Svedberg Open
DISSEMINATION OF KNOWLEDGE

International Journal of Artificial Intelligence and Machine Learning

Publisher's Home Page: https://www.svedbergopen.com/

Research Paper Open Access

Identification And Risk Level Prediction Of Diabetic Retinopathy Using Transfer Learning With Novel Vision Transformer And Grad-Cam Explainable Ai

Anita B. Dombale[1*] , Dr. Premanand P. Ghadekar[2]

[1]Department of Computer Engineering, Vishwakarma Institute of Technology, Savitribai Phule Pune University, Pune, Maharashtra, 411037, India, Email: bhanuseanita@gmail.com

[2]Dept of Computer Science & Engineering (Artificial Intelligence & Machine Learning), Vishwakarma Institute of Technology, Savitribai Phule Pune University, Pune, Maharashtra, 411037, India

*Corresponding Author

Abstract

Diabetic retinopathy (DR) is one of the leading causes of visual impairment that requires correct prediction and early detection. The system suggested herein introduces a hybrid deep learning model that combines statistical feature analysis with a modified version of the Vision Transformer (ViT) architecture to be used in strong classification. Models that had been pre-trained to identify Diabetic Retinopathy were tested and the model with highest performance was integrated into a ViT-inspired architecture to achieve prediction. The improved version comprises of an optimized multi-head attention block and an enhanced transformer block to achieve a high-quality feature extraction and classification accuracy. used explainable AIs, such as a variant of Gradient-weighted Class Activation Mapping (Grad-CAM) tailored to Vision Transformers, to facilitate transparency and make informed decisions. Key performance measures were compared and the assessment revealed an evaluation of 83 % accuracy.

Keywords: Deep Learning, Vision Transformer, Inceptionv3, Transfer Learning, Grad-CAM

1. Introduction

One of the most significant microvascular problems of diabetes is diabetic retinopathy (DR). impacts the retina and is a serious possible danger to sight when not treated [1][2]. It has been forecasted that as the prevalence of diabetes increases across the globe, the burden of DR in the world will increase and therefore early detection and precise risk stratification are particularly significant in the clinical treatment and the prevention of blindness. This is a progressive disease that is due to the damage of the retinal blood vessels and the subsequent vision loss due to mechanisms that involve the macular edema, hemorrhages and neovascularization. Conventionally, diagnosis and evaluation of DR rely primarily on the manual interpretation of retinal fundus images by manual ophthalmologists. This method is very time consuming, subjective and prone to variations in observers. The emerging technologies in Artificial Intelligence have introduced significant opportunities to automate the process of identifying DR and predicting the level of risk [3][4]. These systems would actually go a long way in assisting to reduce the diagnostic workload, improve the accuracy and to assist in offering the timely intervention, especially in those places where specialist care is scarce [5].being based on retinal fundus images, the paper would establish a framework on automated detection of DR and prediction of the level of risk. The proposed methodology employs feature-based risk stratification and interpretable modeling, which is based on the strength of Vision Transformers (ViTs) as well as traditional designs built on transfer learning. The contributions of this work include the following ones: Comprehensive Feature Integration: to enhance the strength of the model and provide a complete consideration. Optimization Architecture: Network optimization of ViTs to use high-resolution

fundus image data. Explainable artificial intelligence (XAI) of Clinical Relevance: It is necessary to ensure that the results of the model can be deciphered and their sense connected to clinical observations to support the decision-making process[27]. Checking Balanced datasets: The Balanced datasets to minimise the bias and improving a greater degree of dependability of the model in diverse degrees of risk. .Proposed System to enhance the accuracy of DR detection and enable personalized care by predicting the level of risk. The study has a prospective to fill the gap between the state-of-the-art AI and clinical practice, which is a part of the overall quest to address the problem of diabetic complications by diagnosing and assessing risks at an early stage.

2. Literature Review

Using convolutional neural networks (CNNs), the authors S. Mishra et al. (2020) investigate an automated method for identifying and assessing DR phases. The study achieved an accuracy of approximately 85% but encountered challenges like uneven dataset distribution and the absence of clarity in the outcomes of the DL model[6].The Authors P. Bidwai et al. (2023) utilize CNN-based deep learning models to identify DR from retinal fundus images.It reached up to around 87% accuracy. However, it needed high-quality images and large data sets for getting the best performance[7]. The OptiDex method was recently introduced by Lalithadevi, B et al. (2023) for detecting DR and grading its severity using XAI. Nevertheless, the method results reached an impressive accuracy of around 90%, though with high computational power to process the explainability models[8]. W. K. Wong et al.(2022) demonstrates the use of transfer learning with pre-trained networks like ResNet and VGG for identification and grading of DR severity. got an accuracy of around 89% but was limited to binary classification and required further fine-tuning for multiclass problems[9]. Naz, H. et al.(2024) integrates supervised and unsupervised learning techniques to improve detection accuracy. The hybrid approach achieved approximately 88% accuracy but increased complexity, raising concerns about real-world applicability[10]. K. K. Y. Tiong et al. researchers combined predictions from multiple CNN architectures to enhance accuracy. This ensemble-based strategy obtained high accuracy at 92% with very high increased computation cost[11][25]. Saket S. Chaturvedi et al.(2019), provides an in-depth survey of various computer-aided systems for detection of DR but, again does not have implementation and real validation, as the article is review based[12]. In a study Hossein Shakibania, Sina Raoufi et al.released in 2023, a two branch deep network is proposed for DR detection and grading. The algorithm attained a promising accuracy close to 91%; however, it failed to generalize with diverse datasets when applied for testing[13]. S. Suganyadevi et al.2020 created an automated pipeline that identified DR based on fundus images using CNNs. The accuracy of the work remained approximated at about 86% although it was sensitive to the image and dataset variability[14]. Finally, Eduard Popescu et al. (2023) applied ensemble classifiers to identify and categorize the severity of DR. The approach attained a performance of approximately 88 percent, yet it was difficult to optimize the hyperparameters among classifiers[15][26]. The need to have a robust, generalized model is highlighted in these studies.

3. Proposed Methodology

Fig.1 represents the suggested system. To obtain precise forecasts on the level of diabetic retinopathy risks. Parameters like DR(Diabetic Retinopathy), AH/HTN(Arterial Hypertension / Hypertension), DN(Diabetic Nephropathy) as well as other medical parameters that are very important in the development and course of diabetic complications were included in these features. Augmentation is done to generate a balanced dataset. And after Augmentation, Train folder has 2829 images of 4 types of classes and Test folder has 856 images of 4 types of classes.

A vision transformer (ViT)-inspired architecture was then created with significant architectural changes to this task alongside InceptionV3. It also modulated the transformer layers, attention mechanisms, and feature extraction methods that could deal with these complexities on retinal fundus images and the various patterns that are characteristic of diabetic retinopathy. The modified ViT model was trained using the risk-level-specific image to ensure that the model learnt subtle features that are associated with the varying levels of diabetic retinopathy.

