© Published by Transactions on
Mathematical Sciences and Computational Engineering.
The Tribhuj Publishers
For permissions, please contact:
tmsce.editorial@gmail.com

Volume 1, Issue I, October 2025
Pages 67-77

Deep Learning–Based Comparative Strategies for Enhanced Diabetic Retinopathy Detection

Rakesh Kumar
Department of Computer Science & Engineering
Siksha 'O' Anusandhan (Deemed to be University)
Bhubaneshwar-751030, India
rakeshkumar@soa.ac.in

Neelam Singha
Centre for Data Science
Siksha 'O' Anusandhan (Deemed to be University)
Bhubaneshwar-751030, India
neelamsingha@soa.ac.in

Abstract

Diabetic Retinopathy (DR) is a progressive ocular disorder associated with diabetes and remains a leading cause of preventable vision loss worldwide. Early and accurate identification of DR is essential to reduce the risk of irreversible blindness. In this work, we present a comprehensive deep learning–based framework for automated DR detection using multiple Convolutional Neural Network (CNN) architectures. Eight state-of-the-art models—ResNet50, DenseNet121, VGG16, AlexNet, InceptionV3, InceptionResNetV2, Xception, and EfficientNetB0—were systematically trained and evaluated on a retinal fundus image dataset annotated according to DR severity stages. To improve image quality and model robustness, Gaussian filtering was applied during preprocessing to suppress noise and enhance discriminative features. Furthermore, transfer learning was adopted to exploit knowledge from pre-trained networks, enabling faster convergence and improved predictive performance. A dynamic learning rate strategy was also employed to ensure stable and efficient optimization during training. The proposed study is motivated by the growing demand for reliable and scalable DR screening systems, particularly in regions where access to ophthalmic specialists is limited. Automated deep learning solutions offer a viable alternative to manual examination, which is time-consuming and prone to inter-observer variability. Experimental results demonstrate that Xception and EfficientNetB0 outperform other evaluated architectures, achieving test accuracies above 80%. EfficientNetB0 delivers superior overall detection performance, while Xception shows enhanced capability in discriminating between multiple DR severity levels. The findings highlight the effectiveness of advanced CNN architectures for automated DR assessment and provide valuable insights into their comparative strengths. This study supports the integration of deep learning models into clinical screening workflows to facilitate early diagnosis and timely intervention for diabetic retinopathy.

Keywords: Diabetic Retinopathy, Deep Learning, Convolutional Neural Networks, Transfer Learning, Gaussian Filtering.

1 Introduction

Diabetic Retinopathy (DR) is a vision-threatening ocular disorder caused by prolonged diabetes (see Fig. 1) and remains one of the predominant causes of blindness among working-age adults worldwide. Despite being largely preventable through early diagnosis and timely treatment, effective screening for DR continues to pose significant challenges. Conventional diagnostic procedures rely heavily on expensive imaging equipment, skilled ophthalmologists, and time-intensive manual evaluation. These requirements make large-scale screening impractical, particularly in low-resource settings, rural regions, and developing countries where access to specialized eye care is limited. As a result, many patients are diagnosed at advanced stages, when irreversible vision damage has already occurred.

Earlier attempts to automate DR detection using traditional machine learning approaches have shown limited success, primarily due to their dependence on handcrafted features and their inability to reliably capture subtle and complex pathological patterns present in retinal fundus images. The increasing availability of large-scale medical imaging data and advances in computational power have positioned deep learning as a powerful alternative. In particular, Convolutional Neural Networks (CNNs) have demonstrated remarkable capability in learning hierarchical visual features directly from raw images, making them well suited for medical image analysis tasks such as DR detection. There is therefore a critical need for an automated, accurate, and efficient DR screening system capable of identifying the disease at its earliest stages, thereby reducing preventable vision loss and easing the burden on healthcare systems.

The primary goal of this study is to design and evaluate a robust deep learning–based framework for automated diabetic retinopathy detection. To achieve this, multiple state-of-the-art CNN architectures—including ResNet, DenseNet, Inception, EfficientNet, and related variants—are systematically investigated to assess their effectiveness in identifying DR from retinal fundus images. Extensive experimental analysis is conducted through careful hyperparameter optimization, along with the application of data augmentation techniques to enhance model generalization and resilience to variations in image quality. The proposed models are evaluated using comprehensive performance metrics such as accuracy, precision, recall, F1-score, and computational efficiency. Furthermore, the most promising model(s) are validated on an independent real-world dataset to assess their suitability for deployment in clinical screening environments. Through this approach, the study aims to deliver a scalable, reliable, and accessible solution for early DR detection across both well-resourced and underserved populations.

