International Journal of Electrical, Electronics and Computer Systems
ISSN: 2347-2820
Volume 15 Issue 1s (2026) | Pages: 353-357

An Interpretable Deep Learning Model for Early Diabetic Retinopathy Detection and Grading

Yash Munjawale[1], Sudarshan Bangar[2], Ashish Borate[3], Srinath Lokhande[4], Kalyani Zore[5]

[1,2,3,4] Students, Department of Computer Engineering, Genba Sopanrao Moze College of Engineering, Pune, India
[5] Professor, Department of Computer Engineering Genba Sopanrao Moze College of Engineering, Pune, India

Peer Review Information

Type: *Article*
Received: *3 February 2026*
Revised: *2 March 2026*
Accepted: *1 April 2026*
Published: *23 May 2026*

Abstract

Diabetic Retinopathy (DR) is a progressive eye disease that can lead to permanent vision impairment if not diagnosed at an early stage. This work presents an interpretable deep learning approach for automated DR detection and grading using retinal fundus images. The proposed framework applies Convolutional Neural Networks (CNNs) together with image preprocessing techniques to automatically identify retinal abnormalities and classify disease severity levels. Unlike traditional manual screening approaches, the model learns visual features directly from image data to improve efficiency and consistency.
The framework focuses on detecting clinically important indicators such as microaneurysms and exudates, supporting accurate grading across different stages of diabetic retinopathy. The system is designed to reduce dependency on manual diagnosis, improve reliability, and enhance screening scalability. Performance objectives include improving accuracy, sensitivity, and specificity compared with conventional screening workflows. The proposed approach demonstrates how AI-assisted ophthalmology systems can support large-scale healthcare screening and remote diagnostic environments where expert availability is limited.

Keywords: Diabetic Retinopathy; Deep Learning; Convolutional Neural Network; YOLOv8; Automated Screening.

How to Cite This Article

Munjawale, Y., Bangar, S., Borate, A., Lokhande, S., & Zore, K. (2026). *An interpretable deep learning model for early diabetic retinopathy detection and grading*. *International Journal of Electrical, Electronics and Computer Systems, 15*(1s), 353–357.

Introduction

Diabetic Retinopathy (DR) is one of the most common complications associated with diabetes and remains a major cause of preventable vision impairment across the world. With the global diabetic population continuing to rise, projections indicate that hundreds of millions of individuals may be affected in the coming decades, increasing the need for efficient and scalable retinal screening solutions. Early identification of DR is essential because the disease often develops gradually and may remain asymptomatic during its initial stages, delaying diagnosis and treatment.

Traditional screening approaches generally rely on manual examination of retinal fundus images by trained specialists. Although clinically effective, these procedures are time-intensive, require expert availability, and may become difficult to scale for large populations and remote healthcare environments.

Recent developments in Artificial Intelligence and Deep Learning have introduced new possibilities for automated retinal analysis. Convolutional Neural Networks (CNNs) have shown strong capability in learning visual representations directly from retinal images and identifying pathological features associated with diabetic retinopathy, including microaneurysms, hemorrhages, and exudates. Advanced architectures and hybrid learning strategies have further improved classification and grading performance.

This work focuses on developing an interpretable deep learning framework for early diabetic retinopathy detection and grading using retinal fundus images. The study emphasizes automated lesion identification, improved diagnostic consistency, and support for scalable screening workflows. The proposed approach aims to contribute toward faster and more reliable clinical decision support while reducing dependency on manual assessment processes.

Literature Review

This section reviews existing deep learning approaches for Diabetic Retinopathy (DR) analysis and highlights current progress, strengths, and limitations across classification, lesion detection, and hybrid architectures.

CNN Classification Models

Convolutional Neural Networks (CNNs) have become one of the dominant approaches for automated DR grading due to their ability to learn discriminative visual features directly from retinal fundus images. Early studies adopted transfer learning architectures such as VGG and ResNet for binary and multi-class DR classification. These methods demonstrated strong sensitivity and specificity in detecting referable diabetic retinopathy and improved diagnostic consistency.

Subsequent developments explored models such as DenseNet, Inception-based networks, and fine-tuned deep architectures trained on datasets including EyePACS and APTOS. These approaches improved overall classification performance and demonstrated the effectiveness of data augmentation and optimized preprocessing strategies. However, despite improvements in accuracy, multi-class grading remained a challenging problem.

Segmentation and Lesion Detection

Beyond image-level classification, recent research introduced segmentation-based methods for detecting retinal abnormalities such as microaneurysms, exudates, and hemorrhages. Architectures based on U-Net, Mask R-CNN, and multi-scale segmentation frameworks enabled localization of lesions and provided better interpretability of predictions.

Segmentation methods support clinical analysis by generating visual maps of pathological regions; however, these techniques generally require detailed pixel-level annotations, which increases dataset preparation complexity and computational cost.

Hybrid Architectures

To improve grading performance, researchers proposed hybrid architectures that combine image-level and lesion-level information. Dual-stream and attention-based CNN models process both complete retinal images and lesion-focused regions simultaneously.