The architecture combines InceptionV3 and Vision Transformer (ViT) with the help of which diabetic retinopathy risk levels are predicted well. The first 30 layers are considered as the backbone, and in the course of early spatial feature extraction, inceptionV3 is employed, and the remaining 30 layers are fine-tuned to identify patterns in the dataset. The extracted features are fed to eight transformer layers, which employ multi-head self-attention (8 heads with key dimension of 128) and MLP layers (256 units with ReLU activation) to identify long-range dependencies. There are layer normalization and residual connections that are used to stabilize and prevent vanishing gradients. Subsequent to transformation, Global Average Pooling (GAP) is conducted twice to down sample features, and then a 512-neuron dense layer with batch normalization, followed by a softmax layer to classify risk levels. This blended architecture provides both strong local feature extraction and global contextual awareness.

The findings indicated that the Proposed system could effectively predict the risk levels that were necessary in real-world application and had achieved these results using other datasets for verification of the generalization ability and robustness of the model. Comparing the performance yields the performance metric of the proposed ViT-inspired model to overcome previous approaches on diabetic retinopathy risk prediction.

Apart from the ViT-Inspired model, proposed system incorporates a novel Explainable Artificial Intelligence (XAI**)** approach. This enhances interpretability in terms of insights regarding the decision-making process behind the model and makes it more transparent for medical analysis.

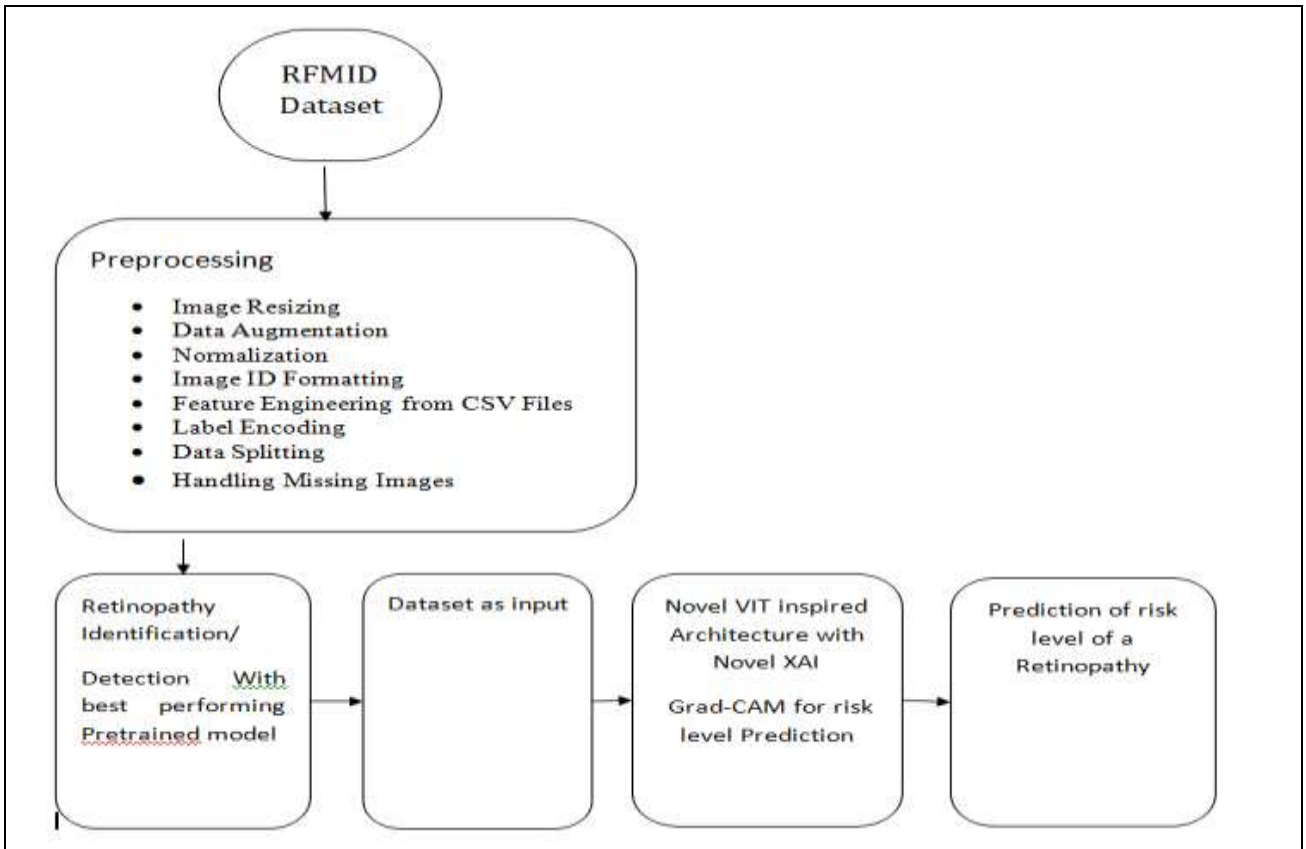

Fig. 1 Identification and Risk level Prediction of Diabetic Retinopathy from retinal fundus images with Statical features using Transfer Learning with novel Vision Transformer and Grad Cam Explainable AI

Accuracy Results Comparison for Diabetic Retinopathy Identification

Fig.2 shows , Among the evaluated models, DenseNet121, Inception V3, and VGG16 demonstrate superior performance, achieving balanced precision, recall, and an F1 Score.

Inception V3 is particularly robust, showing consistency across both classes, making it the most reliable models for use.

Classwise Evaluation

Fig.3 shows the graph of precision, Recall and F1 score values for all 4 models i.e. Inception V3,Resnet152,VGG16 and Densenet121.

