RESEARCH PAPER

Diabetic Retinopathy Detection Using Multimodal Deep Learning and Explainable Ai: A Review

Prashant Dhope[1], Meghana Kulkarni[2]

[1]Research Scholar, Department of Electronics and Communication Engineering, Visvesvaraya Technological University, Belagavi, Karnataka, India - 590018

Email: prashantdhope@gmail.com

[2]Associate Professor, Department of Electronics and Communication Engineering, Visvesvaraya Technological University, Belagavi, Karnataka, India – 590018

Email: meghanak@vtu.ac.in

Received: 17th Mar, 2026 | Revised: 29th Mar, 2026 | Accepted: 19th Apr, 2026 | Available Online: 5th May, 2026

ABSTRACT

Diabetic Retinopathy (DR) is one of the leading causes of vision impairment and blindness among individuals with diabetes worldwide. The condition occurs due to prolonged high blood glucose levels that damage retinal blood vessels, resulting in leakage, hemorrhages, and the growth of abnormal vessels in the retina. Early detection and timely treatment are essential to prevent irreversible vision loss; however, manual screening of retinal images is time-consuming and requires expert ophthalmologists. In recent years, deep learning techniques have shown significant potential in automating DR detection from retinal images. Furthermore, multimodal deep learning approaches that combine multiple data sources, such as retinal fundus images, optical coherence tomography (OCT) scans, and clinical data, have improved diagnostic performance and robustness. Despite their high accuracy, many deep learning models operate as black-box systems, limiting their clinical adoption due to the lack of interpretability. Explainable Artificial Intelligence (XAI) methods have therefore emerged as an important component for providing visual and feature-level explanations that enhance transparency and trust in automated diagnostic systems. This paper presents a comprehensive review of recent advancements in diabetic retinopathy detection using multimodal deep learning and explainable AI techniques. The review summarizes commonly used datasets, deep learning architectures, multimodal fusion strategies, and explainability methods, while also identifying existing challenges and future research directions for developing reliable and clinically interpretable DR detection systems.

Keywords: Diabetic Retinopathy, Disease Detection, Disease Grading, Deep Learning, Explainable Ai

How to cite this article: Dhope P, Kulkarni M., Diabetic Retinopathy Detection Using Multimodal Deep Learning and Explainable Ai: A Review. Int J Drug Deliv Technol. 2026;16(5): 695-707; Doi: 10.25258/Ijddt.16.5.74

INTRODUCTION

1.1 Human Eye

The human eye is a complex sensory organ that allows us to see, converting light into electrical signals that the brain interprets as visual images. It's a roughly spherical structure with various parts working together to capture and process light. The structure of the human eye is depicted in Fig. 1. The key components in human eye include:

- **Cornea:** The clear, outer layer that covers the iris and pupil, helping to focus light.
- **Iris:** The colored part of the eye that surrounds the pupil and controls the amount of light entering the eye.
- **Pupil:** The black circular opening in the iris that allows light to pass through.
- **Lens:** A transparent structure behind the pupil that further focuses light onto the retina.
- **Retina:** A light-sensitive layer at the back of the eye that contains cells (rods and cones) that convert light into electrical signals.
- **Optic Nerve:** A bundle of nerve fibers that carries the electrical signals from the retina to the brain.

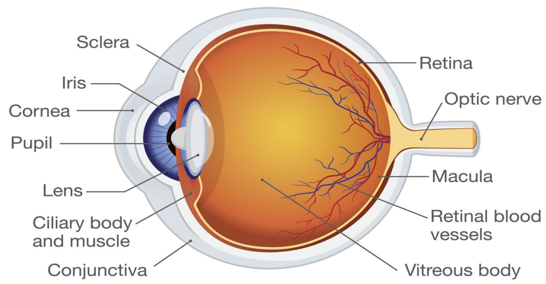

Fig. 1: Structure of the Human Eye

The eye works by capturing light, focusing it through the cornea and lens, and then converting it into electrical signals by the retina. These signals are transmitted to the brain via the optic nerve, where they are interpreted as visual images.

1.2 Diabetic Retinopathy

Retinopathy refers to any disease or damage that affects the retina. It can result from various underlying conditions, most commonly: diabetes, hypertension and any other systemic or vascular disorders. If left untreated, retinopathy can lead to progressive vision impairment and, in severe cases, permanent blindness. Early detection and proper medical intervention are crucial to preventing visual disability associated with retinal diseases.

Diabetic Retinopathy (DR) is an eye condition that can lead to vision loss and blindness in people with diabetes. It occurs when high blood sugar levels damage the blood vessels in the retina. This damage can cause the blood vessels to leak fluid or blood, or new, abnormal blood vessels to grow on the retina's surface. Early detection and treatment are crucial to prevent or reduce vision loss. How diabetic eyes differ from healthy eyes is depicted in Fig. 2.

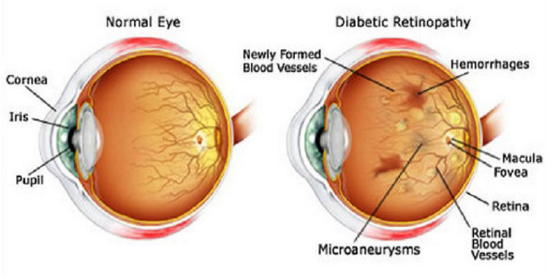

Fig. 2: Healthy and Diabetic Eye

Diabetic Retinopathy progresses through four main stages, each representing increasing damage to the retina. The fundamental stages in diabetic retinopathy are depicted in Fig. 3 and they are categorized as follows:

(a) Mild Non-Proliferative Diabetic Retinopathy (NPDR):

This is the earliest stage of diabetic retinopathy. Tiny balloon-like swellings called microaneurysms form in the retinal blood vessels due to high blood sugar levels. These weak spots may leak small amounts of fluid into the retina.

(b) Moderate NPDR:

As the disease progresses, more retinal blood vessels become damaged, causing microaneurysms to increase in number and size. These vessels may start leaking blood, resulting in hemorrhages within the retina, affecting its normal function.

