International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 09 Issue: 06 | June - 2025 SJIF Rating: 8.586 ISSN: 2582-3930

AI-Based Diabetic Retinopathy Detection using Explainable Deep Learning Models

Geethanjali S K[1], Rupa Kumari[2], Shaheen Naaz[3], Varshitha T L[4].

Saswati Behera5, Dr Krishna Kumar P R[6]

1,2,3,4 Students Department of CSE, SEACET, Bengaluru-49, India

5,6 Faculty Department of CSE, SEACET, Bengaluru-49, India

Abstract

Diabetic Retinopathy (DR) is a critical complication of diabetes, often leading to vision loss if not diagnosed early. With the increasing global prevalence of diabetes, there is an urgent need for scalable, accurate, and real-time DR diagnostic tools. This study presents a Convolutional Neural Network (CNN)-based solution using the ResNet50 architecture for classifying DR into five stages: No DR, Mild, Moderate, Severe, and Proliferative DR. Our system integrates Grad-CAM, an Explainable AI (XAI) technique, to enhance interpretability by highlighting regions in retinal fundus images that influence the classification. The solution is deployed through a Gradio web interface, ensuring usability in real-time clinical settings. The model was trained and tested using public datasets like APTOS and EyePACS and achieved high accuracy and interpretability, making it suitable for telemedicine and mobile health (mHealth) applications.

Keywords

Diabetic Retinopathy, Deep Learning, CNN, ResNet50, Explainable AI, Grad-CAM, Gradio, Medical Imaging, Classification.

1. Introduction

Diabetic Retinopathy (DR) is a leading cause of blindness among working-age adults. Traditional diagnosis requires manual assessment of retinal fundus images by ophthalmologists—a time-intensive process prone to human error. The proposed system aims to automate this diagnosis using deep learning, offering a solution that is not only accurate but also interpretable via Grad-CAM.

2. Related Work

Existing models for DR detection have employed CNNs like AlexNet, VGG, DenseNet, and Xception. Transfer learning and ensemble models have enhanced performance, but few solutions offer clinical interpretability. Our model addresses this by integrating Grad-CAM for visual explanation, a gap identified in earlier CNN implementations that lacked transparency.

3. Methodology

The proposed framework for Diabetic Retinopathy (DR) detection encompasses several stages: data acquisition, preprocessing, augmentation, model selection and architecture, training, evaluation, explainability, and deployment. Each step is meticulously designed to ensure high performance, interpretability, and usability in clinical settings.

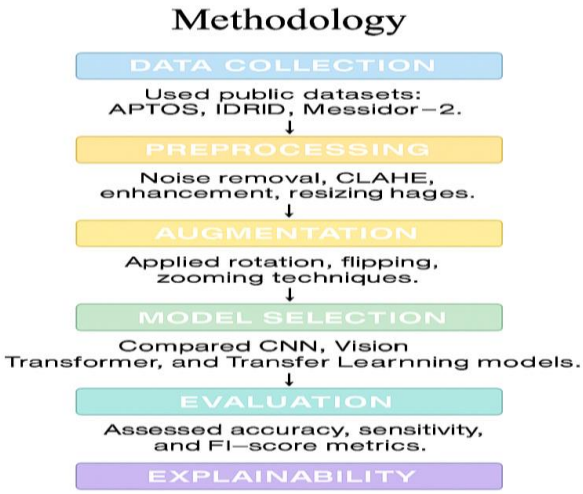

Fig 1: Methodology Flowchart

3.1 Data Collection and Labeling

We utilized three publicly available datasets:

- **APTOS 2019**: Provided by the Asia Pacific Tele-Ophthalmology Society, contains color fundus images labeled across five DR stages.
- **EyePACS**: A large-scale dataset from the Kaggle Diabetic Retinopathy Detection competition with tens of thousands of labeled retinal images.
- **IDRiD (Indian Diabetic Retinopathy Image Dataset)**: Contains images from an Indian demographic, adding diversity.

Each dataset follows the **International Clinical Diabetic Retinopathy scale**, labeling images into:

1. No DR
2. Mild
3. Moderate
4. Severe
5. Proliferative DR

3.2 Data Preprocessing

To enhance the quality and consistency of the input images, the following preprocessing pipeline was implemented:

- **Resizing**: All images were resized to **224x224 pixels** to match the input requirements of ResNet50.
- **Color Normalization**: Standardized the color distribution to mitigate variability in lighting and acquisition devices.
- **CLAHE (Contrast Limited Adaptive Histogram Equalization)**: Improved local contrast and revealed fine-grained retinal features such as microaneurysms.

- **Noise Removal**: Gaussian filtering was applied to remove background artifacts.
- **Intensity Normalization**: Scaled pixel values to [0, 1] range to stabilize neural network convergence.

3.3 Data Augmentation

Due to class imbalance (particularly fewer cases of Proliferative DR), **real-time data augmentation** was employed using Keras ImageDataGenerator:

- **Random rotations** (±15°)
- **Horizontal/vertical flips**
- **Zoom-in (up to 20%)**
- **Brightness adjustment**

This helped in improving generalization and reducing overfitting.