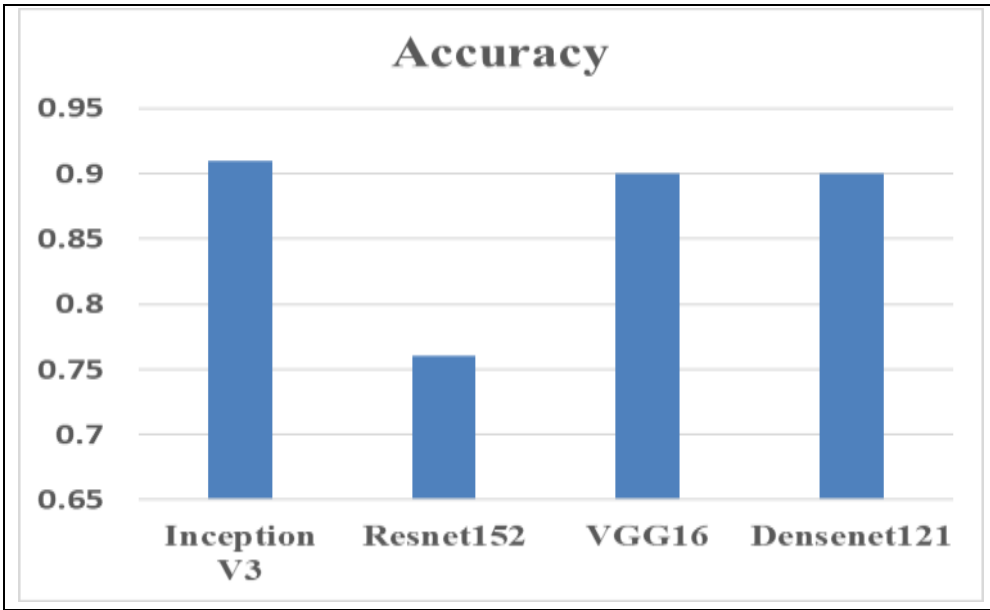

Fig.2 Accuracy Results Comparison

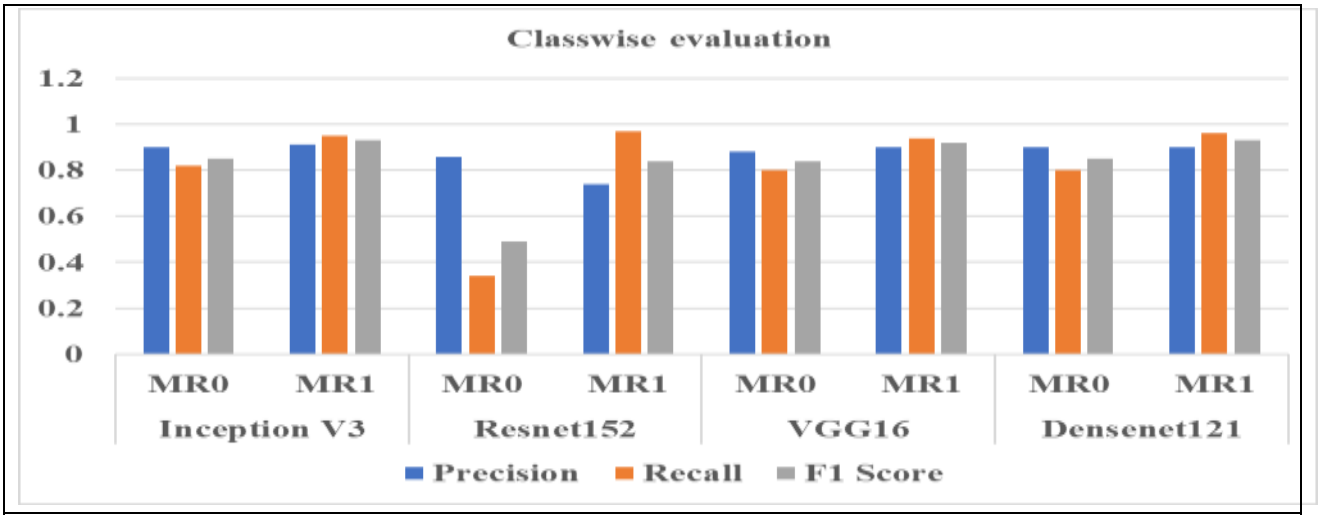

Fig.3 Class wise Evaluation

Predicting Risk level of Retinopathy

Using InceptionV3 and inspired by a Vision Transformer (ViT), a statistical feature analysis model was developed for the prediction of diabetic retinopathy risk levels. Risk levels (0-4) were defined using medical indicators, with images categorized under them. Data augmentation balanced the dataset to comprise 2,829 training images and 856 test images. An architecture inspired by the Vision Transformer (ViT) was then developed with notable architectural modifications tailored to this task in addition to InceptionV3. It optimized the transformer layers, attention mechanisms, and feature extraction methods that were to accommodate such complexities of retinal fundus images and various patterns which are used to define diabetic retinopathy. To obtain the desired results of determining the different stages of diabetic retinopathy, the modified ViT was trained, and the findings showed that the model was able to accurately predict the risk levels that were needed

in the real-world application and had acquired the results with the help of other datasets to validate the results and the robustness of the model.

Gradient-weighted Class Activation Mapping (Grad-CAM) and Overlay Visualization

The Grad-CAM [16][17] technique is used to make deep learning models more interpretable by providing visual explanations for their predictions. This is particularly useful in medical imaging applications [18][19], such as retinal disease detection, as it draws attention to the regions of the image that contribute most to in decision [20][21].

4. Mathematical Model Of Proposed System For Risk Level Prediction

1. Feature Extraction via InceptionV3

Given an input image X of shape (H, W,3), InceptionV3 extracts spatial features through convolutional layers:

$$F = f_{inceptionV3}(X) \quad (1)$$

where F is the feature map of shape (h, w, d).

 H, W, 3 → Original input image dimensions:

- H = Height of the input image (e.g., 224 pixels)
- W = Width of the input image (e.g., 224 pixels)
- 3 = Number of channels (for RGB images)

 h, w, d → Output feature map dimensions after passing through the CNN (e.g., InceptionV3):

- h = Reduced height after convolution and pooling
- w = Reduced width after convolution and pooling
- d = Depth of the feature map (number of filters or channels)

2. Multi-Head Self-Attention (MHSA) in Transformer Block

Each input feature X undergoes self-attention through the following steps:

1. Compute query Q, key K, and value V:

$$Q = XW_Q, K = XW_K, V = XW_V \quad (2)$$

where W_Q , W_K , W_V are learnable weight matrices.

2. Compute attention scores using the scaled dot-product attention:

$$A = softmax\left(\frac{QK^T}{\sqrt{d_k}}\right) \quad (3)$$

Where d_k is the key dimension and T represents the number of tokens..

3.Compute the final attention-weighted output:

$$Z = AV \quad (4)$$

Where Z is Final Output of Self-Attention, A is Attention Matrix (Weights) and V represents the Value vector or matrix.

4.Apply residual connection:

$$X' = Z + X \quad (5)$$

3. Feed-Forward Network (MLP Block)

A two-layer MLP is applied to the attention output:

$$X'' = \sigma(X'W_1 + b_1)W_2 + b_2 \quad (6)$$

Where W_1, W_2 are learnable weights, b_1, b_2 are biases, and σ is the ReLU activation function.

4. Final Classification

The processed feature representation $X^{'}$ is pooled using Global Average Pooling (GAP) and passed through dense layers:

$$y = softmax(W_d\ X^{'}) + b_d \quad (7)$$

where W_d and b_d represent the classification layer parameters.

This model effectively combines CNN feature extraction, transformer-based self-attention, and MLP classification, optimizing for both spatial and global contextual information.

Novelty in Mathematical Model

1. Hybrid ViT-InceptionV3 Feature Extraction Model

Unlike standard ViT models that process raw image patches, proposed approach first extracts deep spatial features using InceptionV3 before feeding them into transformer blocks [22] [23]. This reduces computational complexity while preserving local spatial information, making ViT more efficient for medical image analysis [24].

- **Mathematical Novelty:**

  Instead of raw pixel-based input to ViT, defined the input as:

- $F = f_{inceptionV3}(X)$ (8)

where F is a pre-processed feature map. This hybrid approach leverages CNNs for local feature extraction and Transformers for global attention learning, which is uncommon in standard ViT implementations.

2. Improved Multi-Head Self-Attention (MHSA) Residual Feature Aggregation.

Composed transformer block is not only applying Multi-Head Self-Attention (MHSA) but also instead incorporates residual feature learning at various levels. This guarantees gradient stability and information retention that is a major weakness of pure ViT models [25].

Mathematical Novelty:

Added skip connection and feature fusion instead of the normal X' =MHSA(X)

$$A = softmax\left(\frac{QK^T}{\sqrt{d_k}}\right) \quad (9)$$

$$Z = AV \quad (10)$$

$$X' = Z + X \quad (11)$$

$$X'' = \sigma(X'W_1 + b_1)W_2 + b_2 + X' \quad (12)$$

The additional feature X' guarantees more extensive propagation of the features and provides the model with the capability to acquire long-range dependencies without vanishing gradients.

