4212
Ubiquitous International Volume **16**, Number **1**, March **2025**

Diabetic Retinopathy Detection through Ensemble Averaging Method

Nishu Chowdhury
Department of Computer Science and Engineering,
Southern University Bangladesh
Arefin nagar, Chattogram, Bangladesh
nishuchy11@gmail.com

Samiya Chowdhury
Department of Computer Science and Engineering,
Southern University Bangladesh
Arefin nagar, Chattogram, Bangladesh
nishuchy11@gmail.com

Received November 16, 2024, revised February 13, 2025, accepted February 18, 2025.

Abstract. *An ensemble learning model that successfully integrates many deep learning architectures is used to propose an unique method for the identification of diabetic retinopathy. By combining two different datasets that covered different stages of diabetic retinopathy, a dataset with less bias was suggested. Because of the more thorough training procedure made possible by this improved dataset, the model was able to attain greater classification accuracy over a range of severity levels. A thorough evaluation of the ensemble learning approach's efficacy shows that, in comparison to individual models, performance measures significantly improve. With 98.7% accuracy, 97.7% precision, 98.7% recall, and a 98.2% F1-score, the ensemble model produced remarkable results. Several convolutional neural networks were used in this ensemble approach to enhance the model's ability to identify diabetic retinopathy, demonstrating.*
Keywords: Diabetic Retinopathy, Ensemble Learning, Pre-trained Transfer Learning, Ensemble Average Method.

---

1. **Introduction.** A prominent complication of diabetes and a major cause of vision loss worldwide, diabetic retinopathy frequently goes undetected until it is advanced, underscoring the importance of efficient early identification. Recent developments in computer vision and machine learning have made it possible to use retinal fundus images for automated diagnosis and identification of this problem. This study uses convolutional neural networks (CNNs), a type of deep learning, to create a comprehensive diabetic retinopathy detection system. It uses transfer learning to extract detailed features from images using architectures such as InceptionV3, ResNet50V2, and DenseNet201, allowing for the classification of the severity levels of the condition. Ensemble learning also improves classification accuracy and robustness by combining predictions from several models.

Fundus disease is a significant contributor to blindness and visual loss [1]. Age-related macular degeneration (AMD), cataracts, and diabetic retinopathy (DR) all have a major effect on visual function. There are currently no recognized therapies for these conditions, and vision deteriorates as they worsen. Diabetes affects a sizable portion of the global population, and DR is a frequent side effect. Although it may not initially create any symptoms, DR is one of the top four causes of blindness and can cause blindness [2]. DR is a result of chronic diabetes mellitus and affects the retinal blood vessels [3, 4]. This disease, which causes over 400,000 deaths annually, is frequently diagnosed and treated with fundus images [5].

Due to population aging, DR is now the most common cause of blindness in many nations, and its prevalence rises with age. Vision can quickly decrease without prompt and efficient care, resulting in irreversible disability [6, 7]. Thus, it is essential to diagnose and treat fundus problems as soon as possible.

Using vast amounts of medical data, artificial intelligence (AI) technology can help primary care physicians diagnose and treat eye conditions in primary care settings. It is anticipated that combining AI with ophthalmic treatment would meet the needs of a sizable patient group with fundus problems [8].

Because deep learning (DL) can extract characteristics from training data, it has become popular in many study areas. Through transfer learning (TL), high-performing deep convolutional neural networks (CNNs), which have shown good results in image classification, can be modified for other classification tasks. CNNs have demonstrated their efficacy in machine vision and image classification by independently extracting sophisticated features from images.

To capitalize on the complementing qualities of each unique architecture, we used an ensemble approach. Because each deep learning model has a distinct edge, we may incorporate these varied capabilities by integrating them into an ensemble, which improves overall performance. In general, ensemble approaches increase robustness to changes in input data and reduce the danger of overfitting, which improves classification accuracy. When it came to determining the severity of diabetic retinopathy, our ensemble model outperformed any single model on its own thanks to the integration of the outputs of InceptionV3, ResNet50V2, and DenseNet201. This method efficiently handles the complexity and unpredictability related to retinal pictures while simultaneously increasing diagnostic accuracy.

The main contribution of this work is to improve the accuracy of the system by modifying the dataset so that it can work perfectly in real-world. We extracted comprehensive characteristics from retinal images via transfer learning using state-of-the-art deep learning architectures like InceptionV3, ResNet50V2, and DenseNet201. By using these characteristics, images were categorized into various diabetic retinopathy severity levels, allowing for more individualized treatment plans and better patient results.

2. **Problem Statement and Preliminaries.** Computer-aided diagnostic (CAD) systems were the main method used to identify fundus problems prior to the development of deep learning. These technologies, which analyzed fundus images for early illness diagnosis and screening, closely resembled the diagnostic procedure used by ophthalmologists. CAD systems were successful in detecting anomalies in fundus images, much as manual screening techniques [9]. At intermediate phases, they carried out vital tasks such image improvement [10], [11], and restoration [12, 13], after which they segmented lesions to extract important information. In the end, these systems reduced the intricate diagnostic procedure to tasks involving classification or grouping, which conventional machine learning algorithms could effectively handle. Manual screening by ophthalmologists has long been the gold standard for DR detection. However, this process is labor-intensive, time-consuming, and subject to inter-observer variability, which can lead to inconsistencies in diagnoses[14]. Additionally, reliance on manual screening presents significant challenges in regions where access to eye care professionals and resources is limited [15].

