Deep Feature Extraction and Classification of Diabetic Retinopathy Using AlexNet, InceptionV3, and VGG16 CNN Architectures

Shubhi Shrivastava[1], Shanti Rathore[2], and Rahul Gedam[3]

[1] Princeton University, Princeton NJ 08544, USA
[1]Dr. CV Raman University, Kota Bilaspur, India
[2] ET & T Department, Dr. CV Raman University, Kota Bilaspur, India
[3] ET & T Department, LCIT, Bilaspur, India
shubhi@lcit.edu.in

Abstract. Diabetic retinopathy (DR) is a significant cause of vision impairment and blindness among diabetic patients, characterized by progressive retinal damage. Early and accurate detection is crucial for effective management and treatment. This research explores advanced deep learning techniques to enhance DR detection and classification by leveraging Convolutional Neural Networks (CNNs). We propose a novel methodology incorporating deep feature extraction and classification using three CNN architectures: AlexNet, InceptionV3, and VGG16. Our approach involves extracting deep features from retinal images to capture intricate patterns associated with various DR stages, followed by classification to differentiate between healthy and various stages of DR. The dataset used include publicly available Fundus Image Registration Dataset (FIRE) for comprehensive evaluation. Detailed preprocessing steps ensured data quality and relevance, while feature extraction techniques harnessed the strengths of the selected CNN architectures. The performance of the proposed models was evaluated based on accuracy, sensitivity, precision, and F1-score. Our results demonstrate that AlexNet achieves the highest accuracy at 95.37%, outperforming InceptionV3 and VGG16. This study underscores the effectiveness of CNN-based approaches in DR detection and highlights the potential for further improvements in early diagnosis and treatment strategies.

Keywords: AlexNet, Convolutional Neural Networks, Diabetic Retinopathy InceptionV3, VGG16.

1 Introduction

Diabetic retinopathy (DR) is a severe complication of diabetes, leading to damage to the retina and potentially causing blindness if left untreated. As the prevalence of diabetes rises globally, so does the incidence of DR, making it a critical public health issue. Early detection and treatment are essential to prevent vision loss, but manual screening of retinal images by ophthalmologists is time-consuming and prone to

S. Gupta et al. (eds.), *Proceedings of the 5th International Conference on the Role of Innovation, Entrepreneurship and Management for Sustainable Development (ICRIEMSD 2024)*,
Advances in Economics, Business and Management Research 310,
https://doi.org/10.2991/978-94-6463-612-3_7

variability. Automated detection systems using deep learning have shown promise in addressing these challenges by providing accurate, efficient, and scalable solutions.
Deep learning, particularly convolutional neural networks (CNNs), has revolutionized the field of medical imaging. CNNs are adept at learning hierarchical feature representations from raw images, making them well-suited for tasks such as image classification, object detection, and segmentation. In the context of DR detection, CNNs can be employed to extract deep features from retinal images and classify them into different stages of the disease.

This research paper aims to explore the effectiveness of using deep feature extraction through CNNs followed by classification using state-of-the-art CNN architectures: AlexNet, InceptionV3, and VGG16. By comparing these architectures, the study seeks to identify the most suitable model for accurate and efficient classification of DR stages.

The proposed approach involves a two-step process: feature extraction and classification. First, retinal images are processed through a CNN to extract deep features. These features are then classified into different stages of DR using the same or another CNN architecture. The three architectures under investigation are:

- AlexNet: One of the pioneering deep learning models, AlexNet is known for its relatively simple architecture with five convolutional layers, followed by three fully connected layers. Despite its simplicity, it has demonstrated strong performance in image classification tasks.
- InceptionV3: Part of the Inception family, InceptionV3 introduces the concept of "inception modules," which allow the network to capture multi-scale features through multiple convolutional filter sizes. This architecture is deeper and more complex, designed to improve both the accuracy and efficiency of the model.
- VGG16: Known for its simplicity and depth, VGG16 consists of 16 layers, including 13 convolutional layers and 3 fully connected layers. It uses very small (3×3) convolution filters, which enhances the depth of the network while maintaining a manageable number of parameters.

The primary contributions of this research are as follows:

- Comparative Analysis: Providing a comprehensive comparison of AlexNet, InceptionV3, and VGG16 in the context of DR classification, focusing on their performance metrics, computational efficiency, and suitability for clinical applications.
- Feature Extraction Insights: Evaluating the effectiveness of deep feature extraction using CNNs and its impact on the classification accuracy of DR stages.
- Framework Development: Developing a robust and scalable framework for automated DR detection, which can be integrated into clinical practice to assist ophthalmologists in early diagnosis and treatment planning.