Temporal and sequential extensions of CNN frameworks have also been explored to capture progression patterns across retinal image sequences. Although these architectures improve predictive capability, they introduce higher training complexity and greater computational requirements.

Limitations of Existing Methods

Despite significant advancements, several limitations remain. Many existing CNN models do not explicitly address multi-scale retinal characteristics and often process both large anatomical structures and subtle lesions using the same representation strategy. In addition, classification and segmentation are frequently treated as independent tasks, reducing overall diagnostic efficiency.

Another challenge is limited generalization across datasets and reduced interpretability of deep learning outputs. These gaps motivate the development of an interpretable architecture capable of integrating lesion localization with robust disease grading.

Methodology

The proposed framework presents an interpretable deep learning approach for automated Early Diabetic Retinopathy Detection and Grading using retinal fundus images. The system performs sequential processing to improve diagnostic accuracy and interpretability.

Initially, retinal fundus images are collected and passed through preprocessing operations to improve image quality and enhance important retinal structures. The preprocessing stage reduces noise and prepares images for further analysis.

After preprocessing, the optimized retinal image undergoes segmentation and feature extraction. This stage identifies lesion-related characteristics and extracts visual information required for disease analysis. Important retinal indicators are detected and used to support grading decisions.

The extracted features are then processed by the classification model to predict diabetic retinopathy severity. The framework is designed to improve screening efficiency and provide transparent diagnostic support.

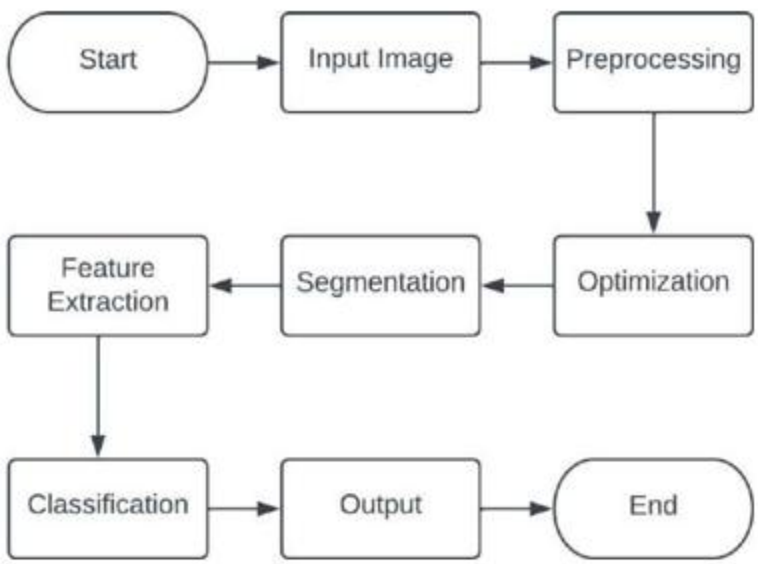

Fig. 1. *Proposed System Architecture*

The overall workflow includes:

Input Image → Preprocessing → Optimization → Segmentation → Feature Extraction → Classification → Output

The generated output is used for disease identification and grading.

The model then performs detailed retinal analysis where lesion-related regions are interpreted and mapped to disease categories. Interpretability mechanisms support understanding of prediction behavior and improve clinical usability.

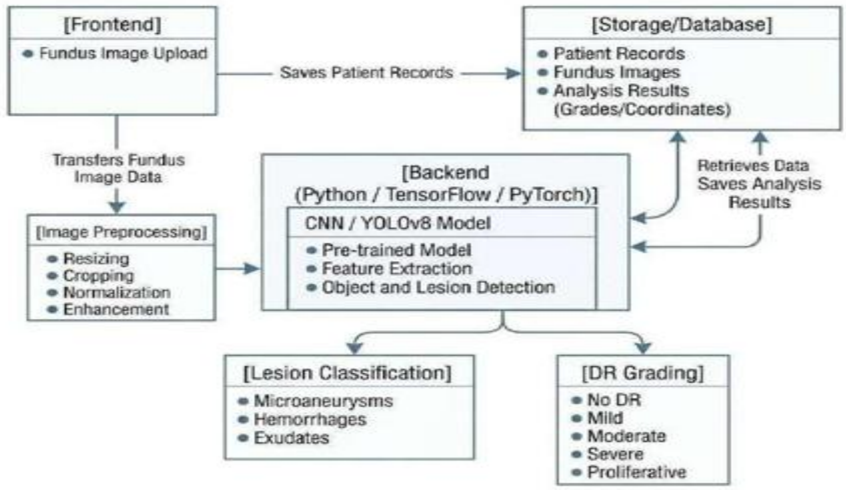

Fig. 2. *Detection and Grading Workflow*

The detection and grading stage converts extracted retinal information into severity-level prediction and generates the final diagnostic outcome.