To overcome these limitations, researchers have increasingly focused on developing automated systems utilizing machine learning and computer vision techniques. Among these advancements, deep learning—a subset of machine learning—has shown great promise for automating the detection of diabetic retinopathy (DR) from retinal images. Convolutional Neural Networks (CNNs), in particular, have demonstrated exceptional performance in image classification tasks, including DR detection[16]. Recent research has shown that deep learning models trained on extensive datasets of retinal images can achieve levels of accuracy comparable to or even exceeding that of human experts [17]. For instance, Gulshan et al. developed a deep learning algorithm for detecting diabetic retinopathy in retinal fundus photographs, achieving high sensitivity and specificity [16]. Similarly, Ting et al. designed a deep learning system capable of detecting DR and other related eye diseases from retinal images of multiethnic populations with diabetes [17]. Abramoff et al. improved upon these findings by integrating deep learning for automated diabetic retinopathy detection using publicly available datasets [18]. Another key study led by Abràmoff et al. evaluated an autonomous AI-based diagnostic system for DR detection in primary care offices, showcasing its potential in real-world applications [19]. Additionally, Niemeijer et al. assessed the cost-effectiveness of automated diabetic retinopathy image assessment software, providing insight into the financial implications of adopting such systems [20]. Treatment outcomes for diabetic patients in a fundus photograph-based screening program [21-23]. Tan et al. added to the field by developing methods for automatic detection of microaneurysms in digital color fundus photographs, laying the groundwork for automated DR screening [24].

Recently, the integration of deep learning in computer vision has yielded promising results in automating diabetic retinopathy detection using retinal images. CNNs have become a robust tool for image analysis and classification tasks, especially in DR detection [25]. Various CNN architectures, such as DenseNet, ResNet, Inception, and EfficientNet, have been explored to enhance performance in this domain [25].Furthermore, transfer learning, which utilizes pre-trained models from large-scale image recognition tasks, has been applied to DR detection, improving both accuracy and efficiency

Several studies have validated the effectiveness of CNN-based approaches for DR detection. For example, Radwan et al. (2024) achieved an impressive accuracy using a deep residual network [26]. Results show that the suggested strategy outperforms cutting-edge techniques, with an accuracy of 99.36%, which is suggestive of improved diagnostic performance and dependability in differentiating DR severity levels. Consequently, this helps to reduce human mistakes and cutting expenses. Although modern methods have shown to be very effective, Their efficacy in identifying diabetic retinopathy (DR) mainly depends on feature extraction, which can be difficult to broaden [25]. To solve these issues, deep learning (DL) algorithms allow fundus images to be automatically. In 2024, Radwan developed a CNN model that achieved 98.5% accuracy in detecting COVID-19 from chest X-rays (CXR). The dataset contained 2541 images, split into training (75%) and test (25%) sets, with data augmentation used to increase the dataset's variety. Validated on Kaggle and GitHub datasets, the model showed improved efficiency with fewer parameters. However, the model's performance is affected by data quality, indicating that further testing across diverse datasets is necessary [26]. Similarly, Radwan et al. (2024) developed an ensemble model for predicting coronary artery disease (CAD) with 98.9% accuracy, utilizing the Z-Alizadeh Sani dataset and SMOTE-ENN for class balancing. The study also proposed potential applications for real-time prediction and highlighted the importance of feature selection and dataset balancing to enhance performance [27].

In 2025, using data fusion and transfer learning, Aftab and Akhtar presented an ensemble classification system for identifying diabetic retinopathy. Their model achieved 96.96% test accuracy on a dataset of 5922 fundus images, created by merging three benchmark datasets. Pre-processing techniques such as CLAHE, SMOTE, and data augmentation improved the dataset's quality and diversity. The ensemble model, averaging predictions from EfficientNetB2, DenseNet121, and ResNet50, outperformed individual models. However, they suggested exploring additional ensemble techniques, such as bagging and boosting, and further dataset fusion to enhance performance [27].

Building on previous research in diabetic retinopathy detection, Abood et al. (2025) developed a multi-label detection system using deep learning models such as VGG-19, DenseNet-121, and EfficientNet-B6. The system was tested on three benchmark datasets: APTOS-2019, IDRiD, and Messidor-2. Pre-processing techniques, including Gaussian Blur and weighted masking, were applied to improve performance. The results, evaluated using sensitivity, precision, F1-score, and accuracy, revealed that EfficientNet-B6 outperformed the other models. The proposed system showed high accuracy in detecting diabetic retinopathy severity, which can aid in early diagnosis. However, the study emphasized the need for lighter models to reduce computational demands for broader implementation [28]

Guefrachi et al. (2024) proposed a multistage training approach using various CNN architectures, including InceptionResNetV2, VGG16, VGG19, DenseNet121, MobileNetV2, and EfficientNet2L, to identify and classify diabetic retinopathy. To reduce overfitting and enhance model robustness, they implemented data augmentation techniques. The training process consisted of two steps: feature extraction followed by fine-tuning through the unfreezing of specific layers. The refined InceptionResNetV2 model achieved an accuracy of 96.61% when tested on Kaggle's diabetic retinopathy dataset, demonstrating the effectiveness of their approach [29]. Similarly, Hussain et al. (2025) developed an enhanced CNN-based technique called P-EDR to detect diabetic retinopathy in both non-proliferative (NPDR) and proliferative (PDR) stages. They preprocessed the high-resolution retinal image dataset using scaling, augmentation, and normalization to enhance picture quality and feature extraction. The P-EDR model was evaluated based on accuracy, sensitivity, specificity, and AUC-ROC, achieving 93% accuracy, 92% sensitivity, 94% specificity, and an AUC-ROC score of 0.97. The P-EDR model outperformed traditional machine learning models like SVM, RF, PNN, and GBM, highlighting its potential for early and precise DR diagnosis. Future work suggested exploring transfer learning and additional benchmark datasets to further enhance performance [30].

Using transfer learning and pre-trained model weights, Chilukoti et al. (2024) introduced computationally efficient ensemble models for the categorisation of diabetic retinopathy (DR). They used Gaussian blur to reduce noise and CLAHE to improve the picture. ReLU activation and dropout in a three-layer classifier reduced overfitting while DR grading feature extraction was taking place. Advanced Quadratic Weighted Kappa (QWK) values of 0.901, 0.967, and 0.944 were attained by the model on the Eyepacs, Aptos, and Messidor datasets, respectively. Despite excellent performance, the study identified shortcomings in addressing the ordered structure of DR grading and identifying all DR phases. Real-time applications require further advancements in processing efficiency and addressing label inconsistencies.[31].