(c) Severe NPDR:

In this stage, many retinal blood vessels are either severely damaged or blocked, leading to large areas of the retina not receiving enough blood (ischemia). This causes more severe hemorrhages and signals the retina to create new blood vessels to make up for the oxygen loss.

(d) Proliferative Diabetic Retinopathy (PDR):

This is the most advanced stage. In response to oxygen shortage, the retina grows newly formed blood vessels. These vessels are abnormal and fragile. They can bleed easily into the eye (vitreous hemorrhage), leading to vision loss or complications like retinal detachment.

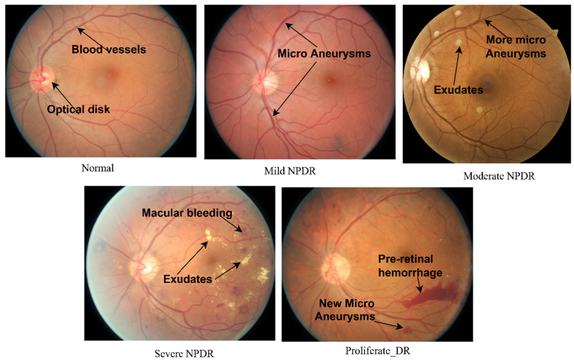

Fig. 3: Fundamental stages in Diabetic Retinopathy

1. LITERATURE SURVEY

2.1 Related Work

Numerous studies on the identification of diabetic retinopathy have been published in the literature, and a number of problem solvers have expressed interest in identifying the various stages of DR. An early diagnosis of DR is essential to minimize its complications. This is because early and accurate detection of diabetic retinopathy has wide-ranging applications in modern healthcare, including improved patient outcomes, reduced treatment costs, and enhanced clinical decision-making.

A deep learning approach using DenseNet121 was developed to classify fundus images into five diabetic retinopathy stages (0 to 4), using the APTOS Kaggle dataset with 3662 training images. The model achieved an accuracy of 96.11% and a quadratic weighted kappa score of 0.8981. A performance comparison with VGG16 showed DenseNet's superiority. The method focuses only on image-based detection, without incorporating multimodal data or offering explainable outputs for clinical interpretability [1].

The methodology developed by K Shankar, et al. classifies diabetic retinopathy severity levels using a Synergic Deep Learning (SDL) model on fundus images. The process includes preprocessing to remove noise, histogram-based segmentation, and final classification using SDL. The model was evaluated on the Messidor dataset, demonstrating improved performance over existing methods. While the approach enhances severity classification, it does not incorporate multimodal clinical data or explainability techniques, limiting clinical decision support [2].

The methodology was developed using Convolutional Neural Networks (CNN) for automated detection of diabetic retinopathy from fundus images. This approach aims to reduce the time, cost, and subjectivity associated with manual diagnosis by ophthalmologists. CNN is applied for image recognition and classification, achieving high accuracy in DR detection. The paper also highlights challenges in existing techniques. However, it lacks multimodal data usage and does not address model explainability, which limits its clinical applicability [3].

The methodology proposed by Chava Harshitha, et al. aimed to automate DR detection by identifying abnormal blood vessels, exudates, and cotton wool spots from fundus images. The images were enhanced using histogram equalization, segmented via k-means clustering, and classified using SVM and Random Forest algorithms. The Random Forest achieved a high recognition accuracy of 96.62%, outperforming SVM. However, the study used a small dataset and lacked multimodal data integration and explainability features [4].

The approach followed by Dolly Das, et al. presents a comprehensive review of DR detection methods, focusing on machine learning (ML) and deep learning (DL) techniques. The paper discusses various feature extraction methods, highlights the limitations of traditional ML models, and emphasizes the superiority of DL in handling large datasets for improved accuracy. While valuable for background understanding, the study does not propose a new model, does not integrate multimodal data, and lacks explainability components [5].

The methodology was developed for intelligent diagnosis of severe diabetic retinopathy using Inception V3 on a large Kaggle fundus image dataset. The model classifies lesion severity using images of two resolutions, with the 896×896 input size achieving superior performance (AUC: 0.968, Sensitivity: 0.925). Lesions like cotton wool spots and exudates are often misclassified, while IRMA is hardest to detect. The approach improves severe DR screening, but lacks multimodal data use and explainable AI techniques for clinical transparency [6].

A solution proposed by KM Fellah, et al. involves a custom Convolutional Neural Network (CNN) for detecting diabetic retinopathy from color fundus images. The input images were preprocessed using various filters to improve quality before training. The proposed model showed better performance compared to existing methods, demonstrating its effectiveness for automated DR detection. However, the approach lacks multimodal data inputs and does not incorporate explainability mechanisms, which limits its use in clinical settings [7].

The use of deep learning-based retinal fundus image segmentation (RFIS) techniques were used to aid early detection of diabetic retinopathy. This systematic review covered 115 studies, focusing on segmentation of blood vessels (66%), lesions (36%), and optic disc/cup (15%) using various DL approaches. The review also evaluated datasets, preprocessing methods, and performance metrics, while highlighting challenges and future research directions. However, most reviewed methods lack multimodal data usage and explainability, limiting clinical reliability [8].

D Das, et al. explored the use of 26 state-of-the-art deep learning architectures for diabetic retinopathy detection using fundus images. The models were trained and evaluated on Kaggle's EyePACS dataset, with EfficientNetB4 emerging as the most reliable, achieving 99.37% training accuracy and 79.11% validation accuracy. The study compared models like InceptionV3, ResNet50, DenseNet169, and others for performance and overfitting. While the approach offers valuable insights into model selection, it lacks multimodal clinical data and explainability, which are essential for real-world deployment [9].

Kalyani, G., et al. implemented a model that uses a reformed capsule network to detect and classify diabetic retinopathy from fundus images. The network extracts features through convolution and capsule layers, followed by a softmax layer for class prediction. Using the Messidor dataset, it achieved high accuracies across DR stages, including 98.64% for stage 3. While the model performs well, it does not include multimodal data or explainability components, which are important for clinical deployment [10].