The remainder of this paper is structured as follows: Section 2 reviews related work in DR detection and deep learning applications in medical imaging. Section 3 details the proposed methodology, including data preprocessing, model architectures, and deep features extraction. Section 4 presents the experimental setup and results. Finally, Section 5 concludes the paper and outlines future research directions.

2 Literature Review

DR is a progressive disease that affects patients with diabetes. This disease goes through several stages that can lead to blindness if it is not detected and treated promptly. In this sense, it is possible to find two-class classification systems or algorithms (healthy patient vs. sick patient) and more detailed ones of up to five classes (healthy, mild, moderate, severe or PDR), all of the above depending on the techniques or data available.

2.1 Work Focused on the Detection of DR

In this sense, the authors of [1] proposes estimating the probability of occurrence of the disease based on the location of two specific lesions in the available images: microaneurysms and hemorrhages. These lesions characterize the initial states of DR; detection is performed using a convolutional neural network (CNN) considering the detection of poorly classified examples, in order to improve the performance of the system that reports a sensitivity of 94% and an area under the characteristic operating curve of 0.912. In [2], the detection of the disease is proposed from a CNN with a radial basis function. The neural network is trained from the characteristics of microaneurysms and hemorrhages, considering the area of the lesions, their perimeter, their circularity, the number of lesions, among other characteristics. The proposed model achieves a sensitivity of 87% for a specificity of 93%. In [3] in a similar way, a CNN with a radial basis function is used, but in this case it is considered to determine the presence of microaneurysms, exudates and the segmentation of the blood vessels of the retina. In this way, an accuracy of 71.2% and 89.4% is reported when using the DIARETDB0 and DIARETDB1 databases respectively. In [4], a method is proposed to reduce the structural complexity of a CNN, in this case, the detection of the disease is based on a hierarchical pruning approach by modifying a VGG-16 network. Feature extraction is performed using a pre-trained model on the ImageNet dataset. When applying the proposed model, a 35% decrease in the feature map used is reported without significantly decreasing the performance of the system. In this case, an accuracy of 92% is reported for a sensitivity of 98.33% and a specificity of 83.7%. In [5], the images are classified as normal or severe non-proliferative, for this a hybrid approach is used by extracting the features with a deep learning approach and classifying from a support vector machine (SVM). This methodology is evaluated on databases 12 and 13 of the Messidor set, reporting an accuracy of 95.83% and 95.24% respectively. It is striking that the behavior of the algorithm is only evaluated on a reduced data set, with more data available within the database itself and the authors do not give details regarding their decision. On the other hand, in [6], a CNN-based approach is proposed, this model is trained and subsequently evaluated using the EyePACS-Kaggle and Messidor-2 databases, the authors report an accuracy of 85.7% and 91% respectively. This proposal has the peculiarity of predicting the importance score of each pixel in the input image, this prediction allows determining the most relevant pixel in the decision and in this way it is easier for experts to verify the results reported by the model. Another approach detects the disease by developing an adaptive

momentum classifier [7]. The algorithm determines the presence of exudates using the Gabor transform and reports an accuracy of 98.4% on the STARE and DRIVE databases. Meanwhile, [8] detects the disease by extracting different features of interest using morphological operations and from these features the classification is performed using SVM.