Desiani et al. (2024) combined the ResNet-50, MobileNet, and EfficientNet architectures using weighted voting in an ensemble learning approach to enhance diabetic retinopathy (DR) classification performance. The APTOS and EyePACS datasets were utilised in the study for both testing and training. The ensemble learning technique outperformed with 93.3% accuracy, 93.42% F1-score, and 0.866 Cohen's Kappa. EfficientNet had the maximum sensitivity at 96.2%, while ResNet-50 had the highest specificity at 99.78%. In contrast to single model classification, the study found that ensemble learning considerably improved accuracy by 28.37% and addressed overfitting. Nevertheless, the ResNet-50 model required more refinement.[32]

3. **Methodology.** In this study, an ensemble learning approach was employed to improve the detection of diabetic retinopathy using a combination of pre-trained convolutional neural networks (CNNs). Diabetic retinopathy is a leading cause of vision loss, and its detection is challenging due to varying image quality, subtle differences between severity stages, and an imbalanced distribution of cases. Accurate detection and classification of the disease's stages are critical for early intervention.

One of the primary challenges in this task is the class imbalance found in most diabetic retinopathy datasets, where the majority of images represent the early or no-disease stages, with fewer images from the severe stages. This imbalance can lead to biased models, especially when training with a single classifier. To address this, the training dataset in this study was balanced in a way that ensured fair representation of all severity stages, though not perfectly equal across classes. This deliberate approach was taken to reflect the natural distribution of cases in real-world scenarios, helping to ensure the model's performance would generalize well in real-time clinical environments.

Furthermore, diabetic retinopathy exhibits subtle differences between stages, making it difficult for models to distinguish between them based solely on image data. To address these issues, ensemble learning was chosen, as it combines the strengths of multiple models to improve overall classification accuracy. By aggregating predictions from multiple models, ensemble techniques help to mitigate individual model weaknesses, such as overfitting to certain classes or image artifacts, and reduce bias by relying on multiple decision sources.

In this study, pre-trained CNN models, including Inception V3, ResNet50 V2, and DenseNet 201, were fine-tuned on a carefully curated dataset of retinal images. These models were selected based on their state-of-the-art performance in image classification tasks and their varying network architectures, which offer complementary feature extraction capabilities. For example, InceptionV3 excels at multi-scale feature extraction, ResNet50V2 handles deep network training through residual learning, and DenseNet201 promotes feature reuse and gradient flow in its densely connected layers.

The predictions from these models were then combined using a simple averaging ensemble technique. This method aggregates the outputs of each model, providing a more robust and accurate classification compared to any individual model. By leveraging the unique strengths of each CNN, the ensemble model enhances classification performance and reduces the risk of overfitting to the training data.

3.1. **Dataset.** For this study, two datasets were used:

- The APTOS2019: Blindness Detection EDA Dataset .
- The Diabetic Retinopathy Detection Dataset .

To prepare the data for training, the images were divided into groups based on the severity of diabetic retinopathy, as determined by clinicians. One of the majonr obstacles in diagnosing diabetic retinopathy is the intrinsic class imbalance, where the majority of images belong to the early or no-disease stages, leaving fewer samples for the more severe stages. The training dataset was meticulously curated to ensure a balanced representation of images across all severity classes. This step was crucial to avoid bias toward majority classes and to ensure accurate predictions at every stage of the disease. Table 1 displays the distribution of images in each class.

n

The APTOS2019 dataset contains retinal images captured using fundus imaging methods under various conditions. Each image is rated by a clinician for the severity of diabetic retinopathy on a scale from 0 to 4. A total of 3,663 images were included in the training

TABLE 1. Distribution of Images by Severity Stage and Dataset

Stage	Severity Level	Dataset 1 (APTOS2019)	Dataset 2 (DR Detection)	Proposed Dataset (Training Set)	Proposed Dataset (Test Set)
0	No DR	1,796	25,802	2,000	400
1	Mild	369	2,438	1,500	300
2	Moderate	995	5,288	1,500	300
3	Severe	193	872	1,000	200
4	Proliferate	295	708	1,000	200

set, and 1,929 images were included in the test set. Figure. 1 describes APTOS2019 dataset.

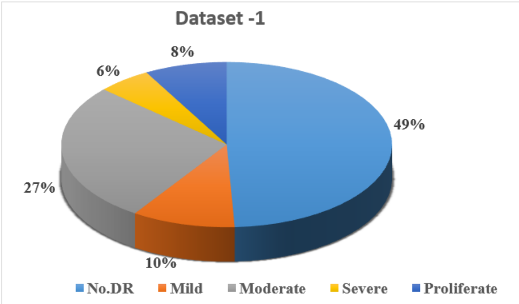

FIGURE 1. Dataset 1 APTOS2019: Blindness Detection EDA Dataset

The Diabetic Retinopathy Detection dataset includes high-resolution retinal images, which have been captured under various imaging conditions, with both left and right fields provided for each subject. Each image has been labeled with a subject ID and assessed by a clinician for the presence of diabetic retinopathy on a scale from 0 to 4. A total of 35,108 images have been included in this dataset. Figure 2 illustrates the distribution of the Diabetic Retinopathy Detection Dataset.

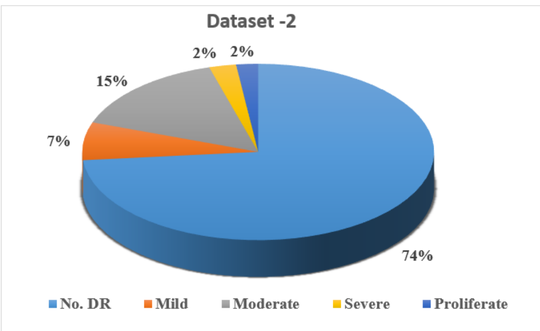

FIGURE 2. Dataset -2- Diabetic Retinopathy Detection Dataset

Images from both the APTOS2019 Blindness Detection EDA Dataset and the Diabetic Retinopathy Detection Dataset were combined to construct a more refined and inclusive dataset for diabetic retinopathy detection. The process involved categorizing the images into classes based on the severity of diabetic retinopathy, as evaluated by clinicians. The representation of images from each severity grade in the training dataset was then meticulously balanced. Figure. 3. describes the proposed dataset.