3. **Adaptive MLP Growth to Improved Feature Representation.**

The Multi-Layer Perceptron (MLP) block in most architectures of the transformer architecture has a fixed expansion ratio. Your model is dynamically adjusted, with the hidden layer dimension being the given twice input size, to allow maximum features to be learned.

Mathematical Novelty:

Rather than fixed MLP, proposed system proposes dynamic expansion of hidden.

$$X'' = \sigma(X'W_1 + b_1)W_2 + b_2 \quad (13)$$

$$\dim(W_1) = 2 \times \dim(X') \quad (14)$$

This hierarchically learns features with an adaptive scale, which enhances feature learning.

4**. Robust Representation Learning by Double Global Average Pooling.**

Instead of applying Global Average Pooling (GAP) only at the CNN output, apply it twice:

1. Following InceptionV3 - To decrease the number of features and increase ViT computational efficiency.
2. After Transformer Blocks - To sum up the last acquired characteristics before classification.

- **Mathematical Novelty:**

Standard GAP:

- $X_{GAP} = \frac{1}{N}\sum_{I=1}^{N} X_i$ (15)

Suggested model will use GAP twice so that the spatial and contextual features are represented better:

$X_{GAP} = \frac{1}{N}\sum_{I=1}^{N} X_{Vit,i}$ (16)

This minimizes overfitting and makes local as well as global features contribute equally to classification.

5. Subtuning of Final 30 Layers in InceptionV3 to Specialized Feature Extraction.

Most transfer learning methods either freeze the full CNN, or fine-tune the entire model, whereas Proposed method determines which 30 last layers of InceptionV3 to strategically fine-tune.

The last 30 layers are only optimized with the proposed model.

Mathematical Novelty:

Instead of full fine-tuning:

$\theta = trainable \quad \forall L$ (17)

model optimizes only the last 30 layers:

$\theta = trainable \quad \forall L \, \epsilon [L_{total} - 30, \; L_{total}]$ (18)

This enhances domain adaption without overfitting.

The proposed system is an effective combination of CNNs and transformers, further develops the residual learning, expands features adaptively, and implements innovative double-pooling processes, thus, being more efficient and effective in the risk level classification of diabetic retinopathy.

5. Result And Discussion

To have a comparative point of view, as well as to achieve the answer to whether the provided ViT-InceptionV3 model is efficient or not, some of the key performance measures (KPMs) are taken into account, and they are classification report, confusion matrix, accuracy graph, sensitivity graph, specificity graph, and explainable AI (XAI) visualizations. The results provide comprehensive explanation of the model skills in termination of rightful categorization of risk amount of diabetic retinopathy.

Classification Report

The model is an elaborate disaggregation of the classification report as shown by Table 1. This will indicate the level to which the model would discriminate the different degree of risk of diabetic retinopathy. The proposed system has a 83 percent accuracy.

Table 1: Classification Report of Risk Level Prediction

	precision	recall	f1-score	support
class0	0.79	0.83	0.81	226
class1	0.77	0.74	0.75	210
class2	0.89	0.83	0.86	210
class3	0.90	0.94	0.92	210
accuracy			0.83	856
macro avg	0.84	0.83	0.83	856
weighted avg	0.83	0.83	0.83	856

2. Confusion Matrix

A larger diagonal dominance (when the correct predictions are clustered around the diagonal) is the evidence of high classification performance. Fig.4 depicts the confusion matrix that indicates strongly classified performance diagonally.

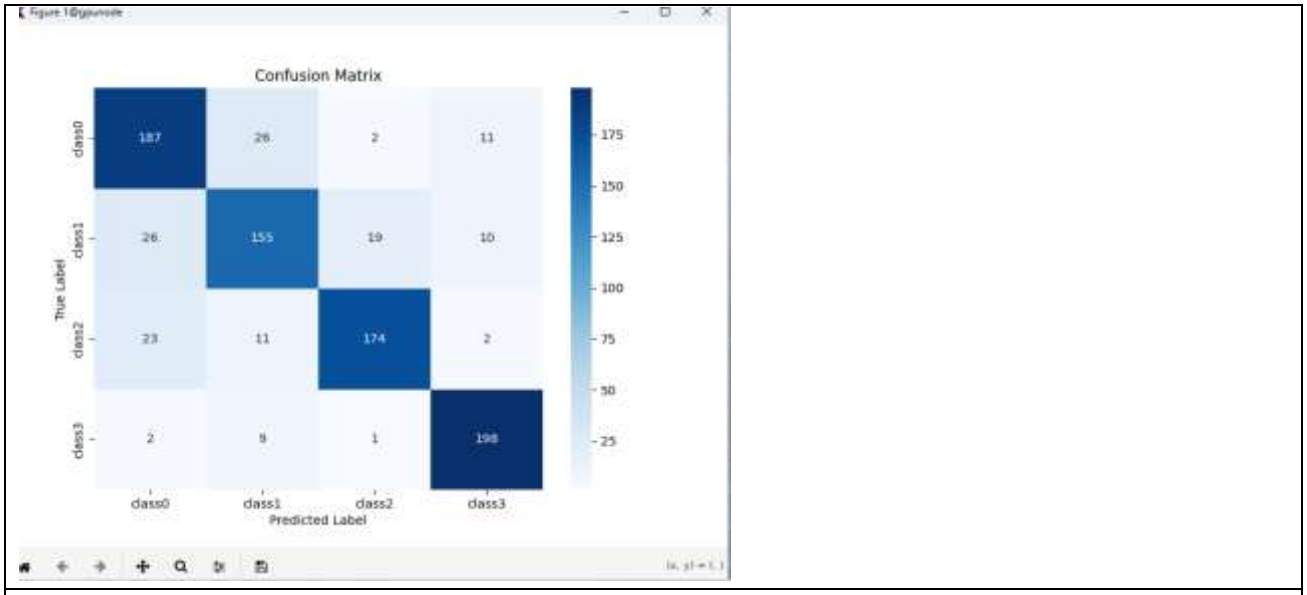

Fig.4 Confusion Matrix

3. Accuracy and Loss Graph

As it appears in Fig.5, The accuracy curve of the proposed system is steadily improving with the final accuracy of the model being high, which verifies the effectiveness of the proposed model in the classification of risk levels.

