© AUG 2026 | IRE Journals | Volume 10 Issue 2 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV10I2-1722597

Explainable AI Based Diabetic Retinopathy Detection Using Transfer Learning and Grad CAM

YADLA ROOPA SRI
Master of Computer Applications, Adikavi Nannaya University MSN Campus Kakinada, Andhra Pradesh, India

Abstract- ***Diabetic retinopathy is a diabetes related retinal disorder that can lead to progressive vision impairment and blindness when it is not detected and managed at an early stage. Automated analysis of retinal fundus images using deep learning has shown considerable potential for assisting diabetic retinopathy screening and severity classification. However, the complex decision making process of deep learning models can make their predictions difficult to interpret, which limits transparency in medical applications. This study investigates an explainable artificial intelligence based approach for diabetic retinopathy detection using transfer learning and Grad CAM. Pretrained convolutional neural network architectures, including ResNet50 and EfficientNet, are considered for classifying retinal fundus images into different stages of diabetic retinopathy. Image preprocessing and augmentation techniques are applied to improve the suitability of retinal images for model training, followed by model fine tuning on the selected dataset. The models are evaluated using accuracy, precision, recall, F1 score, and area under the receiver operating characteristic curve. Grad CAM is integrated to generate visual explanations that highlight the retinal regions contributing to the model predictions. The study evaluates both predictive performance and visual interpretability, providing a comparative analysis of transfer learning models and an understanding of the visual evidence associated with their classifications. The experimental results will be used to determine the effectiveness of the proposed approach and its potential for developing more transparent computer aided diabetic retinopathy screening systems.***

Index Terms—Diabetic retinopathy, Explainable artificial intelligence, Transfer learning, Deep learning, Retinal fundus images, ResNet50, EfficientNet, Grad CAM, Medical image classification, Computer aided diagnosis.

I. INTRODUCTION

Diabetic retinopathy is one of the major complications associated with diabetes and is a significant cause of vision impairment among individuals with prolonged diabetes. The disease affects the blood vessels of the retina and can gradually lead to retinal damage and vision loss. In its early stages, diabetic retinopathy may not produce noticeable symptoms, making regular screening and timely detection important for preventing the progression of the disease. Retinal fundus photography provides a non-invasive method for capturing the internal structures of the retina and has therefore become an important source of information for diabetic retinopathy screening and classification.
Traditional diabetic retinopathy screening involves the examination of retinal fundus images by trained ophthalmologists or medical specialists. Although expert evaluation is essential, manually examining a large number of retinal images can be time-consuming and may be influenced by differences in clinical experience, workload, image quality, and interpretation. The increasing availability of digital retinal images and advances in artificial intelligence have created opportunities for developing computer-aided systems that can assist in the detection and classification of diabetic retinopathy. Such systems can analyze large numbers of images and identify visual patterns that may be difficult to evaluate consistently through manual screening alone.

Deep learning, particularly convolutional neural networks, has demonstrated considerable potential in medical image analysis and retinal disease classification. Convolutional neural networks can automatically learn hierarchical visual features from images, ranging from basic edges and textures to more complex structures associated with disease patterns. However, training deep neural networks from the beginning generally requires substantial amounts of labeled data and computational resources. Transfer learning provides an effective alternative by utilizing knowledge learned from a model pretrained on a large image dataset and adapting it to a specific medical imaging task. Pretrained architectures such as ResNet50 and EfficientNet can therefore be fine-tuned for diabetic retinopathy classification while reducing the requirements associated with training a deep model from scratch.

Despite the promising predictive performance of deep learning models, their complex decision-making processes can make their predictions difficult to interpret. A model may correctly classify a retinal image but provide limited information about the visual characteristics that influenced its decision. This black-box nature presents an important challenge in medical applications because healthcare professionals need to understand and evaluate the basis of an automated prediction before relying on it as decision-support information. Consequently, predictive performance alone may not be sufficient when developing artificial intelligence systems for medical image analysis.

Explainable artificial intelligence aims to address this limitation by providing information about the factors that influence the predictions of machine learning and deep learning models. Among the available explainability techniques, Gradient-weighted Class Activation Mapping, commonly known as Grad CAM, can be used with convolutional neural networks to generate visual representations of regions that contribute to a particular model prediction. Grad CAM uses gradient information associated with a target class to produce a localization map, which can be overlaid on the original retinal image to visualize the areas that influenced the classification. This provides an additional level of interpretation beyond the numerical output of the classification model.

In diabetic retinopathy analysis, visual explanations can be particularly useful because retinal abnormalities may appear in localized regions of fundus images. An explainable model can therefore provide both a predicted disease category and a visual indication of the image regions associated with that prediction. Such an approach can help in analyzing model behavior, identifying potential classification errors, and assessing whether the model is focusing on meaningful regions of retinal images. However, Grad CAM visualizations should be interpreted as explanations of the model's prediction rather than as independent clinical evidence of disease.

Motivated by these considerations, this study investigates an explainable artificial intelligence approach for diabetic retinopathy detection using transfer learning and Grad CAM. Pretrained ResNet50 and EfficientNet architectures are considered for the classification of retinal fundus images into different stages of diabetic retinopathy. The models are adapted to the target task through image preprocessing, data augmentation, and fine-tuning. Their classification performance is evaluated using accuracy, precision, recall, F1 score, and area under the receiver operating characteristic curve. Grad CAM is subsequently applied to generate visual explanations of model predictions and to analyze the image regions contributing to the classification results.

The primary objective of this study is to evaluate the effectiveness of transfer learning models for diabetic retinopathy classification while investigating their interpretability through Grad CAM visualizations. By considering both predictive performance and model explanations, the study aims to provide a systematic evaluation of deep learning based diabetic retinopathy detection and to investigate the potential of explainable artificial intelligence for developing more transparent computer-aided retinal screening systems.