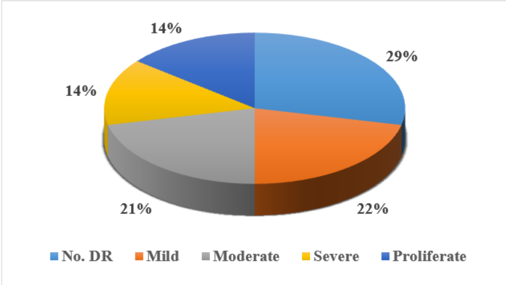

FIGURE 3. Proposed Dataset

3.2. **Model Design.** The diabetic retinopathy detection system used an ensemble of three pre-trained convolutional neural networks (CNNs): InceptionV3, ResNet50V2, and DenseNet201. Each model was chosen for its unique architecture and complementary strengths in feature extraction.

The system begins by feeding the input data—retinal images resized to 224x224 pixels with three color channels (RGB)—through the input layer. This standard size ensures compatibility with the pre-trained models and effective processing. The RGB color channels capture essential visual information related to the retinal features, which is crucial for distinguishing between different stages of diabetic retinopathy.

Each of the three CNNs was fine-tuned using retinal images, allowing them to learn domain-specific features. The models' final dense layers output probability predictions, where each model expresses its confidence in classifying the image into one of the diabetic retinopathy stages. These outputs reflect different perspectives on the retinal image due to the unique architecture of each CNN.

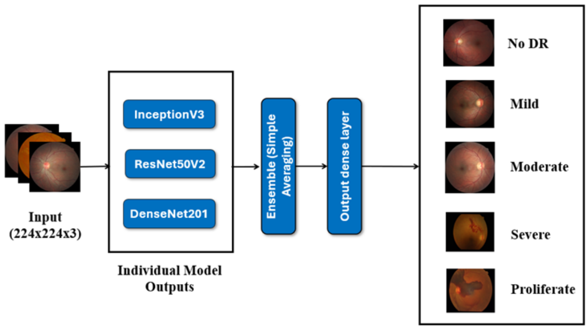

FIGURE 4. Model Summary

Next, the ensemble method combines the predictions from each CNN. A simple averaging technique was used to merge these outputs, creating a unified prediction by averaging the probabilities generated by InceptionV3, ResNet50V2, and DenseNet201. This step minimizes the impact of any potential biases or errors from individual models and leverages the strengths of each network, enhancing the robustness and accuracy of the final prediction.

The final step in the model design involved an output layer with softmax activation, which converts the averaged probabilities into a probability distribution across the possible classes. This ensures the final predictions are interpretable as probabilities, summing to one, and facilitating decision-making by highlighting the most likely stage of diabetic retinopathy.

In terms of monitoring and performance tracking, the system's training and validation accuracy, along with loss curves, were visualized. A confusion matrix was also used to provide deeper insights into the model's classification performance, assisting in the identification of potential issues such as overfitting or underfitting. This allowed for further fine-tuning and optimization of the model's parameters.

In summary, this model design leverages the combined capabilities of multiple CNN architectures, fine-tuning them for retinal image classification and employing ensemble learning to ensure robust and accurate diabetic retinopathy detection.

4. Results and Discussion.

4.1. **Confusion Matrix Analysis and Model Comparison.** The importance of evaluating model performance at the class level is highlighted by the confusion matrix analysis. An N x N matrix, known as a confusion matrix, is used to assess the performance of a classification model, where N represents the number of target classes. This matrix compares the actual target values with those predicted by the machine learning model.In the research, strong performance is demonstrated by InceptionV3, ResNet50V2, and DenseNet201, despite some misclassifications across classes. The ensemble model surpasses the individual models, exhibiting fewer misclassifications overall. The level of misclassification in the ensemble model is significantly lower compared to the individual models, indicating a more consistent performance.The performance of the InceptionV3 model is illustrated in the confusion matrix, showing strong accuracy in Moderate and Proliferative DR stages with 300 and 199 correct classifications, respectively. Some misclassifications are noted, such as No DR being incorrectly identified as Mild DR, indicating minor areas for improvement. Figure. 5. describes the confusion matrix of InceptionV3.

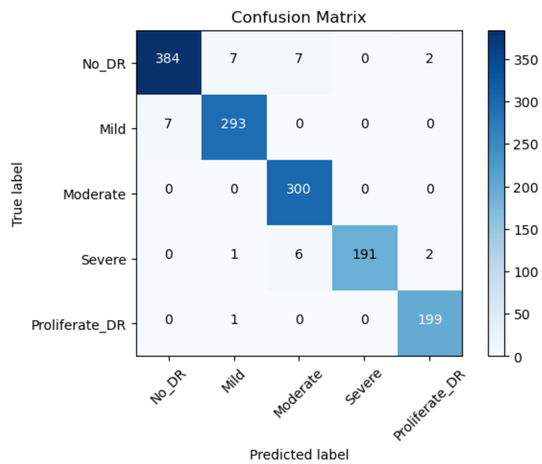

FIGURE 5. Confusion matrix of InceptionV3

The ResNet50V2 model's confusion matrix highlights its high accuracy, particularly in the Mild and Moderate DR stages with 300 and 299 correct predictions. Minimal misclassifications are observed, such as No DR misclassified as Mild, demonstrating the model's overall effectiveness with minor errors. Figure. 6. describes the confusion matrix of ResNet50V2 model.

The DenseNet201 model's confusion matrix reveals high accuracy in No DR and Proliferative DR stages, with 373 and 200 correct classifications. Misclassifications are present, notably between Moderate and Severe DR stages, yet the model performs reliably with some minor errors. Figure. 7. describes the confusion matrix of DenseNet201 model.