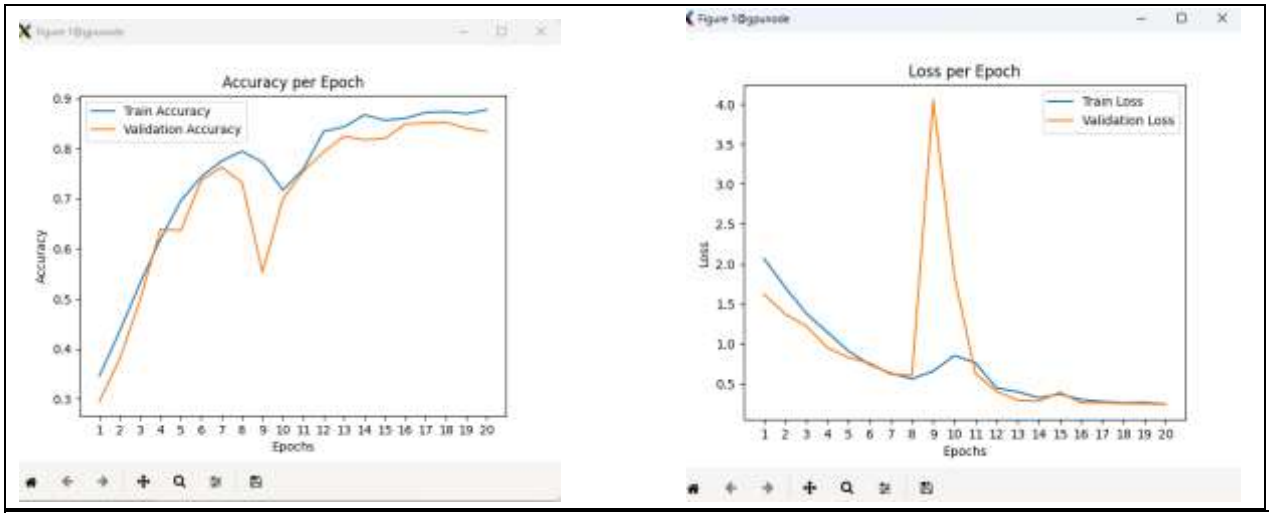

Fig.5 Accuracy and Loss Graph

4. Sensitivity (Recall) and Specificity Graph

The Sensitivity and Specificity Graph of proposed system is obtained as presented in Fig.6. It is also sensitive with severe risk levels and thus the high-risk cases on diabetic retinopathy are not ignored. The confirmations of specificity indicate that the model is able to distinguish the cases of varying levels of severity effectively such that, it does not produce unnecessary alarms in cases where there is minimal risk, and it is highly accurate in cases where the severity is high.

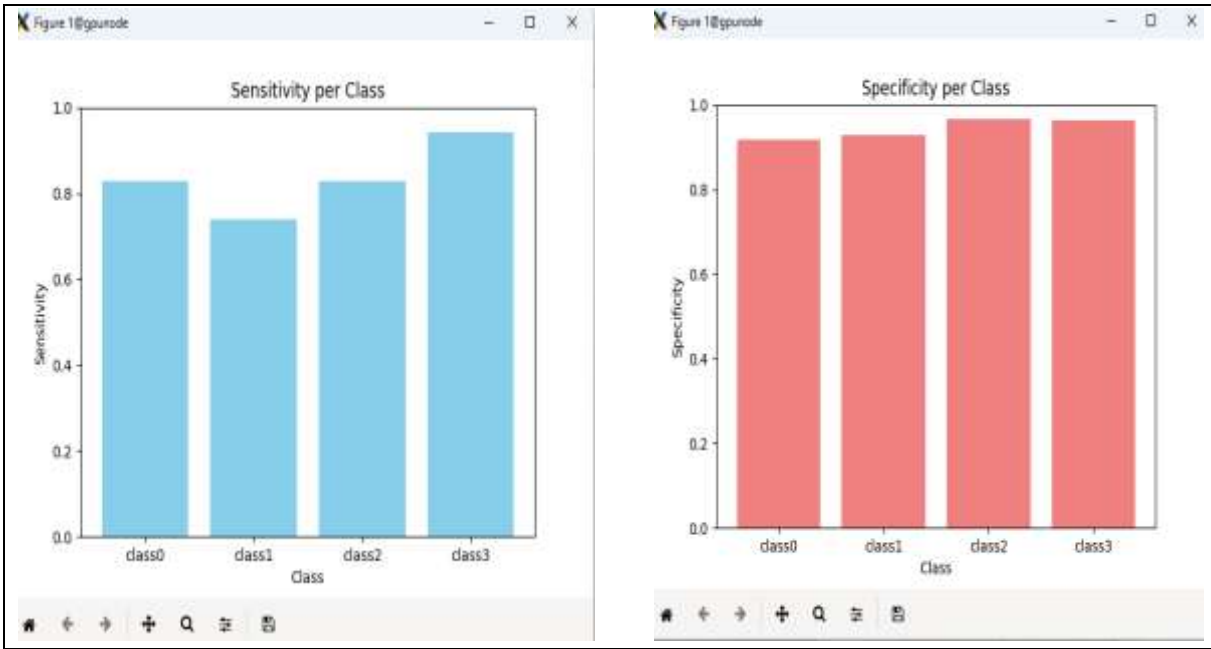

Fig.6 Sensitivity and Specificity Graph

5. Explainable AI (XAI) Visualizations

Grad-CAM is also an explainable AI (XAI) method that was used to outline what essential parts of retinal fundus images contributed to the model. The XAI output is interpretable, and it does not mean that the model is making decisions by depending on the irrelevant parts of the image, but significant retina abnormalities.

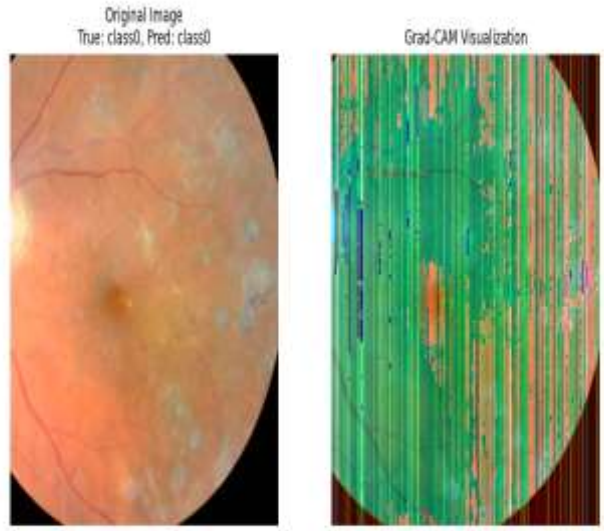

Fig.7 Grad-CAM Visualization 1

Grad-CAM Visualization Analysis

Fig.7 contains two subfigures, supplying information on the model's classification result for a retinal fundus image.

1. Original Image (Left)

The left subfigure of Fig.7 represents the retinal fundus image that was input to the deep learning model. The text label above the image indicates:

- True Class: Class 0 (the actual label assigned to this image).
- Predicted Class: Class 0 (the model correctly classified the image).

Since the prediction matches the ground truth, this indicates a correct classification, meaning the model successfully identified relevant features for this category. The image exhibits vascular structures and retinal features that were used by the model for classification.

2. Grad-CAM Visualization (Right)

The right subfigure for Fig.7 displays the Grad-CAM visualization, which highlights the regions that influenced the model's prediction.

- The green and orange overlay indicates the importance of different areas in decision-making.
- The central and peripheral regions appear to have received higher attention, suggesting that the model focused on vessel structures, lesions, or other retinal abnormalities.
- However, the striped pattern in the heatmap suggests possible diffusion in attention, which might indicate that the model is not fully focusing on localized pathological areas but rather a broader region of the image.