According to Rachapudi, Venubabu, et al., the proposed system aims to improve diabetic retinopathy detection by implementing a multi-stage pipeline comprising preprocessing, segmentation, feature extraction, and classification. Key techniques include CLAHE, Circle Hough Transform for optic disc removal, MEM algorithm for exudate segmentation, and GLCM for feature extraction. A DNN-BOA classifier is used to categorize DR stages, achieving 98.3% accuracy on DIARETDB1 and 98.9% on MESSIDOR, outperforming traditional classifiers. Despite strong performance, the system lacks multimodal inputs and explainable AI features, limiting its clinical interpretability [11].

Palaniswamy, T. and Vellingiri, M. introduced a method based on an IoT and deep learning-enabled system (IoTDL-DRD) for diabetic retinopathy detection using retinal fundus images. The method collects data through IoT devices, preprocesses images, and applies mayfly optimization-based region growing (MFORG) for lesion segmentation. DenseNet is used for feature extraction, while LSTM, optimized via the Honey Bee Optimization algorithm, handles classification. The system shows enhanced performance over traditional models. However, it lacks explainability features and multimodal clinical data, which are important for real-world medical deployment [12].

The study by Raman, et al., focuses on detecting diabetic retinopathy using the EfficientNet-B7 deep learning model, emphasizing early detection through retinal layer thickness analysis and addressing retinal neurodegeneration. The model was trained on the Aptos2019-oversampled dataset, incorporating data augmentation and hyperparameter tuning to mitigate overfitting. It achieved a 94% classification accuracy, showing strong potential for automated DR diagnosis. However, the model lacks multimodal integration and explainable decision support, which are crucial for clinical adoption [13].

According to MR Shoaib, et al., the proposed system aims to enhance diabetic retinopathy diagnosis by applying transfer learning using InceptionResNetv2, InceptionV3, and a custom-built DiaCNN model. The models are trained and tested on the ODIR dataset, achieving testing accuracies of 97.5% and 98.3%, respectively. This deep learning-based approach outperforms traditional diagnostic methods. However, it is limited by its reliance on image-only data and the absence of explainability features, which restrict its clinical acceptance [14].

The study by Malik, et al., focuses on automated diabetic retinopathy detection using a VGG-19-based deep learning classifier. The model is trained on a diverse dataset of retinal fundus images, preprocessed using grayscale conversion, Gaussian filtering, and circular cropping. Through transfer learning and fine-tuning, the classifier achieved 64.5% accuracy for four-class DR severity classification. While the approach demonstrates the potential of VGG-19 in DR detection, its moderate accuracy suggests room for improvement through multimodal inputs and explainability enhancements for clinical application [15].

According to Saranya, P. and K. M. Umamaheswari, the proposed system aims to automate the early detection of bright lesions (exudates) in non-proliferative diabetic retinopathy by employing a deep learning architecture. The model performs image background removal, optic disc elimination, and lesion segmentation to enhance lesion identification. Using the MESSIDOR and e-ophtha Ex datasets, it achieved a maximum accuracy of 97.54%, with strong sensitivity, specificity, and F1-scores. The approach offers a cost-effective, high-accuracy alternative to manual screening, but lacks clinical explainability and multimodal input integration [16].

Guefrachi, Sarra, Amira Echtioui, and Habib Hamam introduced a method based on a deep learning model (ResNet152-V2) integrated into a graphical user interface (GUI) for the detection of diabetic retinopathy using retinal fundus images. The system supports clinicians by automating diagnosis, displaying disease severity, and storing patient information locally. It achieved 100% accuracy, precision, recall, and F1 Score. While the model shows strong diagnostic potential, it lacks multimodal data integration and explainability features, which are important for clinical adoption [17].

Suo, et al., designed a framework using a CS-ResNet-101 model enhanced with a Full Convolution Spatial Attention Module (FCSAM) for automated diabetic retinopathy classification. An image enhancement algorithm was employed to improve DR image quality, and a modified cross-entropy loss function was introduced to reduce overfitting. Trained on the EyePACS dataset and externally validated on APTOS 2019, the model achieved 98.1% accuracy, 99.6% specificity, and demonstrated strong generalization performance. Although the framework excels in classification, it does not incorporate multimodal clinical data or address explainability, which are key for clinical adoption [18].

Rajalingam, et al. explored the use of deep learning techniques for the efficient recognition of diabetic retinopathy using color fundus images. The proposed model employed Convolutional Neural Networks (CNNs) to extract features and classify various DR stages. The study utilized standard retinal image datasets and achieved high specificity and sensitivity, as validated through ROC curve, AUC, and other performance metrics. The method demonstrates significant potential for automated DR screening, enabling earlier interventions and improving clinical outcomes. However, it lacks multimodal input integration and explainability mechanisms, which are key for real-world deployment [19].

The study by K. Ahnaf Alavee et al. focuses on leveraging advanced deep learning and machine learning techniques for the early detection of diabetic retinopathy (DR). It explores various transfer learning models such as DenseNet121, Xception, ResNet50, VGG16, VGG19, and InceptionV3, along with classifiers like SVM and RNN. The authors propose a custom CNN model that achieves superior performance with 95.27% accuracy, outperforming existing state-of-the-art models. The study also emphasizes explainability using XAI methods like Grad-CAM to enhance model transparency and clinical trust [20].

Jabbar, Ayesha, et al. introduced a deep transfer learning-based system for real-time detection of diabetic retinopathy (DR) using fundus cameras, with a focus on improving healthcare access in remote and underserved regions. The proposed system captures retinal images via fundus imaging and transmits them to a processing unit, where VGGNet-based deep learning algorithms classify the severity of DR. It also generates comprehensive diagnostic reports to assist in clinical decision-making. The model achieved a classification accuracy of 97.6%, outperforming several existing methods. By enabling early and automated DR diagnosis, the system significantly reduces diagnostic costs and improves patient outcomes, particularly in resource-limited settings [21].

Ahmed, Fahad, et al. implemented a model that utilizes the ResNet-18 deep learning architecture to detect diabetic retinopathy (DR) from fundus images. Using a dataset sourced from Kaggle, the model achieved a training accuracy of 99.91% and a testing accuracy of 96.65%, showcasing its robustness and precision. The approach aims to assist in early DR identification, streamlining the diagnostic workflow and enhancing timely interventions. This model holds promise for improving clinical decision-making, minimizing the risk of blindness, and optimizing healthcare resource allocation for diabetic patients [22].