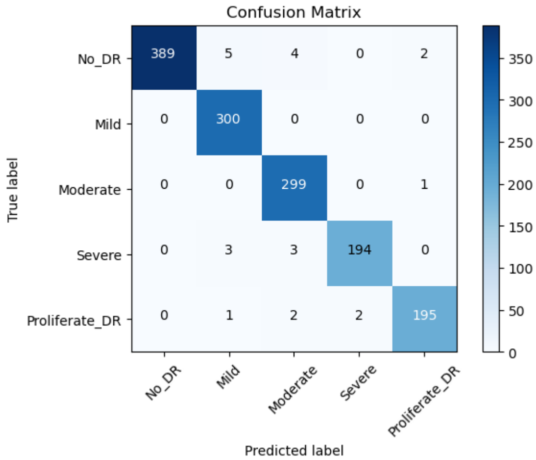

FIGURE 6. Confusion matrix of ResNet50V2 model

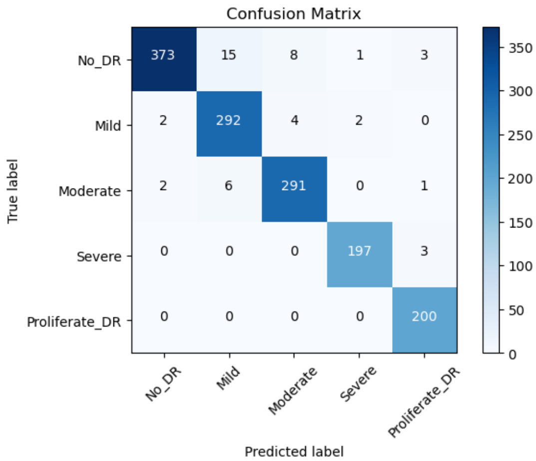

FIGURE 7. Confusion matrix of DenseNet201 Model

The confusion matrix for the ensemble model is noted for its high accuracy, especially in the Moderate and Proliferative DR stages, with a reduced number of misclassifications. It is evident that the strengths of the individual models are effectively combined by the ensemble approach, resulting in enhanced performance relative to the individual models. Figure. 8. describes the confusion matrix of proposed model.

4.2. **Model Performance.** The bar chart below illustrates the individual performance metrics of the InceptionV3, ResNet50V2, DenseNet201, and Ensemble model. Figure. 9. describes the performance of multiple model.

The individual models, InceptionV3, ResNet50V2, and DenseNet201, each exhibited strong performance metrics, including high accuracy and well-balanced precision, recall, and F1-scores. These metrics reflect the effective classification of diabetic retinopathy images into their respective categories with a high degree of accuracy and reliability.

Upon closer examination, ResNet50V2 was identified as the top-performing individual model, demonstrating slightly higher accuracy and precision compared to InceptionV3 and DenseNet201. This superior performance can be attributed to the unique architecture of ResNet50V2, which employs residual connections to facilitate more efficient information flow through the network, thereby enhancing its predictive capabilities.

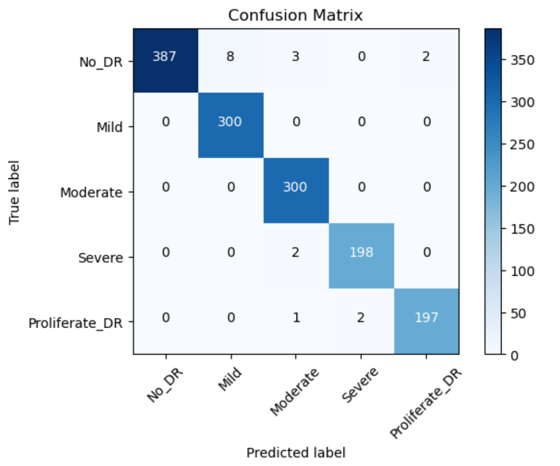

FIGURE 8. Confusion matrix of Ensemble Model

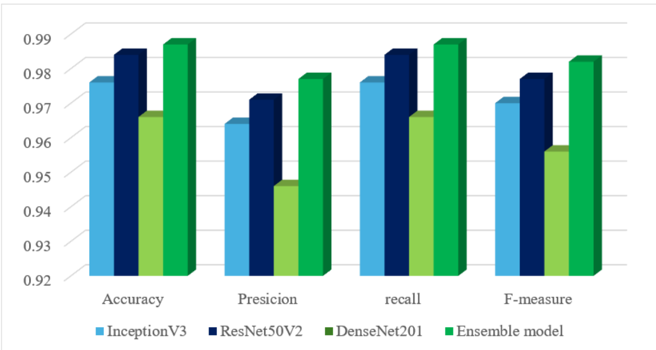

FIGURE 9. Performance analysis of multiple models

However, the ensemble model's performance represents the true highlight of the study. By integrating the diverse strengths of InceptionV3, ResNet50V2, and DenseNet201, the ensemble model achieved exceptional results, surpassing the performance of each individual model. With an accuracy of 98.7%, the ensemble model demonstrated its capability to improve upon the predictive performance of its constituent models.

In conclusion, although each individual model performed admirably, the ensemble model emerged as the clear leader, highlighting the benefits of combining multiple models to achieve superior performance in diabetic retinopathy detection. This underscores the value of employing ensemble methods to improve predictive accuracy and reliability in complex tasks.

4.3. **Accuracy and Loss Graph.** The accuracy curve for the InceptionV3 model shows a steady increase over the training epochs, indicating consistent improvement in the model's ability to correctly classify diabetic retinopathy images. The loss curve demonstrates a gradual decrease, suggesting effective learning and optimization during training. Figure. 10. describes the accuracy and loss graph of InceptionV3 model.