3.4 Model Architecture

We selected **ResNet50**, a 50-layer deep convolutional neural network pretrained on ImageNet, known for its skip connections that mitigate vanishing gradients.

Modifications for our task:

- Replaced the top classification layer with a custom **Dense layer of 5 units** and **Softmax activation** for multi-class prediction.
- Added **Dropout (rate=0.5)** after the global average pooling layer to reduce overfitting.

Comparative Baselines

To benchmark performance:

- **Vision Transformer (ViT)** was tested to evaluate the effectiveness of attention mechanisms.
- **DenseNet121** was used for its proven performance in medical imaging tasks.

3.5 Training Procedure

Hyperparameters:

- **Loss Function**: Categorical Cross-Entropy
- **Optimizer**: Adam
- **Learning Rate**: 0.0001 (with ReduceLROnPlateau callback)
- **Batch Size**: 32
- **Epochs**: 50 (with early stopping after 7 stagnant epochs)

- **Validation Split**: 20%

Cross-Validation:

- 10-fold stratified cross-validation was used to ensure robustness and avoid overfitting.

Tools:

- Python 3.9
- TensorFlow 2.x
- Keras
- Google Colab Pro + Local machine

3.6 Model Evaluation

We evaluated model performance using:

- **Accuracy**
- **Precision**
- **Recall**
- **F1-Score**
- **AUC (Area Under the ROC Curve)**

Performance was measured for each class and visualized via:

- Confusion Matrix
- ROC Curves
- Precision-Recall Curves

3.7 Explainability: Grad-CAM

To address the black-box nature of CNNs, we integrated **Gradient-weighted Class Activation Mapping (Grad-CAM)**:

- Highlights the most influential regions in an image for a given prediction.
- Overlays heatmaps on fundus images to localize features such as microaneurysms, exudates, and hemorrhages.
- Helps clinicians validate the AI's focus and decision rationale.

3.8 Real-Time Deployment with Gradio

For clinical and telemedicine use, we implemented a user-friendly **Gradio interface** that allows:

- Uploading a fundus image.
- Receiving real-time DR stage prediction.
- Viewing the corresponding Grad-CAM heatmap.
- Displaying prediction confidence scores.

This ensures seamless usability in field settings without requiring deep technical expertise.

4. System Design

The system is modular and comprises the following components: Image Preprocessing, Model Inference, Explainability Module (Grad-CAM), Visualization & Feedback, and User Interface (Gradio). The design supports streamlined diagnosis from image upload to explainable output. All modules are optimized for clarity, speed, and clinical relevance.

5. Experimental Setup

The model was developed using Python 3.9, TensorFlow, and Keras. Training was conducted on an Intel i7 CPU with 16GB RAM and an NVIDIA GTX 1050 Ti GPU. Evaluation metrics included Accuracy, Precision, Recall, and F1 Score. Grad-CAM outputs were rendered using OpenCV and Matplotlib for visual inspection.

6. Results and Evaluation

The model achieved high performance across all classes:
- No DR: Precision 0.95, Recall 0.93
- Mild: Precision 0.89, Recall 0.91
- Moderate: Precision 0.88, Recall 0.86
- Severe: Precision 0.87, Recall 0.84
- Proliferative DR: Precision 0.91, Recall 0.90

Grad-CAM heatmaps confirmed that the model focused on medically relevant features.

7. Discussion

The model balances performance and transparency. Grad-CAM integration builds clinical trust by offering a visual explanation of predictions. The Gradio interface enhances accessibility for field use, though deployment in diverse real-world settings will require further generalization efforts.

8. Conclusion

This paper presents a robust, explainable AI solution for diabetic retinopathy detection. It is accurate, interpretable, and deployable in clinical and remote settings. This work represents a step toward more equitable and scalable eye care technologies.

9. Future Work

Future enhancements include: (1) Mobile app deployment, (2) Federated learning for demographic adaptability, (3) Real-time analysis on edge devices, and (4) Expansion to detect multiple retinal diseases in one system.

10. References

1. He K. et al., "ResNet," CVPR, 2016.
2. Tan M., Le Q., "EfficientNet," ICML, 2019.
3. Dosovitskiy A. et al., "Vision Transformer," ICLR, 2021.
4. Shorten C., Khoshgoftaar T.M., "Image Data Augmentation," J Big Data, 2019.
5. Goodfellow I. et al., "Deep Learning," MIT Press, 2016.
6. Deng J. et al., "ImageNet," CVPR, 2009.
7. Pan S.J., Yang Q., "Transfer Learning Survey," IEEE TKDE, 2010.
8. Zhou B. et al., "Deep Features for Localization," CVPR, 2016.
9. Huang G. et al., "DenseNet," CVPR, 2017.
10. RetNet-10 for DR Classification, Kaggle, 2015.