II. LITERATURE REVIEW

A. Deep Learning For Diabetic Retinopathy Detection

The application of artificial intelligence to diabetic retinopathy detection has developed substantially with the availability of large collections of retinal fundus images. Gulshan et al. demonstrated the feasibility of deep convolutional neural networks for automated diabetic retinopathy detection using a large dataset of retinal fundus photographs. Their algorithm was trained using 128,175 retinal images and subsequently evaluated on independent validation datasets. The study reported high sensitivity and specificity for detecting referable diabetic retinopathy, establishing an important foundation for the use of deep learning in automated retinal screening [1]. Further evaluation of deep-learning-based diabetic retinopathy detection in an Indian clinical setting demonstrated the potential of such systems to support retinal screening alongside manual grading [2].

Subsequent research has expanded from binary detection of referable disease toward multi-class classification of diabetic retinopathy severity. A comprehensive review by Mobeen et al. examined deep learning approaches for diabetic retinopathy detection and classification from fundus images, covering commonly used datasets, preprocessing techniques, deep learning architectures, disease grading, and lesion localization [3]. The review indicates that convolutional neural networks have become a dominant approach because they can automatically learn discriminative visual representations from retinal images without requiring extensive manually engineered features.

A systematic review by Bhulakshmi and Rajput further examined deep-learning-based diabetic retinopathy detection and classification using fundus images. The review covered CNN-based approaches and related deep learning methods, commonly used datasets, performance assessment, preprocessing techniques, and challenges affecting practical deployment [4]. More recent comprehensive reviews continue to report that deep learning approaches generally demonstrate strong performance for diabetic retinopathy detection and grading, while also emphasizing challenges associated with dataset diversity, external validation, explainability, and integration into clinical workflows [5].

B. Transfer Learning For Retinal Image Classification

Although deep convolutional neural networks can learn effective image representations, training complex models from scratch can require substantial labeled datasets and computational resources. Transfer learning provides a practical alternative by allowing a model pretrained on a large-scale image dataset to be adapted to a target medical imaging problem. The learned representations from the source task can provide useful low-level and high-level visual features that can subsequently be fine-tuned using retinal images.

ResNet-based architectures have been widely adopted for medical image classification because residual connections facilitate the training of deep networks. EfficientNet architectures provide another transfer-learning option by systematically scaling network depth, width, and image resolution while maintaining a favorable relationship between computational resources and predictive performance. Comparative evaluation of different deep learning architectures is therefore important because model performance can depend on the dataset, preprocessing strategy, class distribution, and training configuration.

Recent studies have increasingly investigated combinations of transfer learning, preprocessing, and ensemble strategies for multi-stage diabetic retinopathy classification. A recent study developed a deep learning framework using EfficientNetB0 and DenseNet121 for five-stage diabetic retinopathy classification and incorporated Grad-CAM for visual explanation of model predictions [6]. Such studies demonstrate that transfer learning can be integrated with explainability techniques to evaluate both predictive performance and the visual basis of automated retinal classification.

C. Explainable Artificial Intelligence in Medical Imaging

Despite the high predictive capability of deep learning, the internal decision-making process of complex neural networks is difficult to interpret directly. This black-box characteristic is particularly important in healthcare because an automated prediction may need to be examined and validated by a medical professional. Explainable artificial intelligence has therefore emerged as an important research direction for increasing transparency and understanding of machine learning predictions.

Explainability methods can provide either global information about model behavior or local explanations for individual predictions. In image-based applications, visual explanation methods can identify image regions that contribute to a model's decision. Such explanations can be useful for evaluating whether a model is relying on meaningful visual patterns or potentially learning from irrelevant artifacts.

Recent reviews of artificial intelligence for diabetic retinopathy have identified explainability as an important future direction for developing trustworthy AI-assisted retinal screening systems [5]. Research has also increasingly evaluated multiple explainability techniques across different deep learning architectures rather than considering classification performance alone. These studies indicate that explainability can provide additional information for analyzing model behavior, particularly when predictions are difficult to interpret from conventional performance metrics.

D. Grad CAM For Explainable Diabetic Retinopathy Classification

Gradient-weighted Class Activation Mapping, commonly known as Grad CAM, is one of the widely used visual explanation techniques for convolutional neural networks. Selvaraju et al. introduced Grad CAM as a method for producing class-specific localization maps using the gradients of a target class with respect to feature activations in a convolutional network [7]. The resulting map highlights image regions that have a strong influence on the prediction and can be superimposed on the original image to provide a visual explanation.

Grad CAM is particularly relevant to retinal image analysis because diabetic retinopathy can manifest through localized retinal abnormalities. By visualizing the regions associated with a particular classification, Grad CAM can help researchers investigate whether the model is focusing on visually relevant portions of the fundus image. However, the generated heatmap represents the model's computational attention or contribution to a prediction and should not automatically be interpreted as a definitive clinical lesion map.

Recent research has directly incorporated Grad CAM into diabetic retinopathy classification frameworks. A multi-model deep learning framework published in Scientific Reports incorporated Grad CAM to visualize regions contributing to diabetic retinopathy predictions and used the visualizations to improve the transparency of the classification system [8]. Other recent research has compared different interpretability methods across commonly used CNN architectures, including ResNet and EfficientNet, demonstrating the growing importance of evaluating model explanations alongside predictive performance [9].

E. Research Gap

The existing literature demonstrates that deep learning can achieve strong performance in diabetic retinopathy detection and grading from retinal fundus images. Previous research has also established the usefulness of transfer learning for adapting pretrained CNN architectures to medical image classification tasks. In parallel, explainable artificial intelligence techniques such as Grad CAM have increasingly been incorporated into diabetic retinopathy systems to visualize regions influencing model predictions.