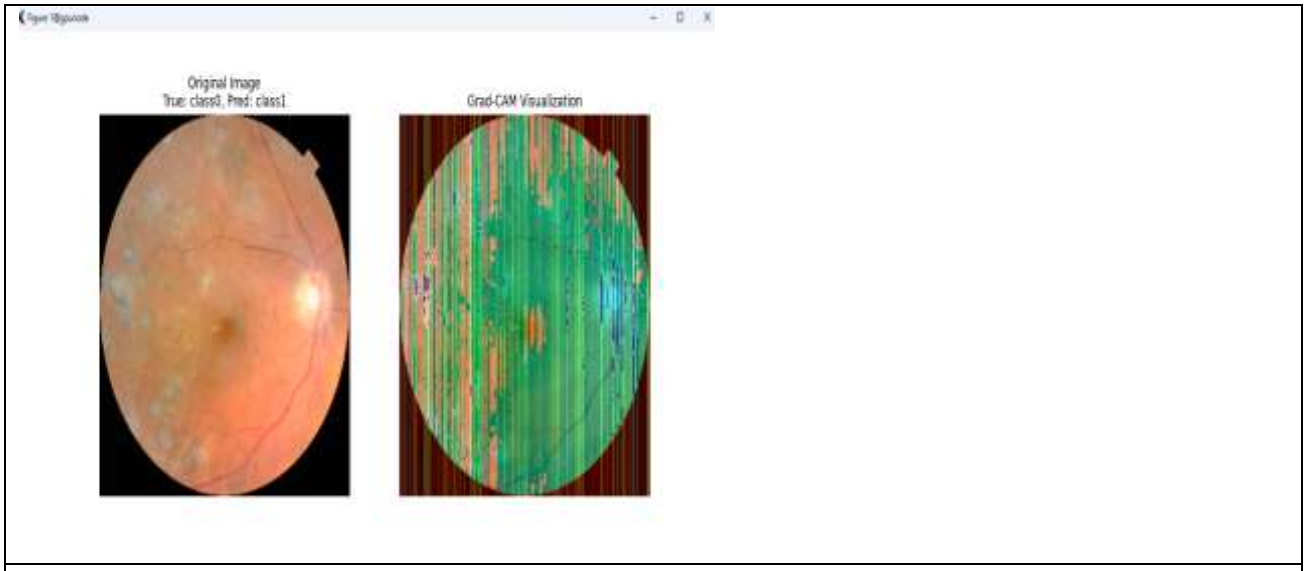

Fig.8 Grad-CAM Visualization 2

Description of Grad-CAM Visualization

Fig.8 consists of two subfigures: the original retinal fundus image (left) and the Grad-CAM visualization (right), which provides an explainable AI-based interpretation for decision-making process of the model.

1. Original Image (Left)
The left subfigure of Fig.8 represents a retinal fundus image, which serves as the input to the DL model for DR risk classification. The label at the top indicates:

- True Class: Class 0 (the actual ground-truth label for the image).

Predicted Class: Class 1 (This image was mistakenly assigned to Class 1 by the model).
This suggests a misclassification, where the model identified certain features in the image that led it to assign the wrong risk level. The image itself contains vascular structures, possible microaneurysms, or exudates, which contributes to the results

2. Grad-CAM Visualization (Right)

The Grad-CAM (Gradient-weighted Class Activation Mapping) output, which shows the areas of the retinal image that affected the model's prediction, is shown in the right subfigure of Fig.8.

- The green and orange overlay represents the attention map, where different intensity levels indicate the importance of various regions in decision-making.
- Ideally, disease-relevant areas such as lesions, microaneurysms, hemorrhages, or exudates should be highlighted in high-risk cases.
- However, the Grad-CAM visualization appears somewhat striped and less focused, suggesting that the model might be attending to irrelevant or inconsistent regions, potentially leading to the observed misclassification.

Fig.9 and Fig.10 shows some more images with their two subfigures i.e. original image and its Grad-CAM visualization.

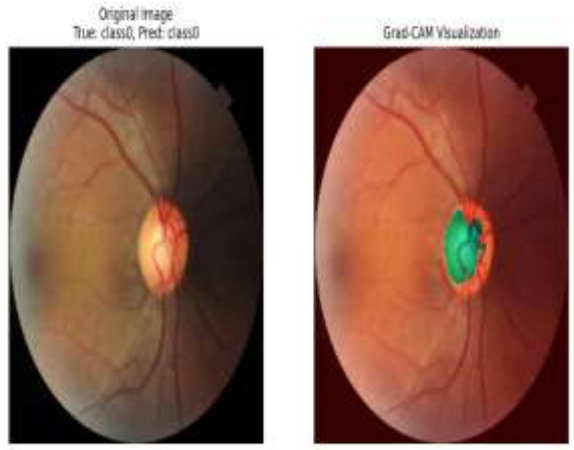

Fig.9 Grad-CAM Visualization 3

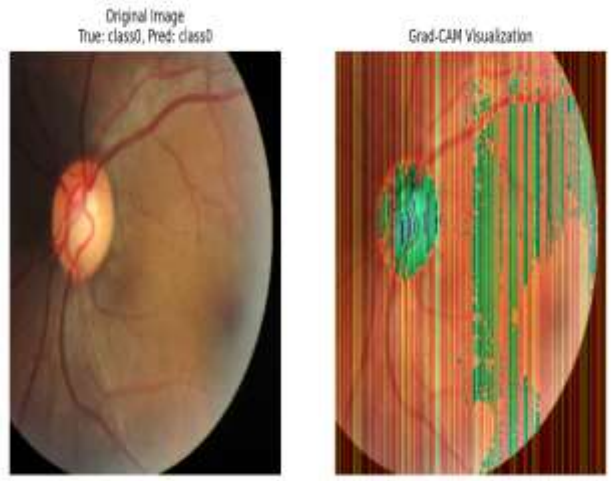

Fig.10 Grad-CAM Visualization 4

6. Comparison of the Proposed System with Other Datasets

To evaluate the generalizability and robustness of the proposed Vision Transformer-based model, conducted an extensive comparative analysis using an alternative dataset in addition to the primary dataset. The key evaluation metrics, including the classification report, confusion matrix, accuracy graph, sensitivity graph, specificity graph, and explainability results (XAI-based Grad-CAM visualizations), were computed for both datasets.

- **Confusion Matrix:**

The confusion matrix visually represents the number of correctly and incorrectly classified instances across all risk levels. A higher diagonal dominance (where correct predictions are concentrated along the diagonal) confirms strong classification performance. Below is the confusion matrix obtained for proposed system for Risk level prediction.

[[178 7 29 12]
[40 131 20 19]
[2 8 194 6]
[13 0 1 196]]

- **Accuracy, Loss, Sensitivity and Specificity Graph**

As can be seen in Fig.11, the accuracy curve of proposed system shows consistent improvement, with the model achieving a high final accuracy, validating its effectiveness in risk-level classification. Higher sensitivity across all risk levels meaning the model successfully detects most positive cases. High specificity means the model is not falsely predicting a high-risk case when it is actually low risk. For a proposed system it is showing high sensitivity, specificity as can be seen in Fig.12.