Das, Uddipan, et al. designed a framework using deep learning algorithms to improve the detection of diabetic retinopathy (DR). Their approach incorporates multiple processing stages: autoencoders for resizing and normalization, a denoising CNN to reduce noise, and CLAHE preprocessing via MATLAB to enhance image quality. For classification, the model uses an InceptionV3-based CNN architecture to perform segmentation, feature extraction, and diagnosis. The system achieves a detection accuracy of 95.6%, offering a robust and reliable solution for early DR identification and aiding in timely clinical interventions [23].

Sushith, et al., introduced a hybrid deep learning framework combining CNNs and RNNs with attention mechanism for early detection and progression monitoring of diabetic retinopathy (DR). The model leverages temporal information across multiple retinal scans to enhance detection accuracy. Spatial features are extracted via CNN with multi-scale enhancement, followed by RNN-based temporal analysis. Evaluated on DRIVE, Kaggle, and Eyepacs datasets, the method achieved accuracy rates of 97.5%, 94.04%, and 96.9%, respectively. While the model shows promising results in temporal DR analysis, it lacks multimodal clinical data integration and explainability features, which are critical for real-world clinical application [24].

Muhammad Faris, et al., explored the use of the VGG-16 deep learning model to classify diabetic retinopathy stages from retinal fundus images. To address class imbalance in the dataset, they applied Synthetic Minority Over-sampling Technique (SMOTE) and Tomek Links. The model was trained on a large dataset of 88,702 images, categorized into five DR stages. Performance metrics such as accuracy, precision, recall, and F1 score were used to evaluate the model's effectiveness. While the approach showed strong classification results, it lacked multimodal integration and explainability, which are crucial for clinical adoption [25].

Henge, Santosh Kumar, et al. proposed a 172-layer deep learning model for early diabetic retinopathy (DR) detection, using a dual-stream approach to process both color and grayscale fundus images. The model incorporates extensive preprocessing, multi-stage convolution layers, and chi-square testing for robust evaluation. It achieved a high detection accuracy of 98.1%, surpassing existing methods. While the model excels in image-based classification, it does not address explainability or incorporate multimodal clinical data—both essential for clinical transparency and decision support. [26].

Vaish, A., and Chandrakar, V. developed a DR classification model using SqueezeNet integrated with Squeeze-and-Excitation (SE) blocks to improve feature extraction and computational efficiency. Their model achieved 90.9% test accuracy, outperforming traditional architectures. The study addresses generalization issues in existing models and highlights potential for clinical deployment. However, it lacks explainability mechanisms and does not incorporate multimodal clinical data, which are vital for real-world diagnostic trust and integration [27].

Shoaib, Mohamed R., et al. proposed the DiaGAN-CNN model, combining GAN-based data augmentation with fine-tuned InceptionResNetv2 and Inceptionv3 architectures for DR diagnosis.

Validated on the ODIR dataset, the model achieved 98% accuracy and precision, outperforming prior approaches. It emphasizes early detection and improved clinical outcomes. While effective, the model does not incorporate multimodal clinical data or explainability tools, limiting transparency and broader diagnostic insight [28].

Sandhya, B. R., and N. Nalini proposed a deep neural network combining DenseNet121 and ResNet50 for classifying DR severity into five stages using high-resolution retinal images. Leveraging GPU-accelerated CNNs, ResNet50 achieved slightly better performance with 97.1% validation accuracy. The model shows strong potential for early-stage DR intervention, especially in low-resource settings. However, the approach relies solely on image data without incorporating multimodal clinical parameters. It also lacks explainability features, limiting clinical interpretability and decision support [29].

Mutawa et al. developed a novel DR detection framework by combining Multi-Scale Discriminative Robust Local Binary Pattern (MS-DRLBP) features with a hybrid CNN-RBF classifier. The model leverages stochastic modeling for efficient, non-iterative learning and uses advanced preprocessing, including noise reduction, morphological filtering, and Otsu's thresholding for vessel segmentation. It achieved 96.10% accuracy, with high precision and specificity. While effective in image-based DR detection, the model lacks multimodal clinical data integration. Additionally, it does not include explainability mechanisms, limiting its interpretability in clinical settings [30].

Sharma, N. and Lalwani, P. proposed a deep learning-based DR detection model using an Adaptive Gabor Filter (AGF) with a Chaotic Map to enhance image quality. Feature extraction combines LBP, SURF, and TEM, while classification is performed using Attention layers, DenseNet blocks, and an Optimized GRU refined via the SANGO algorithm. Grad-CAM is integrated for model explainability. The model achieved high accuracy across DiaRetDB1 (99.01%), APTOS 2019 (98.98%), and EyePACS (99.12%) datasets. While strong in performance and visual interpretability, the method does not incorporate multimodal clinical data [31].

Herrero-Tudela, Maria, et al. proposed a robust DR grading framework based on a modified ResNet-50 architecture, incorporating transfer learning, regularization, data augmentation, and early stopping. The model was evaluated on five public datasets, achieving up to 94.64% accuracy (APTOS-2019). Notably, they integrated SHAP (SHapley Additive exPlanations) to enhance explainability, providing insights into retinal features linked to DR, such as vasculature changes and peripheral lesions. The study effectively addresses class imbalance and contributes significantly to real-world DR diagnosis by merging deep learning with explainable AI, though it does not incorporate multimodal clinical parameters [32].

The study by Zedadra et al. focuses on developing a multimodal deep learning framework, DRdiag, for early detection of diabetic retinopathy (DR). It integrates fundus images with clinical parameters like disease duration to improve diagnostic accuracy. The framework combines a DenseNet121-based CNN for image analysis with a Graph Neural Network (GNN) to model relationships among patient data. Evaluated on APTOS2019 and Messidor-2 datasets, the model achieved accuracy above 97%, showcasing strong performance. While the system effectively leverages multimodal data and deep learning, it lacks explainability features, which are vital for clinical trust and adoption [33].