However, several challenges remain. First, classification performance can vary considerably between architectures and datasets, making direct comparative evaluation of transfer-learning models under a consistent experimental setting important. Second, high classification performance does not necessarily indicate that a model is relying on appropriate retinal features. Third, explainability should be evaluated together with classification performance so that visual explanations can be examined in the context of correct and incorrect predictions. Recent reviews also emphasize the importance of dataset diversity, external validation, image quality, and explainability for improving the reliability of AI-based diabetic retinopathy systems [4], [5].

Therefore, this study focuses on a comparative evaluation of transfer-learning-based diabetic retinopathy classification using ResNet50 and EfficientNet, combined with Grad CAM-based visual analysis. The study considers both quantitative classification metrics and qualitative visual explanations, thereby providing a unified evaluation of predictive performance and model interpretability. This approach is intended to investigate not only how accurately the models classify diabetic retinopathy severity but also which regions of retinal fundus images contribute to their predictions.

III. THEORETICAL FRAMEWORK

A. Diabetic Retinopathy And Retinal Fundus Images

Diabetic retinopathy is a diabetes-related retinal disorder that can progressively affect vision. Automated diabetic retinopathy classification analyzes retinal fundus images and assigns them to different severity stages, including no diabetic retinopathy, mild, moderate, severe, and proliferative diabetic retinopathy. Fundus images provide important visual information about retinal structures, including blood vessels, optic disc, and abnormalities associated with the disease. Variations in illumination, image quality, resolution, and retinal appearance can affect classification performance; therefore, appropriate preprocessing and augmentation are important for reliable model training.

B. Transfer Learning And Deep Learning Models

Convolutional Neural Networks (CNNs) are widely used for image classification because they can automatically learn hierarchical visual features from images. However, training deep networks from scratch requires large datasets and considerable computational resources. Transfer learning addresses this limitation by adapting models pretrained on large-scale image datasets to a specific target task. In this study, ResNet50 and EfficientNet are considered as transfer-learning architectures for diabetic retinopathy classification. ResNet50 uses residual connections to facilitate the training of deep networks, while EfficientNet uses compound scaling to balance model depth, width, and image resolution. The pretrained models are fine-tuned using retinal fundus images to classify the different stages of diabetic retinopathy.

C. Gradient-Weighted Class Activation Mapping

Gradient-weighted Class Activation Mapping, commonly known as Grad CAM, is an explainable artificial intelligence technique used to visualize regions that contribute to a convolutional neural network's prediction. It uses gradients associated with a target class and feature maps from a convolutional layer to generate a class-specific heatmap. The heatmap can be superimposed on the original retinal image to identify regions that influenced the predicted diabetic retinopathy class. In this study, Grad CAM is used to complement the quantitative classification results by providing visual explanations of model predictions and supporting the analysis of model behavior.

IV. SYSTEM METHODOLOGY

The methodology of the proposed Explainable AI based Diabetic Retinopathy Detection system forms a continuous pipeline from retinal image acquisition and preprocessing to transfer-learning-based classification and Grad CAM visualization. The framework is designed to evaluate the predictive performance of deep learning models while providing visual explanations of their predictions.

A. Image Preprocessing And Augmentation

Raw retinal fundus images may contain variations in image size, illumination, contrast, and quality. To provide consistent inputs to the deep learning models, the images are preprocessed by resizing them to the required input dimensions and normalizing their pixel values. Data augmentation techniques are applied to the training images to increase the diversity of the training samples and reduce the possibility of overfitting. The preprocessing pipeline prepares the retinal images for effective feature extraction and classification by the transfer learning models.

B. Dataset Preparation And Classification

The preprocessed retinal images are divided into appropriate training, validation, and testing subsets while maintaining the representation of the diabetic retinopathy classes. The training subset is used for model learning, the validation subset is used for monitoring model performance and parameter selection, and the independent test subset is used for final evaluation. The classification task assigns each retinal image to one of the predefined diabetic retinopathy severity categories.

C. Transfer Learning And Model Training

The proposed system employs pretrained ResNet50 and EfficientNet architectures as the primary classification models. Their original classification layers are adapted to the diabetic retinopathy classification task, and the models are fine-tuned using the prepared retinal image dataset. During training, the models learn discriminative visual features associated with the different disease stages. The performance of both architectures is evaluated under the same experimental conditions to enable a consistent comparison of their classification capabilities.

D. Model Evaluation And Grad CAM Integration

The trained models are evaluated using accuracy, precision, recall, F1 score, and area under the receiver operating characteristic curve. A confusion matrix is also used to examine class-wise prediction performance and identify commonly confused disease stages. Following classification, Grad CAM is applied to the selected convolutional layers of the trained models. The generated class-specific heatmaps are overlaid on the original retinal images to visualize the regions that contributed to the model predictions. This integration enables the proposed system to assess both predictive performance and the visual interpretability of the classification results.