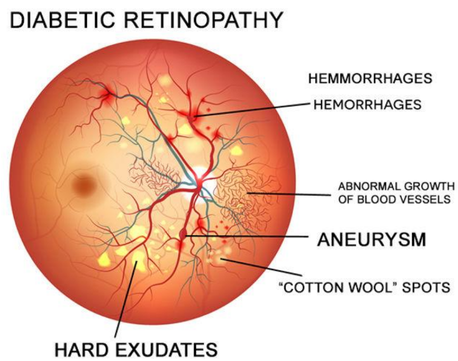

Figure 1: Diabetic Retinopathy Indicators

This work makes several notable contributions to the field of automated diabetic retinopathy detection:

- A Gaussian filtering–based (see Fig. 2) preprocessing strategy is introduced to suppress noise and enhance clinically relevant features in retinal images, thereby improving detection performance.
- Transfer learning is leveraged by utilizing pre-trained CNN models such as VGG, ResNet, and Inception, enabling faster training and improved accuracy compared to training from scratch.
- A dynamic learning rate scheduling mechanism is employed to adaptively adjust optimization parameters during training, leading to improved convergence behavior and enhanced predictive accuracy.

By integrating these techniques, the proposed framework offers a novel and efficient approach for DR detection, with strong potential to support early diagnosis and improve patient outcomes.

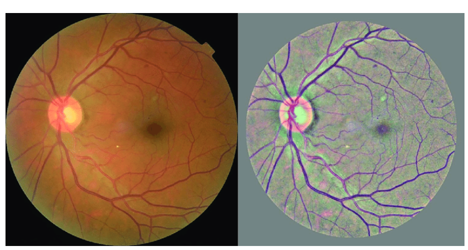

Figure 2: Gaussian Filtered Retinal Image

Recent advances in Artificial Intelligence (AI), particularly deep learning (DL), have significantly transformed the landscape of diabetic retinopathy screening and severity grading. Contemporary research has demonstrated that DL-based models trained on fundus images can achieve performance comparable to, and in some cases exceeding, that of human experts. Numerous studies have reported high accuracy in binary and multi-class DR classification tasks using CNN architectures such as DenseNet, ResNet50, InceptionV3, AlexNet, and VGG16, consistently outperforming conventional machine learning methods. DenseNet-based models, for example, have shown strong performance across multiple benchmark datasets, while ResNet architectures have demonstrated robustness through residual learning mechanisms. Inception-based networks have achieved high sensitivity and specificity, particularly when combined with transfer learning strategies.

More recent investigations [3] have explored advanced architectures such as InceptionResNetV2 and Xception, as well as hybrid and attention-based models, achieving improved performance in DR severity grading tasks. Despite these advancements, several challenges remain unresolved, including limited dataset diversity, reduced generalizability across clinical settings, and a lack of interpretability in deep learning predictions. Addressing these issues is essential for translating AI-driven DR detection systems into real-world clinical practice. Consequently, ongoing research must focus on developing more robust, explainable, and clinically validated models capable of operating reliably under diverse imaging conditions.

2 Methodology

This section describes the proposed deep learning–based framework for automated diabetic retinopathy (DR) detection. The methodology comprises image preprocessing, model architecture selection, transfer learning, training optimization, and performance enhancement techniques designed to ensure robustness and high classification accuracy.

- Image Preprocessing: Retinal fundus images are preprocessed to enhance visual quality and suppress noise that may interfere with feature learning. A Gaussian smoothing filter is applied to each image to reduce high-frequency distortions while preserving critical structural details such as blood vessels and lesion boundaries. This preprocessing step improves the visibility of diagnostically relevant patterns and supports more effective feature extraction by deep learning models.

- Deep Learning Architecture: A variety of deep learning paradigms are examined to identify the most suitable approach for DR detection. Although recurrent and transformer-based models are considered, Convolutional Neural Networks (CNNs) form the foundation of the proposed framework due to their superior performance in image analysis tasks. The CNN architectures consist of convolutional layers for hierarchical feature extraction, pooling layers for spatial reduction, and nonlinear activation functions to model complex visual relationships. This design enables the networks to learn both local and global retinal characteristics.