The accuracy curve for the ResNet50V2 model illustrates a rapid improvement in classification accuracy during the early epochs, followed by a more gradual increase as training progresses. This indicates that the model quickly learns to classify images correctly and continues to refine its performance. The loss curve shows a consistent downward trend, signifying effective convergence and a reduction in classification errors. Overall, the curves

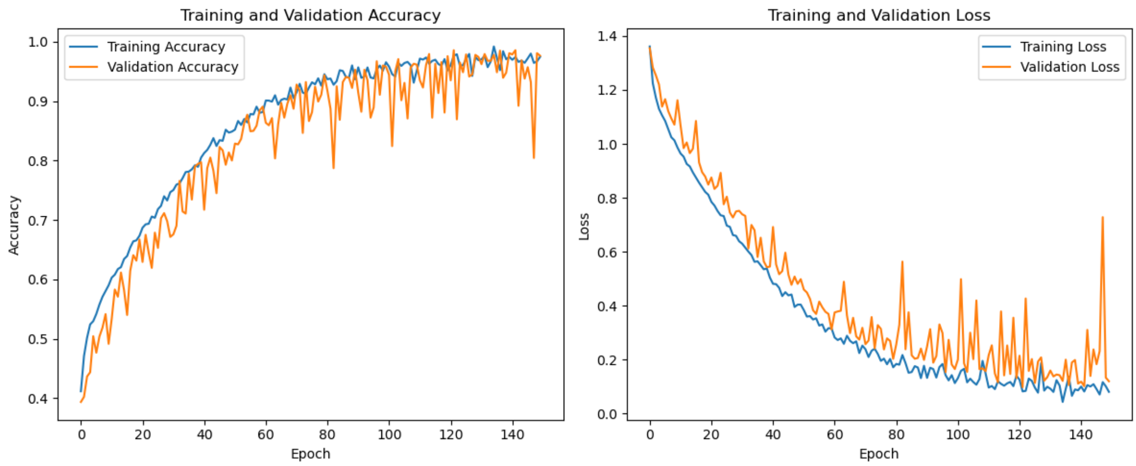

FIGURE 10. Accuracy and loss graph of InceptionV3 Model

demonstrate ResNet50V2's strong performance and effective learning dynamics. Figure. 11. describes the accuracy and loss graph of ResNet50V2 model.

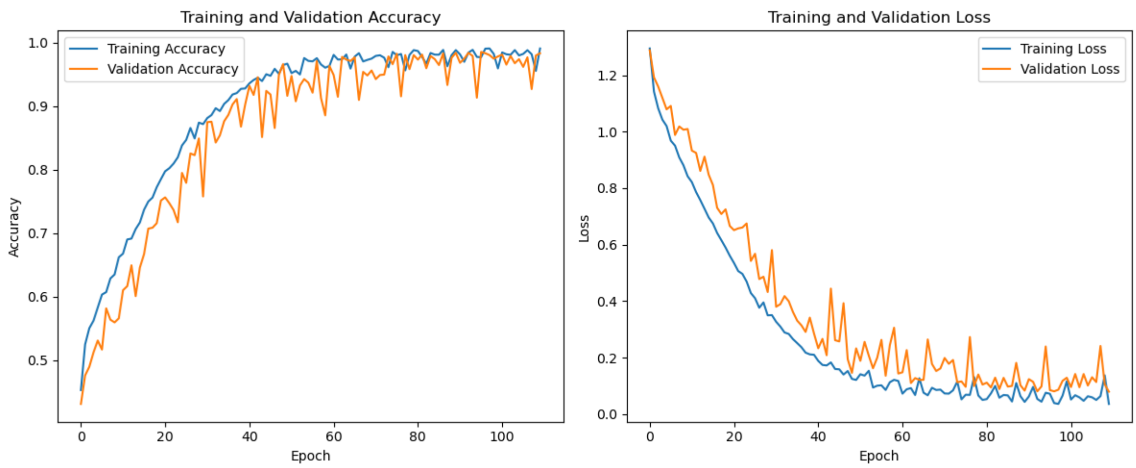

FIGURE 11. Accuracy and loss graph of ResNet50V2 Model

The accuracy curve for the DenseNet201 model exhibits a smooth upward trajectory, indicating continuous enhancement in the model's classification accuracy throughout the training period. The loss curve similarly displays a steady decline, reflecting successful optimization and learning. Figure. 12 describes the accuracy and loss graph of DenseNet201 Model.

The accuracy and loss curves for the Ensemble model are shown to illustrate a consistent improvement in performance throughout the training process. It is indicated by the accuracy curve that the model's ability to correctly classify diabetic retinopathy images steadily enhances, while the loss curve reflects a continuous decrease, suggesting effective learning and optimization. Figure. 13 describes the accuracy and loss graph of proposed Model.

The system effectively identified the presence of diabetic retinopathy and accurately determined its severity in all of the evaluated photos, as expected. These findings emphasize the system's resilience and demonstrate how well it functions as a trustworthy diagnostic tool for differentiating between various DR stages. Comparison between the existing system and the proposed work is given in below: Table 2 describes the comparison

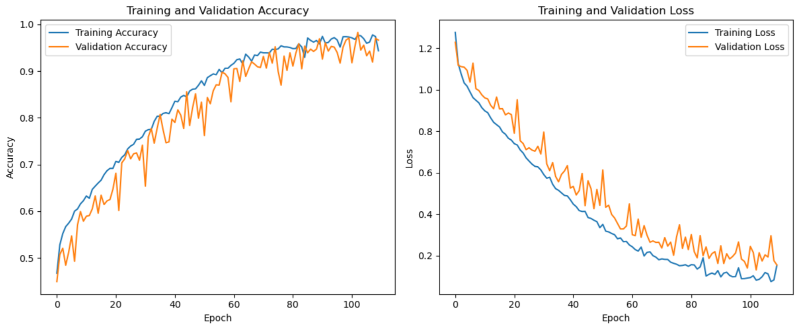

FIGURE 12. Accuracy and loss graph of DenseNet201 Model

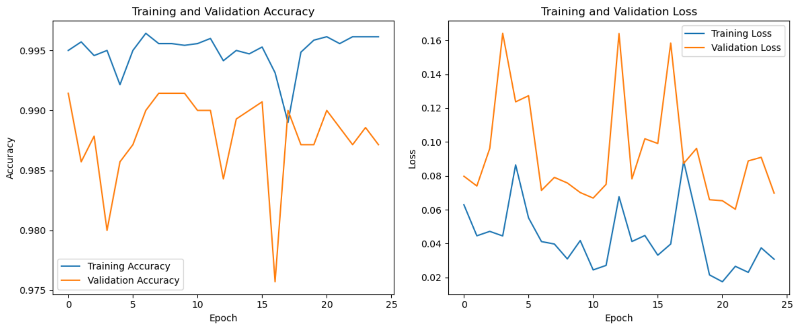

FIGURE 13. Accuracy and loss graph of Ensemble Model (Simple Averaging)