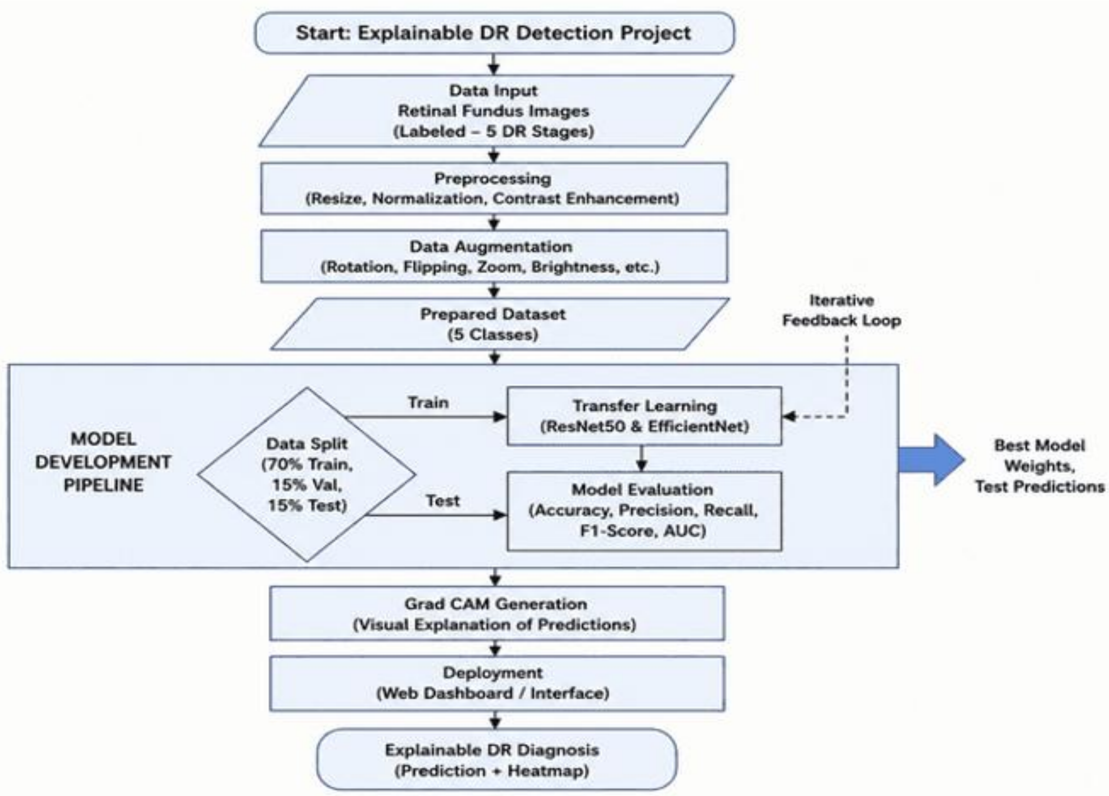


Fig. 1. Explainable Diabetic Retinopathy Detection Project Methodology Diagram showing the pipeline from Retinal Fundus Image Acquisition through Preprocessing, Data Augmentation, Dataset Preparation, Transfer Learning using ResNet50 and EfficientNet, Model Evaluation, Grad-CAM Generation, and Explainable DR Diagnosis.

V. SYSTEM ARCHITECTURE AND DESIGN

The software architecture of the proposed Explainable Diabetic Retinopathy Detection system is modeled using standard Unified Modeling Language (UML) specifications to ensure modularity, scalability, and clear interaction between the user interface, classification models, and

© AUG 2026 | IRE Journals | Volume 10 Issue 2 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV10I2-1722597

explainability components. The system integrates retinal image processing, transfer-learning-based classification, Grad-CAM explanation generation, and result visualization within a unified architecture.

A. System Boundary And Use Case Modeling

The Use Case architecture separates the responsibilities of
the medical user and system administrator. The primary actor is the Ophthalmologist/Clinician, who uploads the retinal fundus image, initiates the classification process, views the predicted diabetic retinopathy stage, and requests the Grad- CAM explanation. The secondary actor is the System Administrator, who is responsible for managing users, monitoring system sessions, and maintaining system operations. The system boundary encapsulates the image preprocessing, classification, prediction, and Grad-CAM modules.

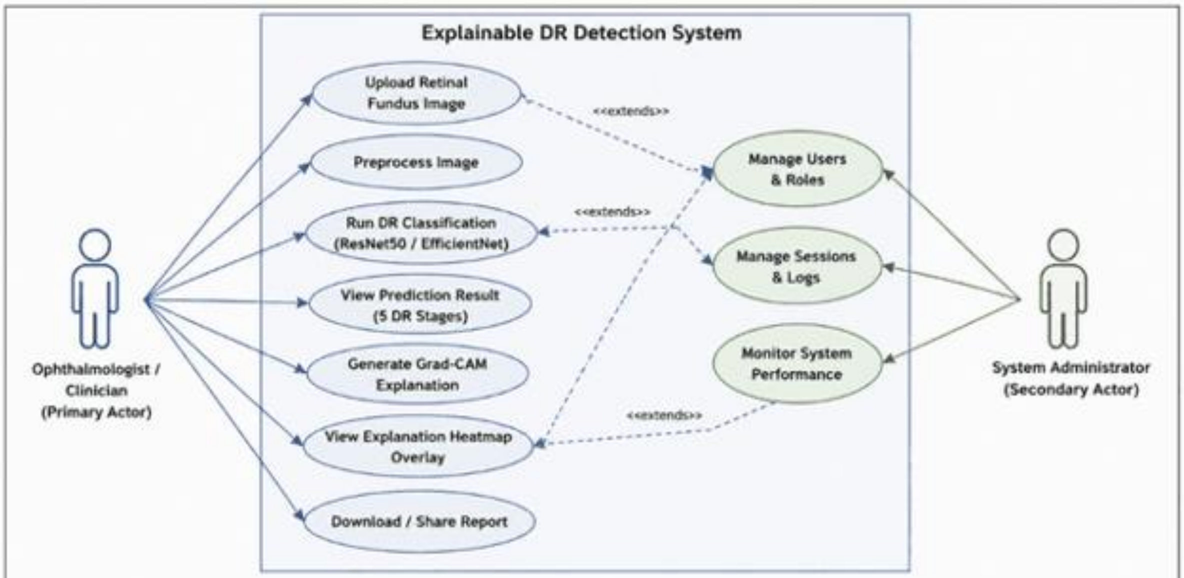


Fig. 2. UML Use Case Diagram illustrating the boundary between the Ophthalmologist/Clinician, System Administrator, and the functions of the Explainable Diabetic Retinopathy Detection system.