The study proposed by Padhmapriya et al., an integrated diabetic retinopathy (DR) detection approach that combines advanced image preprocessing (vignetting correction, denoising, adaptive contrast equalization, and color normalization) with transfer learning using models like AlexNet, GoogLeNet, ResNet, and VGGNet. The framework improves image consistency and model adaptability, addressing feature mismatch and information loss. Evaluation includes confusion matrices and computational efficiency, supporting real-world clinical use. While the model is strong in image-based analysis, it lacks multimodal data integration, which are vital for transparent and comprehensive DR diagnosis [34].

2.2 Comparison of Existing Methods

Ref No.	Authors & Year of Publication	Methods	Results	Gaps	Remarks
[1]	S. Mishra et al. 2020	DenseNet 121 on APTOS Kaggle dataset for 5-stage DR classification	Accuracy: 96.11%, Kappa: 0.8981	No multimodal data, no explainability	Focused on image-based DR detection
[2	K.	Synergic	Outperf	No	Good

Ref No.	Authors & Year of Publication	Methods	Results	Gaps	Remarks
]	Shankar et al. 2020	Deep Learning (SDL) model with histogram-based segmentation on Messidor dataset	ormed existing methods	multimodal data, no explainability	severity classification
[3]	A. Son and A Rai 2021	CNN-based fundus image classification for DR detection	High accuracy	No multimodal data no explainability	Automated, but limited for clinical use
[4]	Chava Harshitha et al. 2021	Histogram equalization + K-means segmentation + SVM & Random Forest classifiers	RF Accuracy: 96.62%	Small dataset, no multimodal data no explainability	RF better than SVM
[5]	Dolly Das et al. 2022	Review of ML & DL approaches for DR detection	Insightful review of feature extraction & DL	No proposed model, no multimodal integration, no explainability	Valuable background understanding
[6]	Zhang et al. 2022	Inception V3 model on high-res fundus images for severe DR classification	AUC: 0.968, Sensitivity: 0.925	Misclassification of lesions, no explainability or multimodal data	Improves severe DR screening
[7	K.M.	Custom	Better	No	Suitable
]	Fellah et al. 2022	CNN with preprocessing filters for DR detection	than existing methods	multimodal data, lacks explainability	for automation, not for clinics
[8]	Vij et al. 2023	DL-based segmentation of blood vessels, lesions, and optic disc/cup (Systematic review of 115 studies)	Covered 66% vessel, 36% lesion segmentation	Lack of multimodal data and explainability in reviewed studies	Highlights trends and gaps
[9]	D. Das et al. 2023	Compared 26 DL models on EyePACS dataset; EfficientNetB4 performed best	Train Acc: 99.37%, Val Acc: 79.11%	No multimodal data, lacks explainability	Useful for model benchmarking
[10]	Kalyani, G. et al. 2023	Reformed Capsule Network with convolution & capsule layers + softmax on Messidor dataset	Stage 3 Accuracy: 98.64%	No multimodal data, lacks explainability	Strong image-based performance
[11]	Rachapudi, Venubabu et al. 2023	Multi-stage pipeline: CLAHE, Hough Transform, MEM GLCM + DNN-BOA classifier	Accuracy: 98.3% (DIARETDB1), 98.9% (MESSIDOR)	No multimodal inputs; lacks explainability	Robust pipeline but limited clinical support

Ref No.	Authors & Year of Publication	Methods	Results	Gaps	Remarks
[12]	Palaniswamy, T. & Vellingiri, M. 2023	IoT-enabled deep learning (IoTDL-DRD), MFORG for segmentation, DenseNet + LSTM classifier	Outperformed traditional models	No multimodal data; lacks explainability	Innovative use of IoT but not yet clinic-ready
[13]	Raman et al. 2023	EfficientNet-B7 on Aptos2019-oversampled data + augmentation + tuning	Accuracy: 94%	No multimodal integration; lacks explainable support	Focuses on early detection via retinal layers
[14]	M.R. Shoaib et al. 2024	Transfer Learning: Inception ResNetv2, Inception V3, DiaCNN on ODIR dataset	Accuracies: 97.5%, 98.3%	Image-only input; lacks explainability	Strong model performance, limited clinical use
[15]	Malik et al. 2024	VGG-19 + preprocessing (grayscale, Gaussian filter, cropping), four-class DR classification	Accuracy: 64.5%	Moderate accuracy; no multimodal input or explainability	Needs enhancement for real-world deployment
[16]	Saranya, P. & Umamaheswari, K.M. 2024	DL model for early bright lesion detection using image preprocessing and segmentation	Accuracy: 97.54%	No multimodal input; lacks explainability	Focuses on exudate detection
[17]	Guefrachi, S. et al. 2024	ResNet152-V2 integrated in GUI for DR detection	Accuracy: 100%	Lacks multimodal data; no explainability	High performance; clinical GUI utility
[18]	Suo et al. 2024	CS-ResNet-101 + FCSAM + enhanced images + modified loss	Accuracy: 98.1%, Specificity: 99.6%	No multimodal data; no explainability	High generalization performance
[19]	Rajalingam et al. 2024	CNN-based DR stage classification using fundus images	High sensitivity & specificity (validated via ROC, AUC)	No multimodal inputs; lacks explainability	Generic CNN pipeline, good diagnostic support
[20]	Ahnaf Alavee, K. et al. 2024	Transfer Learning (ResNet50, DenseNet121, etc.) + XAI (Grad-CAM)	Accuracy: 95.27% (custom CNN)	Incorporates explainability but not multimodal data	Strong transparency emphasis
[21]	Jabbar, Ayesha et al. 2024	Deep transfer learning with VGGNet for real-time DR detection via fundus cameras	Accuracy: 97.6%	No multimodal data; lacks explainability	Useful in remote/underserved settings