TABLE 2. Comparison of Different Models for Diabetic Retinopathy Detection

Authors (Year)	**Dataset (Samples)**	**Applied Models**	**Measures (Proposed Model)**
Hussain, M., Ahmed 2025 [30]	Diabetic Retinopathy Detection dataset	CNN	93%
Guefrachi, S., Echtioui 2024 [29]	Diabetic Retinopathy Detection dataset	InceptionResNetV2	96.61%
Abood, R.H., Hamad 2025 [28]	Diabetic Retinopathy Detection dataset, APTOS-2019	EfficientNetB6	94%
Aftab, S. 2025 [27]	APTOS 2019, IDRiD, and Messidor-2	Ensemble model with Averaging (EfficientNetB2, DenseNet121, and ResNet50)	96.96%
Y. Sun 2019 [4]	Medical Dataset	Random Forest	92%
Proposed Work	APTOS2019, Diabetic Retinopathy Dataset	Bias Prevention, Accuracy Improvement	**Accuracy: 98.7%**, Precision: 97.7%, Recall: 98.7%, F1-Score: 98.2%

5. **Conclusion.** Using a variety of retinal pictures, our work created and assessed a deep learning-based system for automated diabetic retinopathy (DR) identification and categorization. When it came to DR stage classification, individual models such as InceptionV3,

ResNet50V2, and DenseNet201 shown excellent efficacy and accuracy. By integrating predictions from several architectures, the ensemble model fared better than the individual models, improving accuracy and resilience. Confusion matrix analysis revealed opportunities for improvement as well as performance in class-level categorization. The approach proved to be dependable in determining the existence and severity of DR at different phases. To guarantee its clinical usability and efficacy in actual healthcare settings, more improvement and validation are required.

REFERENCES

[1] J. B. Jonas, R. R. Bourne, R. A. White, S. R. Flaxman, J. Keeffe, J. Leasher, K. Naidoo, K. Pesudovs, H. Price, T. Y. Wong et al., *"Visual impairment and blindness due to macular diseases globally: a systematic review and meta-analysis,"* American journal of ophthalmology, vol. 158, no. 4, pp. 808–815, 2014.

[2] J. L. Leasher, R. R. Bourne, S. R. Flaxman, J. B. Jonas, J. Keeffe, K. Naidoo, K. Pesudovs, H. Price, R. A. White, T. Y. Wong et al., *"Global estimates on the number of people blind or visually impaired by diabetic retinopathy: a meta-analysis from 1990 to 2010,"* Diabetes care, vol. 39, no. 9, pp. 1643–1649, 2016.

[3] T. A. Soomro, A. J. Afifi, L. Zheng, S. Soomro, J. Gao, O. Hellwich, and M. Paul, *"Deep learning models for retinal blood vessels segmentation: a review,"* IEEE Access, vol. 7, pp. 71 696–71 717, 2019.

[4] Y. Sun and D. Zhang, *"Diagnosis and analysis of diabetic retinopathy based on electronic health records,"* Ieee Access, vol. 7, pp. 86 115–86 120, 2019.

[5] M. Poostchi, K. Silamut, R. J. Maude, S. Jaeger, and G. Thoma, *"Image analysis and machine learning for detecting malaria,"* Translational Research, vol. 194, pp. 36–55, 2018.

[6] N. M. Bressler, *"Age-related macular degeneration is the leading cause of blindness..."* Jama, vol. 291, no. 15, pp. 1900–1901, 2004.

[7] H. Ye, Q. Zhang, X. Liu, X. Cai, W. Yu, S. Yu, T. Wang, W. Lu, X. Li, H. Jin et al., *"Prevalence of age-related macular degeneration in an elderly urban chinese population in china: the jiangning eye study,"* Investigative ophthalmology & visual science, vol. 55, no. 10, pp. 6374–6380, 2014.

[8] J. M. P. Dias, C. M. Oliveira, and L. A. da Silva Cruz, *"Retinal image quality assessment using generic image quality indicators,"*Information Fusion, vol. 19, pp. 73–90, 2014.

[9] H. Cui, S. Shen, W. Gao, H. Liu, and Z. Wang, *"Efficient and robust large-scale structure-from-motion via track selection and camera prioritization,"* ISPRS Journal of Photogrammetry and Remote Sensing, vol. 156, pp. 202–214, 2019.

[10] G. Molodij, E. Ribak, M. Glanc, and G. Chenegros, *"Enhancing retinal images by extracting structural information,"* Optics Communications, vol. 313, pp. 321–328, 2014.

[11] E. Peli and T. Peli, *"Restoration of retinal images obtained through cataracts,"* IEEE transactions on medical imaging, vol. 8, no. 4, pp. 401–406, 1989.

[12] H. Liu, X. Tang, and S. Shen, *"Depth-map completion for large indoor scene reconstruction,"* Pattern Recognition, vol. 99, p. 107112, 2020.

[13] D. B. Rein, P. Zhang, K. E. Wirth, P. P. Lee, T. J. Hoerger, N. McCall, R. Klein, J. M. Tielsch, S. Vijan, and J. Saaddine, *"The economic burden of major adult visual disorders in the united states,"* Archives of ophthalmology, vol. 124, no. 12, pp. 1754–1760, 2006.