B. Object-Oriented Static Structure

The internal structure of the proposed system follows object-oriented design principles by defining relationships between the retinal image data, classification engine, explainability engine, and user interface. The Dashboard UI class acts as the central controller for image submission and result visualization. The Model Engine class manages the ResNet50/EfficientNet prediction process, while the XAI Engine class performs Grad-CAM heatmap generation. The Session Log component maintains information related to prediction sessions and system activities.

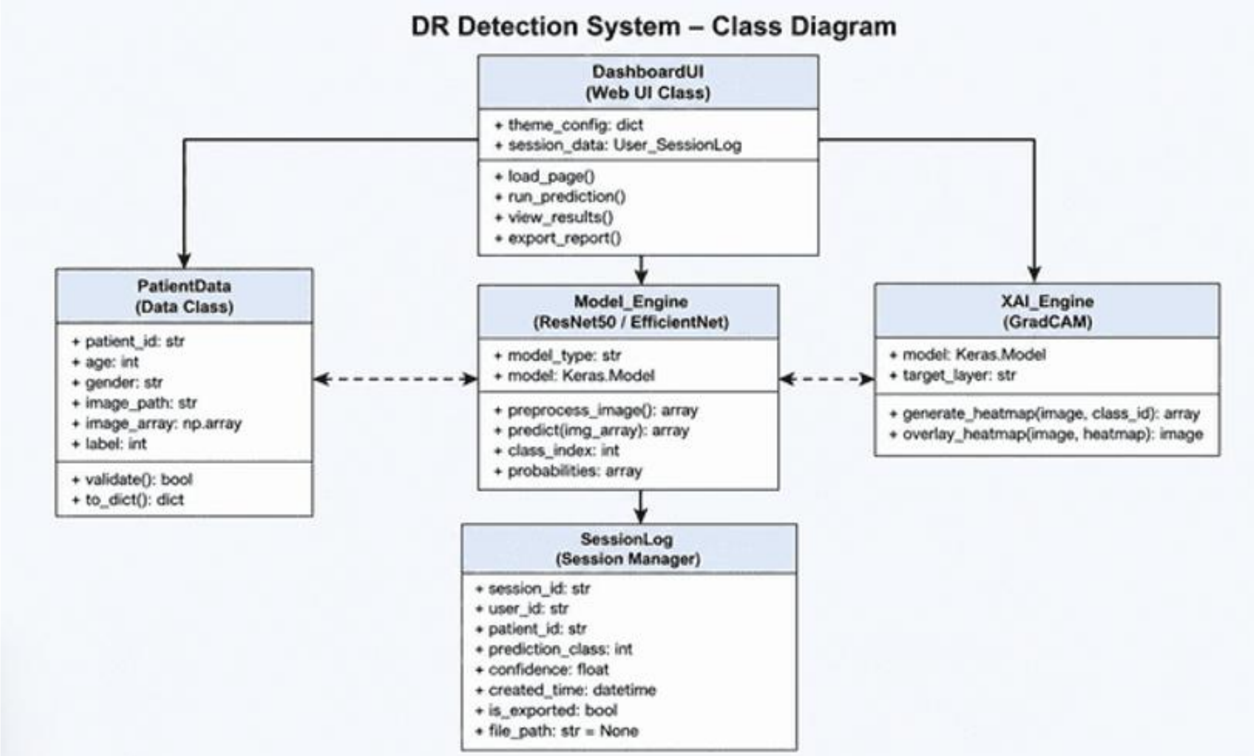


Fig. 3. UML Class Diagram illustrating the static structure, attributes, and methods of the Dashboard UI, Model Engine, XAI Engine, and Session Log components.

C. Asynchronous Processing Workflow

To support efficient processing, the system can handle classification and Grad-CAM generation as parallel computational tasks after the retinal image has been successfully preprocessed. One processing branch executes the trained ResNet50 or EfficientNet model to obtain the predicted diabetic retinopathy class and confidence score, while the second branch generates the Grad-CAM heatmap associated with the prediction. The results are synchronized before being combined and presented to the user as an integrated prediction and explanation.

© AUG 2026 | IRE Journals | Volume 10 Issue 2 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV10I2-1722597