2.2 Works Focused on DR Classification

In [9], the disease is classified as "mild NPDR", "moderate NPDR" or "severe NPDR" for this purpose the morphology extraction of the vascular network is performed, the mean capillary area is determined as a region of interest, the avascular zone of the fovea, its perimeter and its vascular index are included or excluded from the analysis. This method reports an accuracy of 97%. Another region-based approach is presented in [10]. In this case, the retina surface is divided into four areas of interest: central, temporal, nasal, inferior and superior, focusing on the regions concentric to the macula. In these regions, the presence of characteristic lesions is determined. The classification process is carried out by discriminating the location of the lesions from the "severe" to the "mild" state; this discrimination allows to reduce the necessary processing times and reports an overall accuracy of 90% for a sensitivity of 92%, 93.75% and 90.75% for the Messidor, DRIVE and a private databases, respectively. In [11], a classification system is developed using a CNN. The proposed design consists of two convolutional layers and two alternating pooling layers, using ReLu as a nonlinear activation function. The scheme proposes to reduce the number of convolutional layers to avoid overfitting problems in the network parameters. The approach achieves an accuracy of 88% on the EyePACS-Kaggle and DB1 databases. In [12], an approach for RDNP detection and classification is proposed using a CNN from publicly available fundus images on the Internet. The segmentation process is performed using canny edge detection to detect the presence of characteristic lesions. This method reports an accuracy of 90.89%. In [13], it is proposed to use a non-linear label smoothing technique during the CNN training processes and an area under the operating curve of 0.9158 is reported. In this case, a private database is used, so it is not possible to reproduce the research with the same reported data. On the other hand, in [14], an approach is established using a ResNet50 architecture. This strategy allows feature extraction without applying a priori any transformation to improve image quality or enhance specific structures. A transfer learning paradigm is used with a network pre-trained on the ImageNet database. The approach reports an area under the curve of 0.93, 0.81, 0.92 and 0.97 when classified into "mild NPDR", "moderate NPDR", "severe NPDR" and "healthy vs. sick" respectively; although the results are limited to the Messidor database.

In [15], a scheme based on a DenseNet network is proposed on the EyePACS-Kaggle database and reports a Cohens Kappa score of 0.8836 and 0.9809 for the validation and training processes respectively. Another similar proposal is presented in [16], but modifying the architecture of the AlexNet network using a non-linear ReLu activation function. In this case, an accuracy of 96.6%, 96.2% and 96.6% is reported for the "healthy", "mild NPDR", "moderate-severe NPDR" and "proliferative" images respectively; a lower accuracy is obtained in the "moderate-severe NPDR" category

because the database is not adequately balanced and there are fewer examples available for this category.

In [17], a hybrid statistical approach is developed for image classification and DR detection. The method is based on combining the SVM with the scaled Dirichlet distribution on the DRIVE, HRF and private VDIS databases. This approach reports an accuracy of 90.7% for an area under the curve of 0.87. On the other hand, in [18], the texture features are extracted using an oriented gradient histogram technique. This technique is sensitive to the deformations present in the images in the form of lesions and allows them to be represented effectively. The features are supplied to an SVM that performs binary classifications, discriminating each of the classes until the image is placed in the appropriate category. The validation of the model is carried out in the DIARETDB0 database and an accuracy of 85% is reported. Another approach is proposed in [19]. In this case, the algorithm highlights bright lesions through a normalization process, followed by an intensity thresholding for lesion detection; in this way, the robustness of the algorithm and the rejection of false positives are guaranteed. An SVM classifier is used that considers 10 different types of features, performing a quantitative analysis of red, bright lesions and the optic disc for decision making. This approach reports an accuracy of 84% and 92.13% when using the Messidor and DIARETDB1 databases respectively. To classify diabetic retinopathy, the researchers in [20] employed Support Vector Machines (SVM) and k-Nearest Neighbors (KNN), yielding promising comparative results.

2.3 Research Gap

Despite the significant advancements in diabetic retinopathy (DR) detection and classification using deep learning techniques, several gaps remain in the existing literature. Many studies have demonstrated promising results using convolutional neural networks (CNNs) for DR detection by focusing on specific features like microaneurysms and hemorrhages. However, the variability in datasets, feature extraction methods, and classification criteria across different studies limits the generalizability of these findings. For instance, some works achieve high accuracy and sensitivity but are tested on limited or private datasets, making it difficult to reproduce and validate the results on a broader scale. Furthermore, the majority of the research has concentrated on either binary classification (healthy vs. DR) or limited multi-class classification (up to three classes), neglecting the more granular stages of DR, which are crucial for precise treatment planning.

Additionally, while some studies have explored hybrid approaches combining CNNs with other machine learning techniques like support vector machines (SVMs) or radial basis functions, there is a lack of comprehensive comparisons between purely deep learning-based methods and hybrid models. Most studies also fail to address the computational efficiency and scalability of their models, which are critical factors for real-world clinical implementation. Moreover, the influence of different CNN architectures on feature extraction and classification performance remains underexplored, particularly for advanced architectures like InceptionV3 and VGG16 compared to simpler models like AlexNet.

Given these gaps, it is essential to develop a more robust and scalable framework that leverages advanced CNN architectures for deep feature extraction and classification of DR stages. This research aims to fill these gaps by systematically comparing the performance of AlexNet, InceptionV3, and VGG16 in detecting and classifying DR using a comprehensive dataset. By doing so, we seek to provide a clearer understanding of the most effective techniques and architectures for automated DR detection, ultimately contributing to improved early diagnosis and patient outcomes.