[14] M. Abdul Baker Chowdhury, M. Uddin, H. M. Khan, and M. Haque, *"Type 2 diabetes and its correlates among adults in bangladesh: a population based stud,"* 2015.

[15] T. Y. Wong and N. M. Bressler, *"Artificial intelligence with deep learning technology looks into diabetic retinopathy screening,"* Jama, vol. 316, no. 22, pp. 2366–2367, 2016.

[16] V. Gulshan, L. Peng, M. Coram, M. C. Stumpe, D. Wu, A. Narayanaswamy, S. Venugopalan, K. Widner, T. Madams, J. Cuadros et al., *"Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs,"* jama, vol. 316, no. 22, pp. 2402–2410, 2016.

[17] D. S. W. Ting, C. Y.-L. Cheung, G. Lim, G. S. W. Tan, N. D. Quang, A. Gan, H. Hamzah, R. Garcia-Franco, I. Y. San Yeo, S. Y. Lee et al., *"Development and validation of a deep learning system for diabetic retinopathy and related eye diseases using retinal images from multiethnic populations with diabetes,"* Jama, vol. 318, no. 22, pp. 2211–2223, 2017.

[18] M. D. Abràmoff, Y. Lou, A. Erginay, W. Clarida, R. Amelon, J. C. Folk, and M. Niemeijer, *"Improved automated detection of diabetic retinopathy on a publicly available dataset through integration of deep learning,"* Investigative ophthalmology & visual science, vol. 57, no. 13, pp. 5200–5206, 2016.

[19] M. D. Abr'amoff, P. T. Lavin, M. Birch, N. Shah, and J. C. Folk, *"Pivotal trial of an autonomous ai-based diagnostic system for detection of diabetic retinopathy in primary care offices,"* NPJ digital medicine, vol. 1, no. 1, p. 39, 2018.

[20] M. Niemeijer, B. Van Ginneken, M. J. Cree, A. Mizutani, G. Quellec, C. I. S´anchez, B. Zhang, R. Hornero, M. Lamard, C. Muramatsu et al., *"Retinopathy online challenge: automatic detection of microaneurysms in digital color fundus photographs,"* IEEE transactions on medical imaging, vol. 29, no. 1, pp. 185–195, 2009.

[21] M. D. Abr'amoff, J. C. Folk, D. P. Han, J. D. Walker, D. F. Williams, S. R. Russell, P. Massin, B. Cochener, P. Gain, L. Tang et al., *"Automated analysis of retinal images for detection of referable diabetic retinopathy,"* JAMA ophthalmology, vol. 131, no. 3, pp. 351–357, 2013.

[22] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, *"Densely connected convolutional networks,"* in Proceedings of the IEEE conference on computer vision and pattern recognition, 2017, pp. 4700–4708.

[23] K. He, X. Zhang, S. Ren, and J. Sun, *"Deep residual learning for image recognition,"* in Proceedings of the IEEE conference on computer vision and pattern recognition, 2016, pp. 770–778.

[24] M. Tan, *"Efficientnet: Rethinking model scaling for convolutional neural networks,"* arXiv preprint arXiv:1905.11946, 2019.

[25] Rehman, A.U., Lee, C.F.; *CNN-Based Intelligent Disease Detection and Identification Technique Through Chest X-Rays*; J. Inform. Hiding Multim. Signal Process.; 15(4); 319–333; 2024.

[26] Radwan, G.A., Khafagy, M.H., Mabrouk, M.T.M., Mohamed, M.H.;*Coronary Artery Disease Prediction by Combining Three Classifiers*; J. Inform. Hiding Multim. Signal Process.; 15(4); 221–235; 2024.

[27] Aftab, S., Akhtar, S.; *Diabetic Retinopathy Severity Classification Using Data Fusion and Ensemble Transfer Learning*; J. Softw. Eng. Appl.; 18(1); 1–23; 2025.

[28] Abood, R.H., Hamad, A.H.;*Multi-Label Diabetic Retinopathy Detection Using Transfer Learning Based Convolutional Neural Network*; Fusion Pract. Appl.; 2; 279–279; 2025.

[29] Guefrachi, S., Echtioui, A., Hamam, H.; *Diabetic Retinopathy Detection Using Deep Learning Multistage Training Method*; Arab. J. Sci. Eng.; 1–18; 2024.

[30] Hussain, M., Ahmed, H.A., Babar, M.Z., Ali, A., Shahzad, H.M., ur Rehman, S., Khan, H., Alshahrani, A.M.; *An Enhanced Convolutional Neural Network (CNN) Based P-EDR Mechanism for Diagnosis of Diabetic Retinopathy (DR) Using Machine Learning*; Eng. Technol. Appl. Sci. Res.; 15(1); 19062–19067; 2025.

[31] Chilukoti, S.V., Shan, L., Tida, V.S., Maida, A.S., Hei, X. A Reliable Diabetic Retinopathy Grading via Transfer Learning and Ensemble Learning with Quadratic Weighted Kappa Metric; *BMC Med. Inform. Decis. Mak.*; 24(1); 37; 2024.

[32] Desiani, A., Primartha, R., Hanum, H., Dewi, S.R.P., Suprihatin, B., Al-Filambany, M.G., Suedarmin, M.; Weighted Voting Ensemble Learning of CNN Architectures for Diabetic Retinopathy Classification; *J. Infotel*; 16(1); 136–155; 2024

[33] N. Chowdhury, P. P. Choudhury, and S. R. Moon, "Pneumonia stage analyzes through image processing,"*Indonesian Journal of Electrical Engineering and Computer Science*, vol. 36, no. 3, pp. 1778–1786, 2024.

[34] N. Chowdhury, J. Sultana, T. Rahman, T. Chowdhury, F. T. Khan, and A. Chakraborty, "Potato leaf disease detection through ensemble average deep learning model and classifying the disease severity," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 35, no. 1, pp. 494–502, 2024.

[35] N. Chowdhury and M. S. Arefin, "Skyline path queries for location-based services,"*International Journal of Advanced Computer Science and Applications*, vol. 10, no. 5, 2019.