Ref No.	Authors & Year of Publication	Methods	Results	Gaps	Remarks
[22]	Ahmed, Fahad et al. 2024	ResNet-18 for DR detection using Kaggle dataset	Train: 99.91%, Test: 96.65%	No multimodal or explainability	High precision; basic image-only model
[23]	Das, Uddipan et al. 2024	Autoencoder + Denoising CNN + CLAHE + Inception V3	Accuracy: 95.6%	No multimodal data; lacks explainability	Strong layered preprocessing pipeline
[24]	Sushith et al. 2025	Hybrid CNN-RNN model with attention for temporal retinal scan analysis	Accuracy: 97.5% (DRIVE), 94.04% (Kaggle), 96.9% (Eyepacs)	No multimodal data or explainability	Great for progression monitoring
[25]	Muhammad Faris et al. 2025	VGG-16 + SMOTE + Tomek Links for imbalance correction	Five-stage classification on 88,702 images	No multimodal integration or explainability	Good dataset usage; needs clinical expansion
[26]	Henge, Santosh Kumar et al. 2025	172-layer dual-stream DL model processing both color and grayscale images	Accuracy: 98.1%	No explainability or multimodal clinical data	Robust high-accuracy dual-image stream model
[27]	Vaish, A. & Chandrakar, V. 2025	SqueezeNet + SE blocks for improved feature extraction and efficiency	Accuracy: 90.9%	No multimodal integration; lacks explainability	Efficient model with better generalization
[28]	Shoaib, Mohamed R. et al. 2025	DiaGAN-CNN: GAN-based augmentation + fine-tuned Inception ResNetv2/Inceptionv3	Accuracy & Precision: 98%	Lacks multimodal data and explainability	High-performance GAN + transfer learning model
[29]	Sandhya, B. R. & N. Nalini 2025	DenseNet 121 + ResNet50 for 5-stage DR classification using high-res images	Validation Accuracy: 97.1% (ResNet50)	No explainability; no clinical parameters	Well-suited for early intervention
[30]	Mutawa, A. M. et al. 2025	MS-DRLBP features + CNN-RBF hybrid model with vessel segmentation and stochastic modeling	Accuracy: 96.10%, Precision: 96.10%, Specificity: 97.06%	No multimodal data or explainability	Efficient hybrid classifier with strong preprocessing
[31]	Sharma, N. & Lalwani, P. 2025	AGF + LBP + SURF + TEM + Attention + DenseNet + Optimized GRU + Grad-CAM	Accuracy: 99.01% (DiaRetDB1), 98.98% (APTOS), 99.12% (EyePACS)	No multimodal data	Excellent results with explainability via Grad-CAM
[32]	Herrero-Tudela, M. et al. 2025	Modified ResNet-50 + SHAP + transfer	Accuracy: Up to 94.64%; AUC-	No multimodal data	Strong real-world clinical relevance

Ref No.	Authors & Year of Publication	Methods	Results	Gaps	Remarks
		learning, tested on 5 datasets	based analysis		and explainability
[33]	Zedadra Amina et al. 2025	DRdiag: CNN (DenseNet 121) + GNN combining fundus images and disease duration	Accuracy: 0.976 (Messidor-2), 0.980 (APTOS)	Lacks explainability	Only model to use multimodal data effectively
[34]	Padhmapriya, S T. et al. 2025	Image preprocessing + transfer learning (AlexNet, GoogLeNet, ResNet, VGGNet)	Improved consistency, robustness; metrics include confusion matrix, accuracy	No multimodal data	Strong preprocessing focus; real-world diagnostic fit

Table1: Comparison of Existing Methods

2. RESEARCH GAP

Despite significant advancements in the application of deep learning for diabetic retinopathy (DR) detection, critical gaps remain in the areas of explainability and multimodal data integration. Addressing these limitations is essential for improving clinical trust, diagnostic precision and real-world applicability of AI-driven systems.

1. Lack of Multimodal Data Integration:

Most existing approaches rely solely on retinal fundus images, neglecting other clinically relevant parameters such as blood sugar, HbA1c levels, blood pressure, and duration of diabetes. While Zedadra et al. made an initial attempt by incorporating disease duration along with image analysis, such multimodal strategies are rare. This image-centric trend limits the potential for holistic patient profiling and risk prediction.

2. Lack of Explainability in AI Models:

Although several deep learning models have achieved high accuracy in DR classification, explainability remains a significant limitation. Only a few studies such as Sharma et al. (using Grad-CAM) and Herrero-Tudela et al. (using SHAP) have attempted to provide visual or interpretative insights into the model's decisions. However, the majority of the reviewed literature either overlooks explainable AI (XAI) techniques entirely or incorporates them in a limited manner.

3. Lack of Deployment on Embedded Hardware Platforms:

Although many AI models for diabetic retinopathy detection show high accuracy in lab settings, very few are deployed on embedded hardware for real-time use. This limits their practical application, especially in point-of-care scenarios and remote or low-resource settings. Bridging this gap is essential for making AI-driven DR screening more accessible and clinically relevant.

3. PROBLEM STATEMENT

Despite several DL models for DR detection, most are image-only and act like black boxes. This makes it difficult for doctors to understand or trust them. Clinical data like blood sugar, HbA1c, and blood pressure are often ignored.

Hence, there is a need to develop an interpretable, multimodal deep learning framework that not only detects DR with high accuracy, but also provides clinically meaningful explanations and integrates relevant patient health data for improved diagnostic precision and early intervention.

4. RESEARCH OBJECTIVES

1. **To integrate multimodal data** (retinal images and clinical parameters such as blood sugar, HbA1c, and blood pressure) into a hybrid deep learning model for enhanced diabetic retinopathy risk prediction and grading.
2. **To develop explainable AI tools** that help doctors understand and trust the model’s predictions, thereby supporting better clinical decision-making in diabetic retinopathy diagnosis.
3. **To deploy** the proposed multimodal and explainable AI model on an **embedded hardware platform.**

5. AN OVERVIEW OF PROPOSED METHODOLOGY

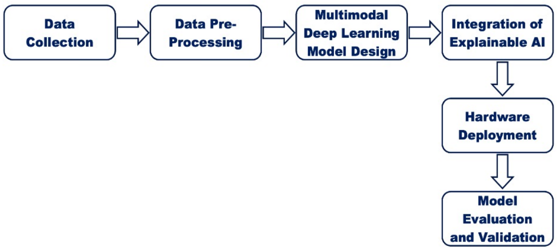

Fig. 4: Block diagram of Proposed Methodology

<u>**Key Stages**</u>**:**

1. Data Collection:

▪ Image Data: Acquire retinal fundus images from public datasets (e.g., APTOS, Messidor, EyePACS).