3 Proposed Methodology

The proposed methodology involves a comprehensive approach to diabetic retinopathy (DR) detection and classification using advanced convolutional neural network (CNN) architectures. The process is divided into several key stages: data collection and preprocessing, model architecture selection, training and validation, feature extraction, classification, and performance evaluation. The CNN architectures under investigation are AlexNet, InceptionV3, and VGG16. Figure 1 shows the flow diagram for proposed DR detection.

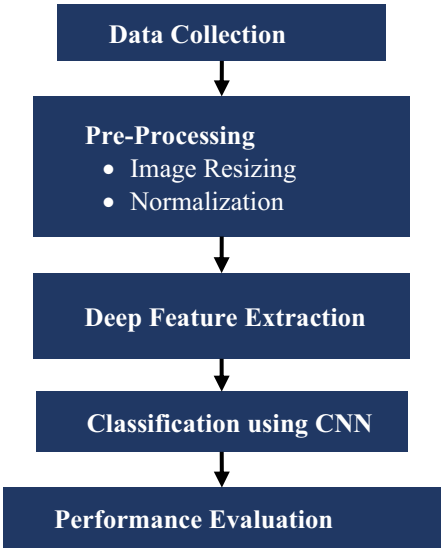

Fig. 1. Flow diagram of proposed work

3.1 Data Collection

The data collection for this study involves sourcing a substantial dataset of retinal images from multiple publicly available databases. The Fundus Image Registration Dataset (FIRE) is utilized to further enrich the dataset. The FIRE dataset comprises 129 retinal images, forming 134 image pairs categorized into three distinct groups based on specific characteristics. These images were captured using a Nidek AFC-210 fundus camera, providing high-resolution images (2912×2912 pixels) with a 45° field of view

in both x and y dimensions. The images were collected at Papageorgiou Hospital, associated with Aristotle University of Thessaloniki, and involve data from 39 patients.

3.2 Preprocessing

It is a crucial step in preparing retinal images for deep learning models. It involves a series of operations to standardize the images, enhance relevant features, and reduce computational complexity. The detailed steps and mathematical formulations for preprocessing are as follows:

3.2.1 Image Resizing

Resize all images to 224×224 pixels to ensure consistency and compatibility with AlexNet, InceptionV3, and VGG16 architectures.

$$I' = Resize(I, 224{,}224) \tag{1}$$

3.2.2 Image Normalization

- Normalize pixel values to a standard range [0, 1].
- Normalize each RGB channel separately.

$$I_n(x, y, c) = \frac{I'(x,y,c) - \mu_c}{\sigma_c} \tag{2}$$

Here, μ_c and σ_c are the mean and standard deviation of pixel values in channel cc across the dataset.

3.3 Deep Feature Extraction

In this research, we employ pre-trained convolutional neural network (CNN) architectures to extract deep features from retinal images. The selected architectures are AlexNet, InceptionV3, and VGG16. The steps for extracting features are detailed below:

3.3.1 Model Architecture Loading

We utilize pre-trained models (AlexNet, InceptionV3, VGG16) and remove the final fully connected (FC) layers to focus on the convolutional layers for feature extraction.

$$Model_{new} = Model_{pretrained} - FC\ layers \tag{3}$$

3.3.2 Image Input and Preprocessing

The input images I are resized and normalized according to the preprocessing steps mentioned earlier. Let I' represent the preprocessed image.

3.3.3 Forward Pass Through CNN

The preprocessed image I' is passed through the convolutional layers of the CNN to obtain feature maps.

$$FeatureMaps = Model_{new}(I') \tag{4}$$

For each architecture, the feature maps are obtained as follows:

AlexNet:

The feature maps are extracted from the fifth convolutional layer (Conv5).