- Transfer Learning: To improve training efficiency and classification performance, transfer learning is employed using pre-trained models such as VGG, ResNet, and Inception. The learned weights from large-scale image datasets are utilized as initialization for the convolutional layers, while the final fully connected layers are restructured and trained for DR-specific classification. This strategy significantly reduces training time and enhances model generalization, particularly when available labeled medical data are limited.

- Hyperparameter Optimization: Model performance is optimized through systematic tuning of key hyperparameters, including learning rate, batch size, and the number of training epochs. The Adam optimization algorithm is selected due to its adaptive learning capability and stable convergence behavior across different network architectures.

- Adaptive Learning Rate Scheduling: A dynamic learning rate scheduler is integrated into the training process to improve convergence and prevent instability. The learning rate is automatically adjusted based on training progression, allowing larger updates during early training stages and finer adjustments as the model approaches optimal weight configurations. This approach enhances training efficiency and minimizes the risk of overshooting optimal solutions.

- Loss Function: For multi-class diabetic retinopathy classification, categorical cross-entropy is adopted as the loss function. This loss formulation effectively quantifies the discrepancy between predicted probability distributions and ground-truth labels, enabling accurate discrimination among multiple DR severity levels.

- Data Augmentation: To address dataset limitations and class imbalance, data augmentation techniques are applied during training. These include random rotations, horizontal and vertical flips, and brightness variations. Augmentation increases data diversity, improves model robustness to real-world variations, and reduces overfitting, thereby enhancing generalization performance.

To determine the most suitable deep learning architecture for automated diabetic retinopathy (DR) detection, this study conducts a comprehensive evaluation of multiple established and advanced neural network models. The selected architectures represent diverse design philosophies, enabling a fair comparison of their feature learning capacity, computational efficiency, and classification performance.

1. ResNet50: ResNet50 is a deep residual network designed to address degradation and vanishing gradient issues through identity-based skip connections. In this work, ResNet50 is enhanced with a Convolutional Block Attention Module (CBAM), which introduces channel-wise and spatial attention mechanisms to emphasize diagnostically relevant retinal features and improve discriminative learning.

2. DenseNet121: DenseNet121 employs dense connectivity, where each layer receives inputs from all preceding layers. This structure encourages feature reuse, strengthens gradient propagation, and supports the learning of fine-grained representations essential for identifying subtle DR-related patterns in retinal images.

3. VGG16: VGG16 is a classical deep CNN characterized by its uniform architecture composed of stacked convolutional layers with small receptive fields. Owing to its simplicity and consistent performance in image classification tasks, VGG16 is adopted as a benchmark model for comparative analysis.

4. AlexNet: AlexNet is one of the earliest deep convolutional architectures that demonstrated the feasibility of large-scale image recognition using deep learning. Despite its relatively shallow structure compared to modern networks, AlexNet is included to provide historical context and serve as a baseline for evaluating architectural advancements.

5. InceptionV3: InceptionV3 incorporates inception modules that process input features through parallel convolutional paths with varying kernel sizes. This multi-scale feature extraction strategy enables the network to capture diverse spatial information while maintaining computational efficiency, making it suitable for complex visual recognition tasks.

6. InceptionResNetV2: InceptionResNetV2 integrates the multi-scale processing capability of Inception modules with residual learning from ResNet architectures. The inclusion of residual connections improves training stability and accelerates convergence in this deep hybrid network.

7. Xception: Xception extends the Inception concept by replacing standard convolutions with depthwise separable convolutions, significantly reducing computational complexity while preserving representational power. Its streamlined architecture has demonstrated strong performance in medical image classification tasks.

8. EfficientNetB0: EfficientNetB0 is a compact and scalable model that employs Mobile Inverted Bottleneck Convolution (MBConv) blocks combined with compound scaling principles. It achieves an effective balance between accuracy and computational cost, making it well suited for deployment in resource-constrained environments.

All experiments were conducted using a hybrid computational setup to ensure efficiency and reproducibility. Model training and evaluation were primarily performed on Google Colaboratory Pro, utilizing an NVIDIA Tesla P100 GPU with 32 GB RAM. Supplementary preprocessing and experimental analyses were carried out on a local workstation equipped with an Intel Core i5-1240P processor, 16 GB RAM, and a 512 GB NVMe SSD running Windows 11 Pro. The software environment included TensorFlow 2.10, CUDA 11.2, cuDNN 8.1, and Python 3.10 within the Anaconda distribution. Model development was implemented using the Keras API, with supporting libraries such as NumPy, Pandas, Matplotlib, and Seaborn employed for data processing, evaluation, and visualization.