▪ Clinical Parameters: Collect associated patient data such as:

o HbA1c

o Blood pressure

o Blood sugar levels

o Duration of diabetes

2. Data Preprocessing:

▪ Preprocessing Techniques:

o Images: Resizing, normalization, contrast enhancement, noise reduction.

o Clinical Data: Missing value imputation, normalization, outlier handling.

3. Multimodal Deep Learning Model Design:

▪ Image Branch: Use a Convolutional Neural Network (CNN) like DenseNet or EfficientNet to extract spatial features from fundus images.

▪ Clinical Data Branch: Use a Deep Neural Network (DNN) or Graph Neural Network (GNN) to process structured clinical features.

▪ Fusion Strategy:

o Combine outputs from both branches using concatenation or attention-based fusion.

o Feed fused features into fully connected layers for DR stage classification or risk grading.

4. Integration of Explainable AI (XAI):

▪ Incorporate explainability methods:

o Grad-CAM: For highlighting lesion regions on retinal images.

o SHAP/LIME: For attributing clinical feature contributions to predictions.

o Output visual and textual explanations to enhance doctor trust and diagnostic insight.

5. Hardware Deployment:

▪ Deploy on embedded platforms like:

o Raspberry Pi Board

o Pynq Z2 Board

6. Model Evaluation and Validation:

▪ Use evalution metrics:

o Accuracy, Precision, Recall, F1-score

o Quadratic Kappa Score

o Explainability metrics (e.g., clinical relevance score, trust scores)

▪ Perform:

o Cross-dataset validation

6. EXPECTED OUTCOMES

1. **Accurate DR Detection:**

Development of a hybrid deep learning model for improved diabetic retinopathy detection and grading using both retinal images and clinical data.

2. **Explainable AI:**

Integration of explainability tools to enhance clinical trust by visualizing and interpreting model decisions.

3. **Hardware Deployment:**

Deployment of the proposed model on embedded hardware platform.

7. REFERENCES

[1] S. Mishra, S. Hanchate, and Z. Saquib, "Diabetic Retinopathy Detection using Deep Learning," in Proc. 2020 Int. Conf. Smart Technol. Comput., Electr. Electron. (ICSTCEE), Bengaluru, India, 2020, pp. 515–520, doi: 10.1109/ICSTCEE49637.2020.9277506.

[2] K. Shankar, A. R. W. Sait, D. Gupta, S. K. Lakshmanaprabu, A. Khanna, and H. M. Pandey, "Automated detection and classification of fundus diabetic retinopathy images using synergic deep learning model," Pattern Recogn. Lett., vol. 133, pp. 210–216, 2020, doi: 10.1016/j.patrec.2020.02.026.

[3] A. Soni and A. Rai, "A Novel Approach for the Early Recognition of Diabetic Retinopathy using Machine Learning," in Proc. 2021 Int. Conf. Comput. Commun. Informat. (ICCCI), Coimbatore, India, 2021, pp. 1–5, doi: 10.1109/ICCCI50826.2021.9402566.

[4] C. Harshitha, A. Asha, J. L. S. Pushkala, R. N. S. Anogini, and K. C., "Predicting the Stages of Diabetic Retinopathy using Deep Learning," in Proc. 2021 6th Int. Conf. Inventive Comput. Technol. (ICICT), Coimbatore, India, 2021, pp. 1–6, doi: 10.1109/ICICT50816.2021.9358801.

[5] D. Das, S. K. Biswas, and S. Bandyopadhyay, "A critical review on diagnosis of diabetic retinopathy using machine learning and deep learning," Multimedia Tools Appl., vol. 81, no. 18, pp. 25613–25655, 2022.

[6] X. Zhang et al., "Automated detection of

severe diabetic retinopathy using deep learning method," Graefe's Arch. Clin. Exp. Ophthalmol., vol. 260, no. 3, pp. 849–856, 2022.

[7] K. M. Fellah, S. Tigane, and L. Kahloul, "Diabetic retinopathy detection using deep learning," in Proc. Int. Symp. Model. Implementation Complex Syst., Cham: Springer, 2022, pp. 234–246.

[8] R. Vij and S. Arora, "A systematic review on diabetic retinopathy detection using deep learning techniques," Arch. Comput. Methods Eng., vol. 30, no. 3, pp. 2211–2256, 2023.

[9] D. Das, S. K. Biswas, and S. Bandyopadhyay, "Detection of diabetic retinopathy using convolutional neural networks for feature extraction and classification (DRFEC)," Multimedia Tools Appl., vol. 82, no. 19, pp. 29943–30001, 2023.

[10] G. Kalyani, B. Janakiramaiah, A. Karuna, and L. V. N. Prasad, "Diabetic retinopathy detection and classification using capsule networks," Complex Intell. Syst., vol. 9, no. 3, pp. 2651–2664, 2023.

[11] V. Rachapudi et al., "Diabetic retinopathy detection by optimized deep learning model," Multimedia Tools Appl., vol. 82, no. 18, pp. 27949–27971, 2023.

[12] T. Palaniswamy and M. Vellingiri, "Internet of Things and Deep Learning Enabled Diabetic Retinopathy Diagnosis Using Retinal Fundus Images," IEEE Access, vol. 11, pp. 27590–27601, 2023, doi: 10.1109/ACCESS.2023.3257988.

[13] D. R. Raman, S. Nishanthi, and P. Babysha, "Diagnosis of Diabetic Retinopathy by using EfficientNet-B7 CNN Architecture in Deep Learning," in Proc. 2023 Int. Conf. Sustain. Comput. Smart Syst. (ICSCSS), Coimbatore, India, 2023, pp. 430–435, doi: 10.1109/ICSCSS57650.2023.10169453.