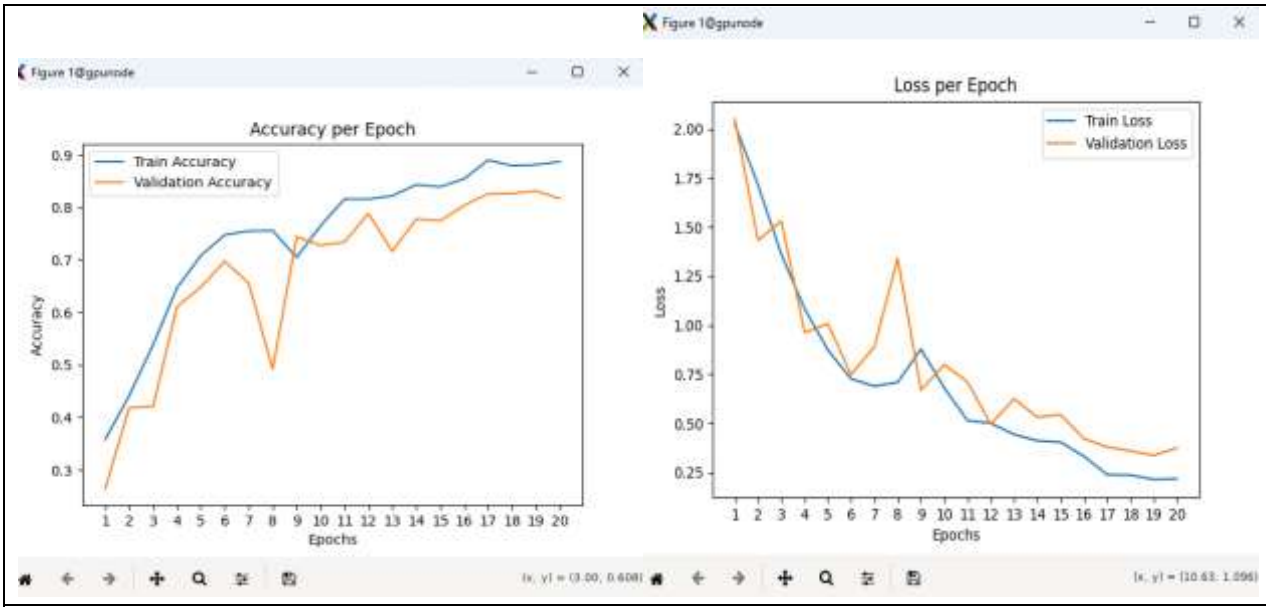

Fig.11 Accuracy and Loss Graph of Second Dataset

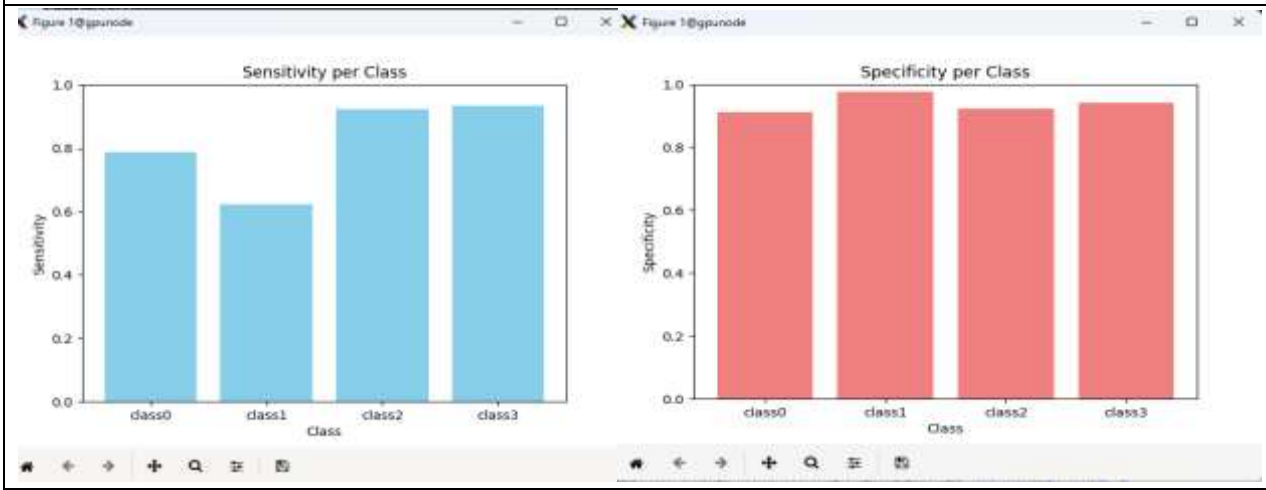

Fig.12 Sensitivity and Specificity Graph of Second Dataset

7. Conclusion

The findings prove that the suggested ViT inspired-InceptionV3 model, which has been adjusted by changes to

its architecture, has been efficient in classifying the risk levels of diabetic retinopathy. Its performance is confirmed through the classification report, the confusion matrix, the accuracy, the sensitivity and the specificity graphs, whereas the reliability of its use as a clinical tool is validated with the help of the XAI visualizations.

This paper extends the use of a ViT-based model to predict the risk level in diabetic retinopathy by using statistical features, pre-trained models, and XAI. The model obtains hierarchical representations of the images according to the predefined risk levels obtained on the basis of statistical characteristics and is additionally enhanced by optimized multi-head attention mechanisms, and a fine-tuned existing model to make the classification to be more accurate and robust.

Testing of other dataset demonstrated that the suggested method is consistent and reliable, and it is highly accurate, sensitive, and specific. Moreover, the addition of a modified version of Grad-CAM to ViT enhances interpretability, which makes the decision-making process of the model more transparent so as to increase its clinical uses. The generalizability and stability of the proposed system was confirmed by comparison with other datasets.

In general, Proposed System demonstrates the selection of the pretrained models, and an augmented ViT architecture to predict the risk of diabetic retinopathy. These findings indicate that this method can be an effective way of early diagnosis and risk evaluation, therefore enabling the implementation of clinical measures in time. Additionally, Combining diverse medical data may enhance the robustness of predictions and facilitate thorough disease evaluation.

Author Contribution ABD wrote the manuscript. ABD and PPG authors revised the work critically for important intellectual content and approved the final version to be published.

Funding No Funding

Conflict of interest- No

References

1. Ishtiaq, Uzair & Rahayu, Erma & Ishtiaque, Zubair. (2023). Ishtiaq, U.; Abdullah, E.R.M.F.; Ishtiaque, Z. A Hybrid Technique for Diabetic Retinopathy Detection Based on Ensemble-Optimized CNN and Texture Features. Diagnostics 2023, 13, 1816.. Diagnostics. 13. 10.3390/diagnostics13101816.
2. Parthiban, K. & Mari, Kamarasan. (2022). Diabetic retinopathy detection and grading of retinal fundus images using coyote optimization algorithm with deep learning. Multimedia Tools and Applications. 82. 10.1007/s11042-022-14234-8.
3. Kamble, Vaibhav & Kokate, Rajendra. (2020). Automated diabetic retinopathy detection using radial basis function. Procedia Computer Science. 167. 799-808. 10.1016/j.procs.2020.03.429.
4. Ryu, Gahyung & Lee, Kyungmin & Park, Donggeun & Park, Sang & Sagong, Min. (2021). A deep learning model for identifying diabetic retinopathy using optical coherence tomography angiography. Scientific Reports. 11. 10.1038/s41598-021-02479-6.
5. Kanadath, J. A. A. Jothi and S. Urolagin, "CViTS-Net: A CNN-ViT Network With Skip Connections for Histopathology Image Classification," in IEEE Access, vol. 12, pp. 117627-117649, 2024, doi: 10.1109/ACCESS.2024.3448302.
6. S. Mishra, S. Hanchate and Z. Saquib, "Diabetic Retinopathy Detection using Deep Learning," 2020 International Conference on Smart Technologies in Computing, Electrical and Electronics (ICSTCEE), Bengaluru, India, 2020, pp. 515-520, doi: 10.1109/ICSTCEE49637.2020.9277506
7. P. Bidwai, S. Gite, K. Patwa, K. Maheshwari, T. S. Bais and K. Batavia, "Detection of Diabetic Retinopathy using Deep Learning," 2023 IEEE 8th International Conference for Convergence in Technology (I2CT), Lonavla, India, 2023, pp. 1-8, doi: 10.1109/I2CT57861.2023.10126384.
8. Lalithadevi, B., Krishnaveni, S. Diabetic retinopathy detection and severity classification using optimized deep learning with explainable AI technique. Multimed Tools Appl 83, 89949–90013 (2024). https://doi.org/10.1007/s11042-024-18863-z