$$FeatureMaps_{AlexNet} = Conv5(I') \tag{5}$$

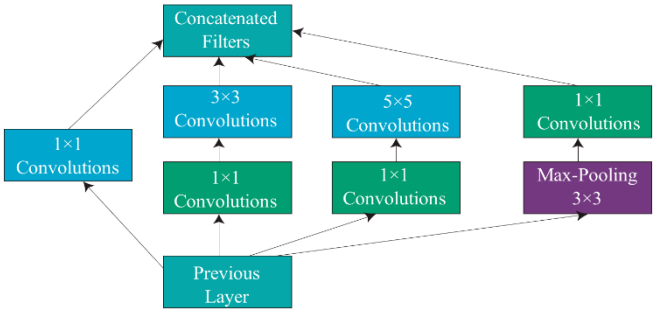

Fig. 2. AlexNet architecture

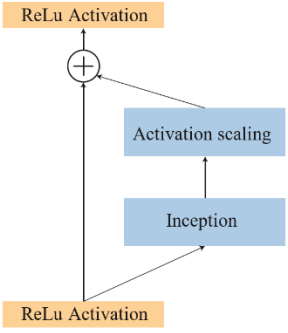

Fig. 3. InceptionV3 architecture

InceptionV3:

The feature maps are extracted from the final inception module (Mixed10).

$$FeatureMaps_{InceptionV3} = Mixed10(I') \tag{6}$$

VGG16:

The feature maps are extracted from the last convolutional layer (Conv13).

$$FeatureMaps_{VGG16} = Conv13(I') \tag{7}$$

3.3.4 Flattening Feature Maps

The obtained feature maps are flattened into a one-dimensional vector to create a fixed-length feature representation for each image.

Given a feature map F of dimensions $H \times W \times D$ (height, width, depth), the flattened feature vector f is obtained as:

$$f = flatten(F) \tag{8}$$

Where:

$$f = [f_1, f_2, \ldots, f_{H \times W \times D}] \tag{9}$$

3.3.5 Feature Vector Extraction

For each image, the feature vector f is extracted as follows:

AlexNet:

The feature vector is derived from the Conv5 layer.

$$f_{AlexNet} = flatten(FeatureMaps_{AlexNet}) \tag{10}$$

InceptionV3:

The feature vector is derived from the Mixed10 layer.

$$f_{InceptionV3} = flatten\left(FeatureMaps_{InceptionV3}\right) \tag{11}$$

VGG16:

The feature vector is derived from the Conv13 layer.

$$f_{VGG16} = flatten(FeatureMaps_{VGG16}) \tag{12}$$

3.3.6 Feature Vector Usage

The extracted feature vectors f are subsequently used for the classification stage. These vectors serve as inputs to the classifier, which is trained to categorize the images into different stages of diabetic retinopathy (DR).

By employing the above methodology, deep features extracted from AlexNet, InceptionV3, and VGG16 effectively capture the essential patterns and characteristics of retinal images, providing a robust foundation for accurate DR classification.

3.4 Classification Using CNN

The classification phase involves using the extracted deep features from the CNN architectures (AlexNet, InceptionV3, and VGG16) to classify retinal images into different stages of diabetic retinopathy (DR). This process includes constructing a classifier on top of the extracted features and fine-tuning the entire model for optimal performance. The steps and mathematical formulations are detailed below:

3.4.1 Fully Connected Layers Addition

After extracting the deep features, fully connected (FC) layers are added to perform classification. This involves adding dense layers, dropout layers to prevent overfitting, and an output layer with a softmax activation function for multi-class classification.

Let f be the flattened feature vector obtained from the CNN:

$$f = [f_1, f_2, \ldots, f_N] \tag{13}$$

Where N is the length of the feature vector.

The fully connected layers are defined as follows:

First Fully Connected Layer:

$$h_1 = \sigma(W_1 f + b_1) \tag{14}$$

Where:

- W_1 is the weight matrix of the first FC layer.
- b_1 is the bias vector of the first FC layer.

- σ is an activation function, typically ReLU (Rectified Linear Unit).

Dropout Layer: Dropout is applied to the activations to prevent overfitting during training.

$$h_1^{dropout} = Dropout(h_1, p) \tag{15}$$

Where p is the dropout rate.

Second Fully Connected Layer:

$$h_2 = \sigma\left(W_2 h_1^{dropout} + b_2\right) \tag{16}$$

Where:

- W_2 is the weight matrix of the second FC layer.
- b_2 is the bias vector of the second FC layer.

Output Layer:

$$y = \text{softmax}(W_o h_2 + b_o) \tag{17}$$

Where:

- W_o is the weight matrix of the output layer.
- b_o is the bias vector of the output layer.
- y is the output vector representing the probabilities of each class.

3.4.2 Loss Function

The categorical cross-entropy loss function is used for multi-class classification. It measures the difference between the predicted class probabilities and the actual class labels.