[14] M. R. Shoaib et al., "Deep learning innovations in diagnosing diabetic retinopathy: The potential of transfer learning and the DiaCNN model," Comput. Biol. Med., vol. 169, p. 107834, 2024.

[15] T. Malik, V. Nandal, and P. Garg, "Deep Learning-Based Classification of Diabetic Retinopathy: Leveraging the Power of VGG-19," in Proc. 2024 Int. Conf. Emerg. Innov. Adv. Comput. (INNOCOMP), Sonipat, India, 2024, pp. 645–651, doi: 10.1109/INNOCOMP63224.2024.00111.

[16] P. Saranya and K. M. Umamaheswari, "Detection of exudates from retinal images for non-proliferative diabetic retinopathy detection using deep learning model," Multimedia Tools Appl., vol. 83, no. 17, pp. 52253–52273, 2024.

[17] S. Guefrachi, A. Echtioui, and H. Hamam, "Automated diabetic retinopathy screening using deep learning," Multimedia Tools Appl., vol. 83, no. 24, pp. 65249–65266, 2024.

[18] Y. Suo, Z. He, and Y. Liu, "Deep learning CS-ResNet-101 model for diabetic retinopathy classification," Biomed. Signal Process. Control, vol. 97, p. 106661, 2024.

[19] R. S, K. S, B. U, and M. Chouksey, "Revolutionizing Diabetic Retinopathy Recognition with Advanced Deep Learning Techniques," in Proc. 2024 Int. Conf. Comput., Sci. Commun. (ICCSC), Ghaziabad, India, 2024, pp. 1–5, doi: 10.1109/ICCSC62048.2024.10830327.

[20] K. A. Alavee et al., "Enhancing Early Detection of Diabetic Retinopathy Through the Integration of Deep Learning Models and Explainable Artificial Intelligence," IEEE Access, vol. 12, pp. 73950–73969, 2024, doi: 10.1109/ACCESS.2024.3405570.

[21] A. Jabbar et al., "Deep transfer learning-based automated diabetic retinopathy detection using retinal fundus images in remote areas," Int. J. Comput. Intell. Syst., vol. 17, no. 1, p. 135, 2024.

[22] F. Ahmed, A. Fatima, M. Mamoon, and S. Khan, "Identification of the Diabetic Retinopathy Using ResNet-18," in Proc. 2024 2nd Int. Conf. Cyber Resilience (ICCR), Dubai, UAE, 2024, pp. 1–6, doi: 10.1109/ICCR61006.2024.10532925.

[23] U. Das, J. Haloi, P. Hassam, S. Kashyap, and M. Sharma, "Detection Of Diabetic Retinopathy Using Deep Learning Algorithms," in Proc. 2024 6th Int. Conf. Energy, Power Environ. (ICEPE), Shillong, India, 2024, pp. 1–5, doi: 10.1109/ICEPE63236.2024.10668922.

[24] M. Sushith, A. Sathiya, V. Kalaipoonguzhali, and V. Sathya, "A hybrid deep learning framework for early detection of diabetic retinopathy using retinal fundus images," Sci. Rep., vol. 15, no. 1, p. 15166, 2025.

[25] M. Faris, F. Sohail, C. Pepe, M. F. Ali, and S. M. Zanoli, "Detection of Diabetic Retinopathy Using Deep Learning," in Proc. 2025 26th Int. Carpathian Control Conf. (ICCC), Slovakia, 2025, pp. 1–4, doi: 10.1109/ICCC65605.2025.11022858.

[26] S. K. Henge et al., "Detection of Diabetic Retinopathy Using a Multi-Decision Inception-ResNet-Blended Hybrid Model," IEEE Access, vol. 13, pp. 8988–9005, 2025, doi: 10.1109/ACCESS.2024.3525154.

[27] A. Vaish and V. Chandrakar, "Improving Classification Accuracy For Diabetic Retinopathy With Deep Learning," in Proc. 2025 IEEE Int.

Conf. Interdiscip. Approaches Technol. Manage. Social Innov. (IATMSI), Gwalior, India, 2025, pp. 1–6, doi: 10.1109/IATMSI64286.2025.10985144.

[28] M. R. Shoaib et al., "Revolutionizing diabetic retinopathy diagnosis through advanced deep learning techniques: Harnessing the power of GAN model with transfer learning and the DiaGAN-CNN model," Biomed. Signal Process. Control, vol. 99, p. 106790, 2025.

[29] S. B. R and N. N, "Leveraging Retinal Images for Detection and Classification of Diabetic Retinopathy Using DenseNet121 and ResNet50," in Proc. 2025 Int. Conf. Comput. Sustain. Intell. Future (COMP-SIF), Bangalore, India, 2025, pp. 1–6, doi: 10.1109/COMP-SIF65618.2025.10969964.

[30] A. M. Mutawa, G. R. Hemalakshmi, N. B. Prakash, and M. Murugappan, "Randomization-Driven Hybrid Deep Learning for Diabetic Retinopathy Detection," IEEE Access, vol. 13, pp. 38901–38913, 2025, doi: 10.1109/ACCESS.2025.3546359.

[31] N. Sharma and P. Lalwani, "A multi model deep net with an explainable AI based framework for diabetic retinopathy segmentation and classification," Sci. Rep., vol. 15, no. 1, p. 8777, 2025.

[32] M. Herrero-Tudela et al., "An explainable deep-learning model reveals clinical clues in diabetic retinopathy through SHAP," Biomed. Signal Process. Control, vol. 102, p. 107328, 2025.

[33] A. Zedadra, O. Zedadra, M. Y. Salah-Salah, and A. Guerrieri, "Graph-Aware Multimodal Deep Learning for Classification of Diabetic Retinopathy Images," IEEE Access, vol. 13, pp. 74799–74810, 2025, doi: 10.1109/ACCESS.2025.3564529.

[34] S. T. Padhmapriya, M. Sharmila, A. K. Shree, and J. B. Premkumar, "Explainable Deep Learning Models for Clinical Decision Support in Diabetic Retinopathy Detection," in Proc. 2025 IEEE Int. Conf. Interdiscip. Approaches Technol. Manage. Social Innov. (IATMSI), vol. 3, pp. 1–6, 2025.