2.1 Proposed Algorithm

The proposed framework (see Fig. 3) for automated diabetic retinopathy (DR) detection follows a structured pipeline consisting of dataset analysis, image enhancement, data enrichment, model training, and performance assessment. Each stage is designed to ensure reliable learning, robust generalization, and accurate classification of DR severity.

The process begins with an initial examination of the retinal image dataset to assess class distribution and identify potential imbalance among DR categories. Representative fundus images from each class are visually inspected to capture characteristic patterns associated with both healthy and diseased retinas. This exploratory analysis provides essential insights that inform preprocessing and model selection strategies.

In the preprocessing stage, a Gaussian smoothing operation is applied to each image to suppress noise while preserving important retinal structures, such as blood vessels and lesion boundaries. All images are subsequently resized to a uniform spatial resolution of 224×224 pixels to ensure compatibility with pre-trained deep learning architectures. The dataset is then partitioned into training, validation, and testing subsets using a 70%–20%–10% split, respectively, enabling effective model optimization and unbiased performance evaluation.

To improve model robustness and mitigate overfitting, data augmentation techniques are applied exclusively to the training set. These include random horizontal flipping to simulate orientation variability and random brightness modulation to account for differences in illumination conditions commonly observed in fundus imaging. Augmentation increases sample diversity and enhances the model's ability to generalize to unseen data.

The core learning phase utilizes a deep Convolutional Neural Network (CNN) enhanced through transfer learning. Pre-trained architectures such as VGG, ResNet, and Inception are employed as feature extractors, with their final classification layers modified to accommodate DR-specific categories. Fine-tuning is performed to adapt the learned representations to retinal imaging characteristics. Model optimization is carried out using the Adam optimizer in conjunction with an adaptive learning rate scheduler, enabling stable convergence and efficient training. For multi-class DR grading, categorical cross-entropy is adopted as the objective function.

Model evaluation is conducted on the independent test set using a confusion matrix to quantify classification outcomes, including true positives, true negatives, false positives, and false negatives. Standard performance metrics derived from these values are used to assess overall effectiveness. The proposed method is further compared against existing approaches using the same dataset to establish relative performance gains. The optimal model is selected based on its superior accuracy and consistent performance across all evaluation metrics.

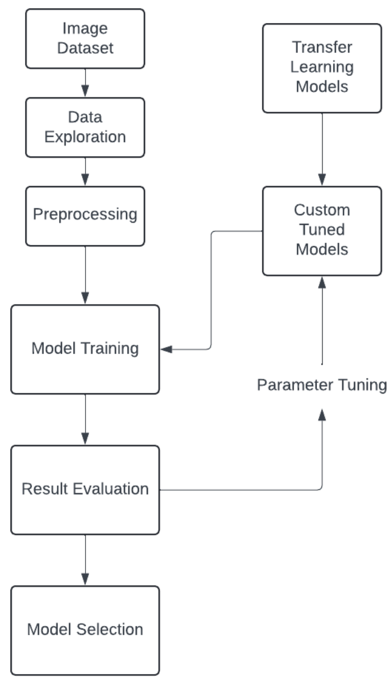

Figure 3: Work Flow Diagram

3 Experimentation and Model Evaluation

3.1 Depiction Results

The performance of all evaluated deep learning architectures was systematically analyzed to assess their suitability for automated diabetic retinopathy (DR) detection. Confusion matrices were generated for each model to provide a detailed breakdown of classification outcomes, including true positives, true negatives, false positives, and false negatives. In addition, training and validation curves were plotted over successive epochs, illustrating metrics such as accuracy, loss, cosine similarity, and Euclidean distance. These visualizations offered insights into the models' learning dynamics, helping to identify signs of overfitting or underfitting and to monitor convergence behavior. By examining these trends, appropriate adjustments to hyperparameters and training strategies were implemented to enhance overall performance and stability.

3.2 Validation

Model validation was conducted using an independent subset of retinal images not exposed during training. This step ensured that the trained networks could generalize to unseen data and provided a realistic assessment of their predictive reliability. Standard evaluation metrics including accuracy, precision, recall, F1-score, and area under the ROC curve were computed for each architecture to facilitate a comprehensive comparison. Performance across different DR severity levels was also examined to evaluate the models' sensitivity to subtle pathological variations.