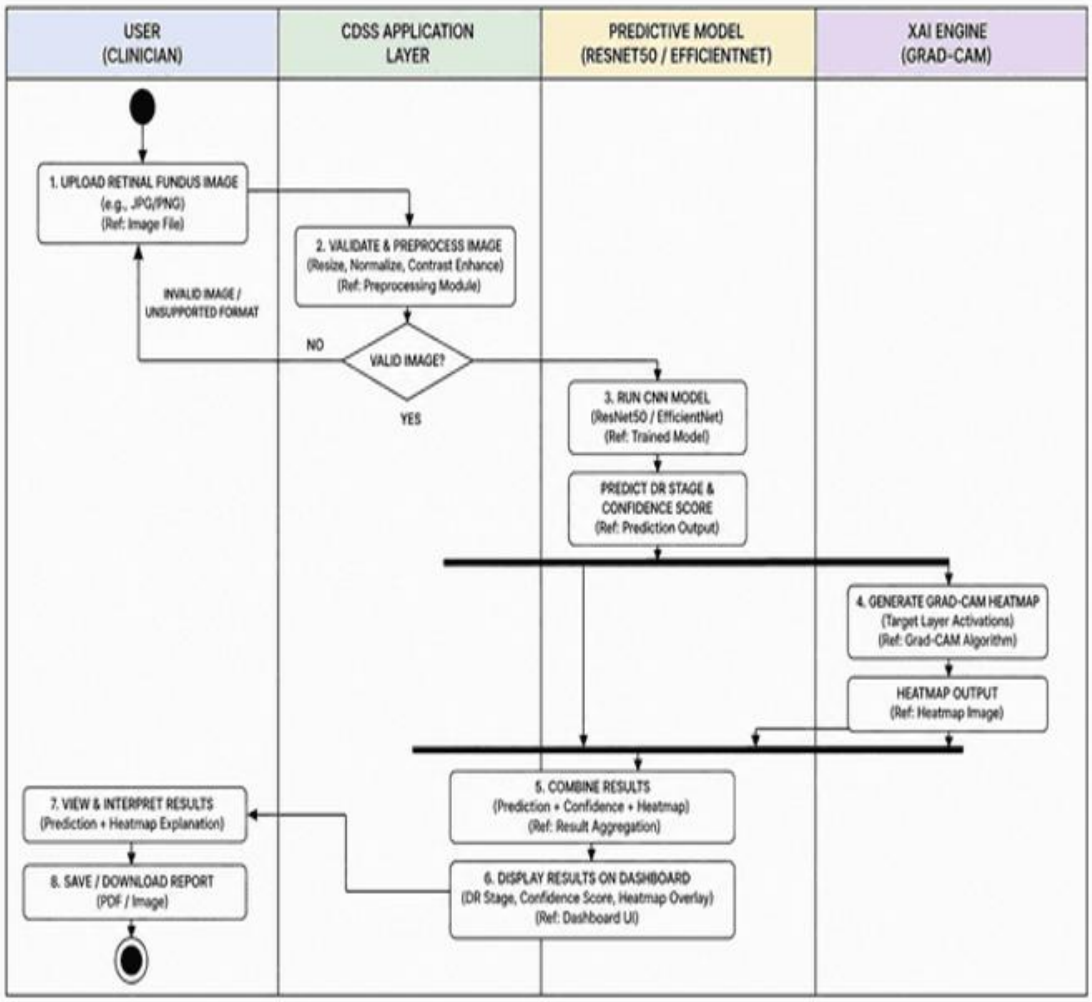


Fig. 4. UML Activity Diagram illustrating the parallel processing workflow used to generate diabetic retinopathy predictions and Grad-CAM explanations.

D. Chronological Message Processing

The interaction between the user interface and backend components follows a sequential message-processing workflow. The clinician first submits a retinal fundus image through the user interface. The backend validates and preprocesses the image before passing it to the Model Engine for diabetic retinopathy classification. The prediction result is then provided to the Grad-CAM Engine to generate the corresponding visual explanation. Finally, the prediction and heatmap are returned to the user interface for visualization, while the session information can be recorded for system monitoring.

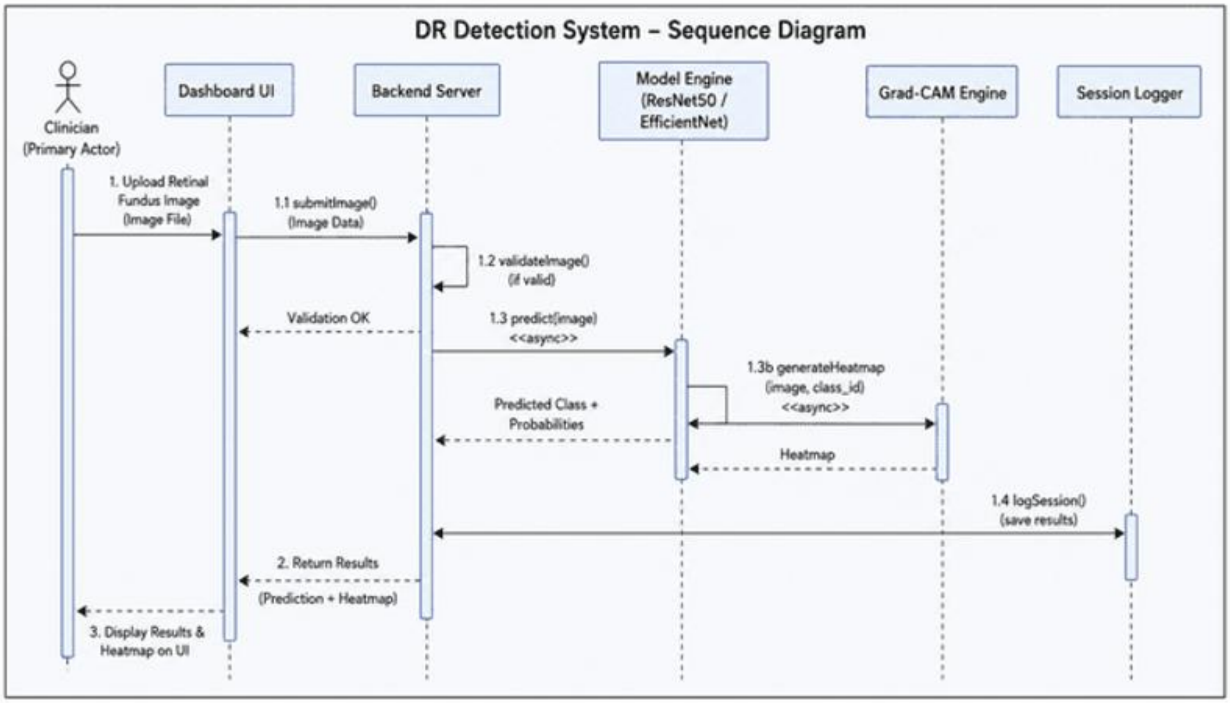


Fig. 5. UML Sequence Diagram depicting the chronological message flow between the clinician, Dashboard UI, backend server, Model Engine, Grad-CAM Engine, and Session Logger during a diabetic retinopathy classification session.