Given a true label vector t and the predicted probability vector y, the loss L is defined as:

$$L = -\sum_{i=1}^{C} t_i \log(y_i) \tag{18}$$

Where C is the number of classes, and t_i and y_i are the true label and predicted probability for class i, respectively.

3.4.3 Training the Classifier

The entire model, including the convolutional base and the newly added FC layers, is trained end-to-end on the retinal image dataset. This training process involves iterating through multiple epochs, where each epoch consists of forward passes, loss computation, gradient calculation, and parameter updates for all batches in the dataset.

4 Results and Discussion

4.1 Evaluation Parameters

Table 1. Evaluation Parameters

TP (True Positive)	"Represents the number of retinal images where the model correctly identifies the presence of diabetic retinopathy at a given stage."
TN (True Negative)	"Indicates the number of retinal images correctly classified as not having diabetic retinopathy or being in a healthy stage."
FP (False Positive)	"Represents the number of retinal images incorrectly classified as having diabetic retinopathy when it is not present."
FN (False Negative)	"Indicates the number of retinal images where diabetic retinopathy is present but the model fails to detect it or misclassifies it as a healthier stage."

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \tag{19}$$

$$Precision = \frac{TP}{TP+FP} \tag{20}$$

$$Sensitivity = \frac{TP}{TP+FN} \tag{21}$$

$$Specificity = \frac{TN}{TN+FN} \tag{22}$$

$$ErrorRate = \frac{FP+FN}{TP+TN+FP+FN} \tag{23}$$

$$False\ Positive\ Rate(FPR) = \frac{FP}{FP+TN} \tag{24}$$

$$F-Score = \frac{2TP}{2TP+FP+FN} \tag{25}$$

4.2 Results

Table 2. Performance evaluation of proposed approach using various CNN architectures

Parameters	AlexNet	InceptionV3	VGG16
Accuracy	95.37%	93.12%	94.79%
Error Rate	4.63%	6.88%	5.21%
Sensitivity	95.37%	93.12%	94.79%
Specificity	98.62%	97.79%	98.24%
Precision	95.29%	93.42%	94.68%
False Positive Rate	1.38%	2.21%	1.76%
F1-Score	95.33%	93.19%	94.73%
MCC	94.61%	90.66%	93.40%
Kappa Statistics	89.76%	84.21%	87.65%

Table 2 presents a detailed performance evaluation of three CNN architectures—AlexNet, InceptionV3, and VGG16—expressed in percentages. Accuracy measures the proportion of correctly classified instances and shows AlexNet achieving 95.37%,

InceptionV3 93.12%, and VGG16 94.79%, indicating high performance across all models. The Error Rate, which reflects the proportion of incorrect classifications, is lowest for AlexNet at 4.63% and highest for InceptionV3 at 6.88%. Sensitivity, the ability to correctly identify true positives, matches the accuracy values for each model. Specificity, the proportion of correctly identified true negatives, is highest for AlexNet at 98.62% and lowest for InceptionV3 at 97.79%. Precision, which measures the proportion of true positives among all positive predictions, is very close across the models, with AlexNet at 95.29% and VGG16 at 94.68%. The False Positive Rate, representing the proportion of negative instances incorrectly identified as positive, is lowest for AlexNet at 1.38% and highest for InceptionV3 at 2.21%. The F1-Score, which balances precision and recall, is highest for AlexNet at 95.33% and lowest for InceptionV3 at 93.19%. The Matthews Correlation Coefficient (MCC), reflecting the overall quality of the binary classification, is highest for AlexNet at 94.61% and lowest for InceptionV3 at 90.66%. Lastly, Kappa Statistics, measuring agreement between predicted and observed classifications, is highest for AlexNet at 89.76% and lowest for InceptionV3 at 84.21%. These results collectively demonstrate that while all models perform well, AlexNet generally achieves the highest metrics across most evaluation criteria.