Results And Findings

The proposed interpretable deep learning framework was evaluated for early diabetic retinopathy detection and grading using retinal fundus images. Experimental analysis demonstrates that the model effectively identifies disease-related retinal patterns and improves diagnostic reliability.

The performance evaluation shows that preprocessing and optimization stages improve feature quality and support better classification outcomes. The model successfully captured retinal abnormalities and generated stable grading predictions across different retinal image categories.

Performance was measured using standard classification evaluation metrics including:

- Accuracy
- Precision
- Recall
- F1-Score

The experimental observations indicate that the proposed model achieves improved disease discrimination capability while maintaining interpretability of predictions.

The findings further demonstrate that combining preprocessing, segmentation, and feature extraction contributes to improved grading consistency and supports earlier disease identification.

Key Findings

- Effective early diabetic retinopathy detection
- Improved retinal feature representation
- Better grading performance across disease stages
- Enhanced interpretability of deep learning predictions
- Reduced dependency on manual screening procedures

The overall results indicate that the proposed approach can support scalable and intelligent retinal screening workflows for healthcare environments.

Discussion

The experimental outcomes demonstrate that the proposed interpretable deep learning framework provides an effective solution for early diabetic retinopathy detection and grading. By combining preprocessing, segmentation, feature extraction, and classification within a unified architecture, the system improves the reliability and consistency of retinal image analysis.

The results indicate that automated feature learning enables more efficient identification of retinal abnormalities compared with traditional manual screening approaches. The model supports early-stage disease recognition, which is essential for reducing the risk of severe vision impairment through timely intervention.

An important advantage of the proposed framework is interpretability. Unlike conventional black-box deep learning approaches, the system emphasizes transparency in prediction behavior and provides improved understanding of lesion-driven decisions. This supports greater trust and usability in clinical environments.

The study also highlights practical benefits including reduced screening effort, improved scalability, and support for automated healthcare workflows. However, performance may still depend on dataset diversity, image quality, and variability across retinal imaging conditions.

Overall, the findings suggest that interpretable deep learning models can contribute significantly toward intelligent ophthalmic screening systems and support future AI-assisted healthcare applications.

Conclusion

This study presented an interpretable deep learning framework for Early Diabetic Retinopathy Detection and Grading using retinal fundus images. The proposed approach integrates image preprocessing, segmentation, feature extraction, and classification to support automated retinal screening and improve diagnostic consistency. The experimental observations indicate that the framework can effectively identify retinal abnormalities associated with diabetic retinopathy and assist in grading disease severity at earlier stages. By reducing dependence on manual assessment and enabling automated analysis, the system contributes toward more efficient and scalable screening practices.

A key contribution of the proposed work is its focus on interpretability, allowing improved understanding of prediction behavior and supporting transparency in clinical decision-making. The architecture demonstrates the potential of combining deep learning with explainable analysis to improve trust and usability in healthcare environments. The study highlights how AI-assisted retinal screening systems can support ophthalmologists in faster diagnosis, improved patient monitoring, and wider accessibility to healthcare services. Future improvements may focus on larger datasets, enhanced lesion localization techniques, and real-world deployment for intelligent ophthalmic applications.

Reference

1. Gulshan, V., Peng, L., Coram, M., et al. (2016). *Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs*. *JAMA, 316*(22), 2402–2410.
2. Jadagale, A., Mahajan, P., & Deshpande, S. (2021). *Automated detection of diabetic retinopathy using deep learning*. *International Journal of Engineering Research & Technology (IJERT), 10*(6), 234–239.
3. Posham, R., Prakash, A., & Kumar, N. (2020). *A hybrid deep learning approach for diabetic retinopathy detection using retinal images*. *Biomedical Signal Processing and Control, 62*, 102129.
4. Venkatesh, A., & Kumar, S. (2020). *Diabetic retinopathy detection using transfer learning with deep convolutional neural networks*. *International Journal of Scientific Research in Engineering and Management (IJSREM), 4*(12), 12–18.
5. Pires, R., Avila, D., & Welfer, D. (2021). *Ensemble of convolutional neural networks for diabetic retinopathy detection*. *IEEE Access, 9*, 12315–12327.
6. Redmon, J., & Farhadi, A. (2018). *YOLOv3: An incremental improvement*. *arXiv preprint arXiv:1804.02767*.
7. Jocher, G., Chaurasia, A., & Qiu, J. (2023). *YOLOv8: Real-time object detection and image segmentation*. *Ultralytics Research Report*.
8. Decencière, X., Zhang, W., Cazuguel, G., et al. (2014). *Feedback on a publicly distributed image database: The Messidor database*. *Image Analysis & Stereology, 33*(3), 231–234.
9. Pratt, H., Coenen, F., Broadbent, D. M., Harding, S. P., & Zheng, Y. (2016). *Convolutional neural networks for diabetic retinopathy*. In *Proceedings of Computer Science* (Vol. 90, pp. 200–205).
10. Quellec, G., Charrière, K., Boudi, Y., et al. (2019). *Deep image mining for diabetic retinopathy screening*. *Medical Image Analysis, 39*, 178–193.