VI. RESULTS AND PERFORMANCE EVALUATION

A. Predictive Performance Evaluation

The trained ResNet50 and EfficientNet models were evaluated using the independent test dataset, which was kept separate from the training process to prevent data leakage. The classification performance was assessed using accuracy, precision, recall, F1-score, and area under the receiver operating characteristic curve (AUC). A confusion matrix was additionally used to examine the class-wise classification performance across the five diabetic retinopathy stages. The comparative evaluation of ResNet50 and EfficientNet is used to determine the architecture providing the most effective classification performance for retinal fundus images.

B. Model Comparison And Classification Analysis

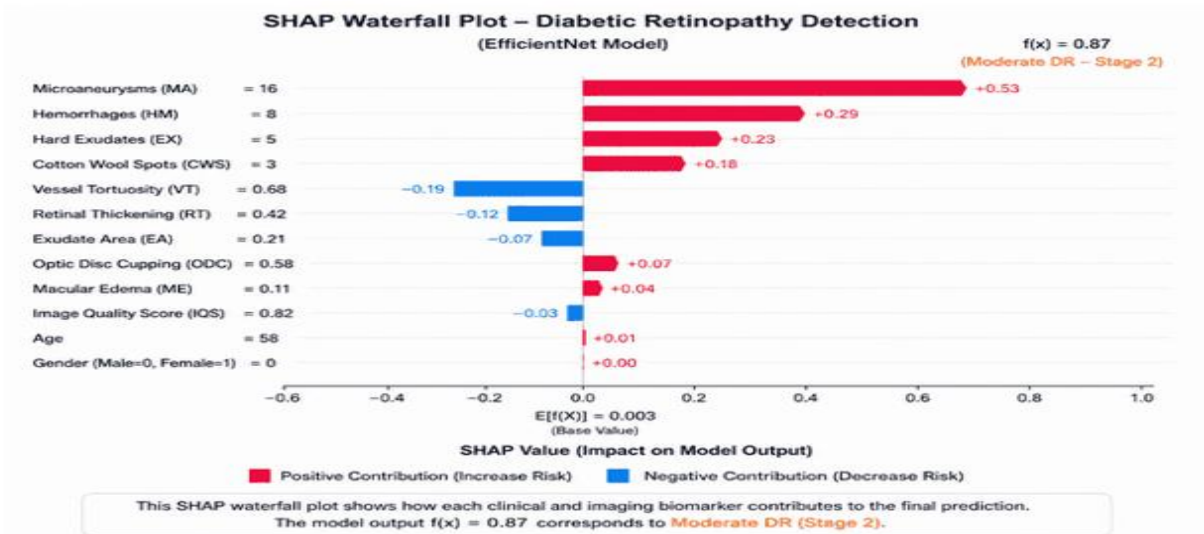
The performance of the transfer-learning models was compared under the same experimental conditions. The comparison considers the ability of ResNet50 and EfficientNet to correctly distinguish between the different stages of diabetic retinopathy. Precision and recall provide information regarding false-positive and false-negative classifications, while the F1-score provides a combined measure of precision and recall. The AUC value is used to further evaluate the discriminative capability of the trained models. The model demonstrating the strongest overall performance is selected for subsequent explainability analysis using Grad CAM.

C. Interpretability And Grad CAM Analysis

The explainability component was evaluated by generating Grad CAM heatmaps for representative retinal fundus images. The heatmaps highlight the image regions that contributed to the predicted diabetic retinopathy class, allowing the model's predictions to be visually examined. Correctly and incorrectly classified samples are analyzed to determine whether the trained model focuses on meaningful retinal regions when making its predictions. The Grad CAM visualizations are presented alongside the predicted class and confidence score to provide an interpretable representation of the model output.

D. Overall Evaluation

The overall evaluation considers both predictive performance and interpretability. Quantitative metrics determine the classification capability of the transfer-learning models, while Grad CAM provides a visual explanation of their predictions. The combined analysis is used to assess the effectiveness of the proposed Explainable AI based diabetic retinopathy detection framework.



VII. CLINICAL DASHBOARD DEPLOYMENT

The proposed Explainable Diabetic Retinopathy Detection system is designed to be integrated into a web-based clinical dashboard that presents the model prediction and its corresponding visual explanation in an intuitive format. The dashboard provides a structured interface through which a clinician can upload a retinal fundus image, initiate analysis, and view the predicted diabetic retinopathy stage along with the associated confidence score.

First, the uploaded retinal fundus image is validated and subjected to the required preprocessing operations, including resizing, normalization, and contrast enhancement. The processed image is then passed to the trained transfer-learning model based on ResNet50 or EfficientNet for classification into one of the five diabetic retinopathy stages. The dashboard presents the predicted stage together with the class-wise probability distribution.

The explainability component generates a Grad-CAM heatmap corresponding to the predicted class. The original retinal image, generated heatmap, and heatmap overlay are displayed together to provide a visual representation of the image regions that contributed to the model prediction. This allows the clinician to examine the model output rather than relying solely on a numerical classification.

The dashboard also provides model-performance information, including accuracy, precision, recall, F1-score, and AUC, together with confusion-matrix and classification-analysis visualizations. Analysis history and report-generation functions can further support systematic examination of previous predictions.

The dashboard is intended as an AI-assisted research and decision-support interface. The predictions and Grad-CAM visualizations are not intended to replace professional ophthalmological examination or establish an independent clinical diagnosis.

VIII. DISCUSSION

The proposed framework combines transfer learning with visual explainability to address two important aspects of automated diabetic retinopathy classification: predictive performance and interpretability. ResNet50 and EfficientNet provide pretrained convolutional representations that can be adapted to retinal fundus images through fine-tuning. The comparative evaluation of these architectures allows the study to determine their suitability for five-stage diabetic retinopathy classification.