Table 3. Comparison of the proposed approach with previous research works

Parameters	**CNN-AlexNet (Proposed)**	**KNN [20]**
Accuracy	95.37%	94.44%
Error Rate	4.63%	5.56%
Sensitivity	95.37%	94.44%
Specificity	98.62%	98.15%
Precision	95.29%	95.45%
False Positive Rate	1.38%	1.85%
F1-Score	95.33%	94.56%
MCC	94.61%	93.05%
Kappa Statistics	89.76%	85.19%

Table 3 presents a comparative evaluation of the proposed CNN-AlexNet approach against a K-Nearest Neighbors (KNN) algorithm described in reference [20]. The Accuracy of the CNN-AlexNet model is 95.37%, surpassing the KNN model's 94.44%, indicating a higher overall correctness in classification. The Error Rate for CNN-AlexNet stands at 4.63%, which is lower than the KNN's 5.56%, suggesting fewer misclassifications. Sensitivity is identical between the two models, with CNN-AlexNet at 95.37% and KNN at 94.44%, reflecting a similar ability to correctly identify true positives. In terms of Specificity, CNN-AlexNet slightly outperforms KNN with 98.62% versus 98.15%, demonstrating a better capability to correctly identify true negatives. The Precision of CNN-AlexNet is 95.29%, marginally lower than KNN's 95.45%, indicating a slightly reduced rate of false positives. The False Positive Rate is lower for CNN-AlexNet at 1.38% compared to KNN's 1.85%, pointing to fewer

incorrect positive predictions. The F1-Score, which balances precision and recall, is higher for CNN-AlexNet at 95.33% compared to KNN's 94.56%, suggesting better overall performance in balancing precision and recall. The Matthews Correlation Coefficient (MCC) for CNN-AlexNet is 94.61%, exceeding KNN's 93.05%, reflecting a superior overall classification quality. Lastly, the Kappa Statistics for CNN-AlexNet is 89.76%, higher than KNN's 85.19%, indicating a greater level of agreement between the observed and predicted classifications. Overall, the CNN-AlexNet approach shows improved performance across most metrics compared to the KNN model, demonstrating its enhanced effectiveness for diabetic retinopathy detection.

5 Conclusion

This research successfully demonstrates the efficacy of using Convolutional Neural Networks (CNNs) for the detection and classification of diabetic retinopathy (DR). By applying deep feature extraction and classification through AlexNet, InceptionV3, and VGG16 architectures, we have achieved significant advancements in identifying and differentiating between various stages of DR. Among the evaluated models, AlexNet emerged as the most effective, achieving an accuracy of 95.37%, which surpasses both InceptionV3 and VGG16. The improved performance of CNN-based methods over traditional approaches, such as k-Nearest Neighbors (KNN), highlights the potential of deep learning techniques in enhancing diagnostic precision. The analysis of metrics such as sensitivity, specificity, and F-score further corroborates the robustness of CNNs in handling the complexities of DR detection. The findings advocate for the integration of advanced deep learning models into clinical practice to facilitate early diagnosis and intervention, potentially reducing the risk of vision loss in diabetic patients. Future work may focus on optimizing these models, exploring additional CNN architectures, and incorporating larger and more diverse datasets to further enhance the accuracy and reliability of DR detection systems.

References

1. Zago, G. T., Andreão, R. V., Dorizzi, B., & Teatini Salles, E. O. (2023). Diabetic retinopathy detection using red lesion localization and convolutional neural networks. Computers in Biology and Medicine, 116, 103537. https://doi.org/10.1016/j.compbiomed.2024.103537
2. Kumar, S., Adarsh, A., Kumar, B., & Singh, A. K. (2023). An automated early diabetic retinopathy detection through improved blood vessel and optic disc segmentation. Optics & Laser Technology, 121 (September 2024), 105815. https://doi.org/10.1016/j.optlastec.2024.105815
3. Kamble, V. V, & Kokate, R. D. (2023). Automated diabetic retinopathy detection using radial basis function. Procedia Computer Science, 167(2024), 799–808. https://doi.org/10.1016/j.procs.2023.03.429
4. Hajabdollahi, M., Esfandiarpoor, R., Najarian, K., Karimi, N., Samavi, S., & Reza Soroushmehr, S. M. (2024). Hierarchical Pruning for Simplification of Convolutional Neural Networks in Diabetic Retinopathy Classification. 2024 41st Annual International

Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), 970–973. https://doi.org/10.1109/EMBC.2024.8857769