Model Name	Training Accuracy	Validation Accuracy	Cosine Similarity	Validation Cosine Similarity	Euclidean Distance	Validation Euclidean Distance
ResNet50	84.5	54.7	0.76	0.55	3.388	4.485
DenseNet-121	87.4	74.3	0.78	0.68	3.264	4.049
VGG16	79.6	65.1	0.71	0.70	3.791	3.660
AlexNet	70.6	63.8	0.69	0.68	3.937	3.777
Inception-V3	79.4	72.6	0.80	0.73	3.166	3.882
InceptionRes-NetV2	89.4	78.1	0.82	0.77	2.890	3.331
Xception	97.6	82.2	0.89	0.81	2.268	3.125
EfficientNet-B0	92.4	83.1	0.85	0.80	2.7387	3.129

3.3 Discussion Contribution

The comparative analysis revealed notable differences in performance among the evaluated CNN architectures. Xception and EfficientNetB0 consistently demonstrated superior results, achieving high classification accuracy (83%) and strong generalization capabilities. Their architectural designs enabled effective extraction of fine-grained features, which are critical for distinguishing subtle DR-related retinal patterns. In contrast, deeper networks such as ResNet50 and DenseNet121 exhibited overfitting tendencies, particularly in this multi-class classification scenario, indicating sensitivity to training data distribution. Shallow models, including VGG16 and AlexNet, struggled to capture complex retinal features due to limited representational capacity. Inception-based networks achieved moderate performance; however, the absence of residual connections limited their ability to efficiently propagate information through deeper layers.

A closer examination of class-level performance highlighted that Xception excelled in DR severity grading, whereas EfficientNetB0 delivered higher overall detection accuracy across all categories. These findings underscore the critical role of architectural depth, connectivity patterns, and feature extraction mechanisms in achieving robust DR detection. The study provides a detailed comparative understanding of the strengths and limitations of diverse CNN models, offering guidance for selecting and designing AI-driven frameworks for early DR diagnosis and automated screening systems.

4 Conclusion

In this work, we conducted a comprehensive evaluation of multiple Convolutional Neural Network (CNN) architectures for automated diabetic retinopathy (DR) detection. Our comparative analysis demonstrated that Xception and EfficientNetB0 consistently outperformed other models, owing to their ability to effectively capture subtle and complex features in retinal fundus images. These results underscore the potential of advanced deep learning models to improve the accuracy and reliability of AI-driven DR screening systems.

For future investigations, integrating model ensembling strategies could be beneficial, combining the high overall detection accuracy of EfficientNetB0 with the superior DR severity discrimination capabilities of Xception. Moreover, expanding the dataset to include larger, balanced, and more diverse retinal images, as well as incorporating relevant clinical metadata, could further enhance model robustness and generalizability across varied populations. Ultimately, the deployment of these optimized models in clinical workflows could facilitate early detection of DR, improve patient outcomes, and contribute to more accessible and effective eye care, particularly in resource-limited and underserved regions.

Acknowledgment

The authors convey their sincere thanks to the Siksha 'O' Anusandhan (deemed to be) University, Bhubaneswar, India, for the continuous support and providing the best research facility.

References

[1] G. Eason, B. Noble, and I. N. Sneddon, "On certain integrals of Lipschitz-Hankel type involving products of Bessel functions," Phil. Trans. Roy. Soc. London, vol. A247, pp. 529–551, April 1955.

[2] J. Clerk Maxwell, A Treatise on Electricity and Magnetism, 3rd ed., vol. 2. Oxford: Clarendon, 1892, pp.68–73.

[3] I. S. Jacobs and C. P. Bean, "Fine particles, thin films and exchange anisotropy," in Magnetism, vol. III, G. T. Rado and H. Suhl, Eds. New York: Academic, 1963, pp. 271–350.

[4] K. Elissa, "Title of paper if known," unpublished.

[5] R. Nicole, "Title of paper with only first word capitalized," J. Name Stand. Abbrev., in press.

[6] Y. Yorozu, M. Hirano, K. Oka, and Y. Tagawa, "Electron spectroscopy studies on magneto-optical media and plastic substrate interface," IEEE Transl. J. Magn. Japan, vol. 2, pp. 740–741, August 1987 [Digests 9th Annual Conf. Magnetics Japan, p. 301, 1982].

[7] M. Young, The Technical Writer's Handbook. Mill Valley, CA: University Science, 1989.