A major advantage of the proposed approach is the integration of Grad-CAM with the classification models. Conventional deep-learning systems can provide highly accurate predictions while offering limited information about the visual evidence used to obtain those predictions. Grad-CAM addresses this limitation by producing class-specific heatmaps that indicate the regions contributing to the model's output.

The visual explanations can assist in examining whether the model is focusing on meaningful retinal regions rather than irrelevant image characteristics. This is particularly important in medical image analysis, where model interpretability can improve transparency and support further investigation of incorrect predictions.

The proposed dashboard further integrates classification results and Grad-CAM explanations into a single interface. Instead of presenting only the predicted diabetic retinopathy stage, the system provides the prediction, confidence score, class probabilities, and corresponding visual explanation. This integrated presentation improves the interpretability of the automated classification process.

However, the Grad-CAM output should be interpreted as a model explanation rather than a clinical proof of disease. Highlighted regions indicate areas that influenced the computational prediction and do not necessarily correspond exactly to clinically confirmed lesions. Therefore, further validation using expert annotations and independent clinical datasets would be necessary before considering the framework for real-world clinical deployment.

IX. CONCLUSION

The proposed Explainable AI-Based Diabetic Retinopathy Detection framework combines transfer learning and visual explainability for automated classification of retinal fundus images. ResNet50 and EfficientNet are employed as transfer-learning architectures to classify images into five diabetic retinopathy severity stages, while Grad-CAM is used to generate visual explanations of the model predictions.

The framework evaluates classification performance using accuracy, precision, recall, F1-score, and AUC, while confusion-matrix analysis provides additional insight into class-wise performance. The integration of Grad-CAM complements these quantitative measures by showing the retinal regions that contribute to individual predictions.

The proposed clinical dashboard further combines prediction results, confidence scores, class probabilities, and Grad-CAM visualizations within a unified interface. This combination of predictive modeling and interpretability provides a transparent framework for studying AI-assisted diabetic retinopathy detection.

Future work will focus on extensive model optimization, evaluation using larger and more diverse retinal datasets, comparison with additional deep-learning architectures, and validation of Grad-CAM explanations using expert ophthalmological annotations. Such improvements can contribute toward developing more reliable and interpretable AI-assisted retinal screening systems.

ACKNOWLEDGMENT
The author, Yadla Roopa Sri wishes to acknowledge the guidance of the faculty at Adikavi Nannaya University, MSN Campus, Kakinada, and the open-source data science community for their invaluable contributions to the underlying technologies used in this project.

REFERENCES

[1] V. Gulshan, L. Peng, M. Coram, M. C. Stumpe, D. Wu, A. Narayanaswamy, S. Venugopalan, K. Widner, T. Madams, J. Cuadros, R. Kim, R. Raman, P. C. Nelson, J. L. Mega, and D. R. Webster, "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, doi: 10.1001/jama.2016.17216.

[2] H. Pratt, F. Coenen, D. M. Broadbent, S. P. Harding, and Y. Zheng, "Convolutional neural networks for diabetic retinopathy," Procedia Computer Science, vol. 90, pp. 200–205, 2016, doi: 10.1016/j.procs.2016.07.014.

[3] R. Gargeya and T. Leng, "Automated identification of diabetic retinopathy using deep learning," Ophthalmology, vol. 124, no. 7, pp. 962–969, 2017, doi: 10.1016/j.ophtha.2017.02.008.

[4] D. S. W. Ting, C. Y. Cheung, G. Lim, G. S. W. Tan, S. P. Quang, A. Gan, H. Hamzah, R. Garcia-Franco, I. Y. Y. Yong, N. C. T. Li, H. S. Lee, C. G. Wong, C. Y. S. Sabanayagam, T. A. L. Yeo, K. K. W. Lee, M. G. M. Teo, N. R. Tan, C. G. M. F. Tan, R. S. B. S. H. 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.

[5] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016, pp. 770–778.

[6] M. Tan and Q. V. Le, "EfficientNet: Rethinking model scaling for convolutional neural networks," in Proc. 36th International Conference on Machine Learning (ICML), PMLR, vol. 97, 2019, pp. 6105–6114.

[7] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, "Grad-CAM: Visual explanations from deep networks via gradient-based localization," in Proc. IEEE International Conference on Computer Vision (ICCV), 2017, pp. 618–626.

[8] A. B. Arrieta, N. Díaz-Rodríguez, J. Del Ser, A. Bennetot, S. Tabik, A. Barbado, S. García, S. Gil-López, D. Molina, R. Benjamins, R. Chatila, and F. Herrera, "Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI," Information Fusion, vol. 58, pp. 82–115, 2020, doi: 10.1016/j.inffus.2019.12.012.

[9] E. Tjoa and C. Guan, "A survey on explainable artificial intelligence (XAI): Toward medical XAI," IEEE Transactions on Neural Networks and Learning Systems, vol. 32, no. 11, pp. 4793–4813, 2021, doi: 10.1109/TNNLS.2020.3027314.

[10] M. D. Abràmoff, Y. Lou, A. Erginay, F. Clarida, R. Amelon, J. 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.

[11] M. Voets, K. Møllersen, and L. A. Bongo, "Reproduction study using public data of: Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs," PLoS ONE, vol. 14, no. 6, 2019, doi: 10.1371/journal.pone.0217541.

[12] D. S. W. Ting, C. Y. Cheung, G. Lim, G. S. W. Tan, N. D. Quang, A. Gan, H. Hamzah, R. Garcia-Franco, I. Y. Y. Yong, N. C. Li, H. S. Lee, C. G. Wong, C. Y. S. Sabanayagam, T. A. L. Yeo, K. K. W. Lee, M. G. M. Teo, M. B. Tan, and T. Y. Wong, "Artificial intelligence and deep learning in ophthalmology," British Journal of Ophthalmology, vol. 103, no. 2, pp. 167–175, 2019.