5. Qomariah, D. U. N., Tjandrasa, H., & Fatichah, C. (2024). Classification of Diabetic Retinopathy and Normal Retinal Images using CNN and SVM. 2024 12th International Conference on Information & Communication Technology and System (ICTS), 152–157. https://doi.org/10.1109/ICTS.2024.8850940
6. de la Torre, J., Valls, A., & Puig, D. (2023). A deep learning interpretable classifier for diabetic retinopathy disease grading. Neurocomputing, 396(xxxx), 465–476. https://doi.org/10.1016/j.neucom.2023.07.102
7. Washburn, P. S., Mahendran, Dhanasekharan, Periyasamy, & Murugeswari. (2023). Investigation of severity level of diabetic retinopathy using adaboost classifier algorithm. Materials Today: Proceedings, 33(xxxx), 3037–3042. https://doi.org/10.1016/j.matpr.2023.03.199
8. Saman, G., Gohar, N., Noor, S., Shahnaz, A., Idress, S., Jehan, N., Rashid, R., & Khattak, S. S. (2023). Automatic detection and severity classification of diabetic retinopathy. Multimedia Tools and Applications, 79(43–44), 31803–31817. https://doi.org/10.1007/s11042-020-09118-8
9. Abdelsalam, M. M. (2023). Effective blood vessels reconstruction methodology for early detection and classification of diabetic retinopathy using OCTA images by artificial neural network. Informatics in Medicine Unlocked, 20, 100390. https://doi.org/10.1016/j.imu.2023.100390
10. Dutta, M. K., Parthasarathi, M., Ganguly, S., Ganguly, S., & Srivastava, K. (2017). An efficient image processing based technique for comprehensive detection and grading of nonproliferative diabetic retinopathy from fundus images. Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization, 5(3), 195–207.
11. Birajdar, U., Gadhave, S., Chikodikar, S., Dadhich, S., & Chiwhane, S. (2023). Detection and Classification of Diabetic Retinopathy Using AlexNet Architecture of Convolutional Neural Networks. In Proceeding of International Conference on Computational Science and Applications (pp. 245–253). https://doi.org/10.1007/978-981-15-0790-8_25
12. Saranya, P., & Prabakaran, S. (2023). Automatic detection of non-proliferative diabetic retinopathy in retinal fundus images using convolution neural network. Journal of Ambient Intelligence and Humanized Computing, 0123456789. https://doi.org/10.1007/s12652-020-02518-6
13. Galdran, A., Chelbi, J., Kobi, R., Dolz, J., Lombaert, H., ben Ayed, I., & Chakor, H. (2023). Non- uniform Label Smoothing for Diabetic Retinopathy Grading from Retinal Fundus Images with Deep Neural Networks. Translational Vision Science & Technology, 9(2), 34. https://doi.org/10.1167/tvst.9.2.34
14. Martinez-Murcia, F. J., Ortiz, A., Ramírez, J., Górriz, J. M., & Cruz, R. (2023). Deep residual transfer learning for automatic diagnosis and grading of diabetic retinopathy. Neurocomputing. https://doi.org/10.1016/j.neucom.2023.04.148
15. Samanta, A., Saha, A., Satapathy, S. C., Fernandes, S. L., & Zhang, Y.-D. (2023). Automated detection of diabetic retinopathy using convolutional neural networks on a small dataset. Pattern Recognition Letters, 135, 293–298. https://doi.org/10.1016/j.patrec.2023.04.026
16. Shanthi, T., & Sabeenian, R. S. (2024). Modified Alexnet architecture for classification of diabetic retinopathy images. Computers & Electrical Engineering, 76, 56–64. https://doi.org/10.1016/j.compeleceng.2024.03.004
17. Bourouis, S., Zaguia, A., & Bouguila, N. (2023). Hybrid Statistical Framework for Diabetic Retinopathy Detection. In A. Campilho, F. Karray, & B. ter Haar Romeny (Eds.),

International Conference Image Analysis and Recognition (pp. 687–694). Springer International Publishing. https://doi.org/10.1007/978-3-319-93000-8_78
18. Sarwinda, D., Siswantining, T., & Bustamam, A. (2023). Classification of Diabetic Retinopathy Stages using Histogram of Oriented Gradients and Shallow Learning. 2023 International Conference on Computer, Control, Informatics and Its Applications (IC3INA), 83–87. https://doi.org/10.1109/IC3INA.2023.8629502
19. Issac, A., Dutta, M. K., & Travieso, C. M. (2023). Automatic computer vision-based detection and quantitative analysis of indicative parameters for grading of diabetic retinopathy. Neural Computing and Applications, 32(20), 15687–15697. https://doi.org/10.1007/s00521-018-3443-z
20. Rehman, A., Harouni, M., Karimi, M., Saba, T., Bahaj, S.A. and Awan, M.J., 2022. Microscopic retinal blood vessels detection and segmentation using support vector machine and K-nearest neighbors. Microscopy research and technique, 85(5), pp.1899-1914.