9. W. K. Wong, F. H. Juwono and C. Apriono, "Diabetic Retinopathy Detection and Grading: A Transfer Learning Approach Using Simultaneous Parameter Optimization and Feature-Weighted ECOC Ensemble," in IEEE Access, vol. 11, pp. 83004-83016, 2023, doi: 10.1109/ACCESS.2023.3301618.
10. Naz, H., Ahuja, N.J. & Nijhawan, R. Diabetic retinopathy detection using supervised and unsupervised deep learning: a review study. Artif Intell Rev 57, 131 (2024). https://doi.org/10.1007/s10462-024-10770-x
11. K. K. Y. Tiong, W. K. Wong, F. H. Juwono and I. M. Chew, "Diabetic Retinopathy Detection: Improving Accuracy Using Multiple Transfer Learning Features from Pre-trained Deep Learning Networks," 2022 International Conference on Green Energy, Computing and Sustainable Technology (GECOST), Miri Sarawak, Malaysia, 2022, pp. 171-175, doi: 10.1109/GECOST55694.2022.10010468.
12. Advances in Computer-Aided Diagnosis of Diabetic Retinopathy by Saket S. Chaturvedi, Kajol Gupta, Vaishali Ninawe, Prakash S. Prasad https://doi.org/10.48550/arXiv.1909.09853
13. Dual Branch Deep Learning Network for Detection and Stage Grading of Diabetic Retinopathy by Hossein Shakibania, Sina Raoufi, Behnam Pourafkham, Hassan Khotanlou, Muharram Mansoorizadeh https://doi.org/10.48550/arXiv.2308.09945
14. S. Suganyadevi, K. Renukadevi, K. Balasamy and P. Jeevitha, "Diabetic Retinopathy Detection Using Deep Learning Methods," 2022 First International Conference on Electrical, Electronics, Information and Communication Technologies (ICEEICT), Trichy, India, 2022, pp. 1-6, doi: 10.1109/ICEEICT53079.2022.9768544.
15. Detecting diabetic retinopathy severity through fundus images using an ensemble of classifiers by Eduard Popescu, Adrian Groza, Ioana Damian https://doi.org/10.48550/arXiv.2307.16622
16. Obayya M, Nemri N, Nour MK, Al Duhayyim M, Mohsen H, Rizwanullah M, Sarwar Zamani A, Motwakel A. Explainable Artificial Intelligence Enabled TeleOphthalmology for Diabetic Retinopathy Grading and Classification. Applied Sciences. 2022; 12(17):8749. https://doi.org/10.3390/app12178749
17. T. Shahzad, M. Saleem, M. S. Farooq, S. Abbas, M. A. Khan and K. Ouahada, "Developing a Transparent Diagnosis Model for Diabetic Retinopathy Using Explainable AI," in *IEEE Access*, vol. 12, pp. 149700-149709, 2024, doi: 10.1109/ACCESS.2024.3475550
18. K. Duvvuri, S. Chethana, S. S. Charan, V. Srihitha, T. K. Ramesh and S. K S, "Grad-CAM for Visualizing Diabetic Retinopathy," *2022 3rd International Conference for Emerging Technology (INCET)*, Belgaum, India, 2022, pp. 1-4, doi: 10.1109/INCET54531.2022.9824598.
19. Amna Ikram, Azhar Imran,ResViT FusionNet Model: An explainable AI-driven approach for automated grading of diabetic retinopathy in retinal images,Computers in Biology and Medicine,Volume 186,2025,109656,ISSN 0010-4825,https://doi.org/10.1016/j.compbiomed.2025.109656.
20. Mohammad Shorfuzzaman, M. Shamim Hossain, and Abdulmotaleb El Saddik. 2021. An Explainable Deep Learning Ensemble Model for Robust Diagnosis of Diabetic Retinopathy Grading. ACM Trans. Multimedia Comput. Commun. Appl. 17, 3s, Article 113 (October 2021), 24 pages. https://doi.org/10.1145/3469841
21. Vasireddi, H.K., Devi, K.S. & Reddy, G.N.V.R. DR-XAI: Explainable Deep Learning Model for Accurate Diabetic Retinopathy Severity Assessment. Arab J Sci Eng 49, 12899–12917 (2024).https://doi.org/10.1007/s13369-024-08836-7
22. J. Rohankar, S. Jagtap, C. Nimje, and P. Padole, "RetinAI-DR Fusion: Dual Deep Learning Framework for Diabetic Retinopathy Detection," International Journal on Advanced Computer Engineering and Communication Technology, vol. 14, no. 3s, pp. 213–219, 2025, doi: 10.65521/ijacect.v14i3s.1623.
23. S. Shirkande, S. B. Bhange, K. B. Sable, P. R. Salunkhe, and N. H. Shendage, "Exploring Deep Learning Architectures for Diabetic Retinopathy Detection," International Journal on Advanced Computer Engineering and Communication Technology, vol. 14, no. 1, pp. 694–697, 2025, doi: 10.65521/ijacect.v14i1.791.
24. P. Anu Rakhi, R. Vijayamahisha, V. Mathy P., and B. L. Priyadharshini, "A Robust Deep Learning-Based Classification Framework for Diabetic Eye Diseases Using Inception V3," International Journal on Advanced Computer Engineering and Communication Technology, vol. 15, no. 1, pp. 156–159, 2026, doi: 10.65521/ijacect.v15i1.2336.
25. G.F. Freire, "Federated Learning Models for Privacy-Preserving Healthcare Data Analysis", *Journal of Wireless Intelligence and Spectrum Engineering*, vol. 2, no. 1, pp. 19–25, Apr. 2025.
26. P.Dineshkumar. (2026). Intelligent Protection Schemes for Modern Power Systems Using Deep Learning-Based Fault Classification. *National Journal of Intelligent Power Systems and Technology*, 34-41.
27. Dinfe Egash, & Rane Kuma. (2025). Explainable Artificial Intelligence (XAI): State-of-the-Art, Challenges, and Research Trends. *Innovative Reviews in Engineering and Science*, *3*(2), 154-162.