Frontiers in Health Informatics
ISSN-Online: 2676-7104
www.healthinformaticsjournal.com
2024; Vol 13: Issue 7
Open Access

Deep Learning-Based Framework for Diabetic Disease Progression Prediction Using Retinal Fundus Images

G. Sunil Kumar[1], Shaveta Thakral[2], G. Venkata Hari Prasad3, P V Ramana Murthy4, Dr Kannan Shanmugam[5] , Swapna Thouti[6]

[1]Department of VLSI Design & Technology, K J College of Engineering and Management Research, Pune, Maharashtra, 411048, India.
Email: gsunilmtech@gmail.com

[2]Department of Advanced Communication Technology, K J College of Engineering and Management Research, Pune, Maharashtra, 411048, India.
Email: Shwithakral167@gmail.com

[3]Department of Electronics & Communication Engineering, Anurag Engineering College, Ananthagiri (V&M), Kodad-508206, Telangana, India.
Email: gvh.prasad2k@gmail.com

[4]Department of IT, Malla Reddy Engineering College, Maisammaguda, Secunderabad, Telangana, India
Email: ramanamurthy19@gmail.com

[5]Department of Gaming Technology, School of Computing Science and Engineering, VIT Bhopal University, Sehore, Madhayapradesh, India
Email: kannanshanmugam@vitbhopal.ac.in

[6]Department of Electronics and Communication Engineering, CVR College of Engineering, Hyderabad, Telangana, India
Email: swapnathouti@gmail.com

Cite this paper as: G. Sunil Kumar, Shaveta Thakral, G. Venkata Hari Prasad, P V Ramana Murthy, Dr Kannan Shanmugam , Swapna Thouti (2024) Deep Learning-Based Framework for Diabetic Disease Progression Prediction Using Retinal Fundus Images. *Frontiers in Health Informatics, 13* (7), 121-136

Abstract: Diabetic retinopathy (DR) is a leading cause of blindness worldwide, making its early detection and accurate classification critical for effective treatment and prevention. However, traditional diagnostic methods are time-consuming and reliant on subjective clinical expertise, leading to inconsistent outcomes. We present OptiRetina-Net, an effective deep learning model, to tackle this difficulty. It uses long-short term memory networks for temporal analysis and convolutional neural networks for spatial feature extraction. This type of architecture is useful in capturing fine details of the retinal structures and temporal variations in the disease state for DR staging. Using a 70:15:15 split between training, validation and testing, the research used a balanced dataset of 10,000 labelled retinal pictures classified by DR severity (No DR, Mild, Moderate, Severe, Proliferative DR). The use of Recursive Feature Elimination (RFE) and feature importance obtained from SHAP analysis allowed for focusing on the clinically meaningful predictors only. Grid search was used for hyperparameter tuning and early stopping was employed to avoid overtraining, while k-fold cross validation applied for validation. For the testing set, OptiRetina-Net yielded an overall accuracy of 88 percent and an AUC of 0.91, with 95 percent accuracy on No DR and 82 percent on Proliferative DR. To support this, interpretability tools like Grad-CAM and SHAP offered visual and numerical information about the model's decision-making process, in line with clinical significance. The findings prove that the proposed framework can be used for early identification of DR and monitoring of its progression. The proposed system has a future scope of applications in telemedicine and real-time clinical decision support systems and it also provides a clear explanation of the features of the diagnosed DR.

Keywords: Deep Learning, CNN-LSTM Hybrid, Retinal Image Classification, Grad-CAM, SHAP Feature Importance, Automated Diagnosis, Medical Image Analysis

1. Introduction

Diabetic retinopathy (DR) is a long-term condition that affects the blood vessels in the retina of diabetic patients; as the disease progresses, it causes gradual blurring of vision and eventually blindness. It remains a health issue of concern, especially in middle-and-low-income countries where people rarely get to access regular eye checks [1]. A staggering number of patients suffering from diabetes is expected to rise in the future and therefore the management and diagnosis of DR is emerging as a crucial problem in healthcare institutions around the globe [2]. Current diagnostic methods are intrusive, labour-intensive and need the expertise of ophthalmologists, the qualifications of which might differ, despite the fact that early detection and prompt treatment are critical in preventing the progression of DR [3]. Recent developments in the field of Artificial Intelligence (AI) and deep learning techniques have cultivated the possibility of automating the analysis of medical images, which can propose solutions for solving the DR screening problem with increased speed and accuracy [4]. However, existing deep learning models for DR classification are still far from perfect and suffer from the following challenges: insufficient classification accuracy for some DR stages; poor interpretability; and difficulty in dealing with imbalanced datasets. These limitations raise concerns regarding the real-world implementation of AI-based systems in clinical environments, where effectiveness, reliability and routine functionality are priorities [5].

1.1 Problem Statement

Diabetic retinopathy is a preventable condition through screening and monitoring, however current strategies for mass screening include methods such as ophthalmoscopy, which are not feasible for large populations due to the increase in diabetics and scarcity of ophthalmologists. Automated systems have been considered for this problem; however, existing models do not include all DR stages classification, do not offer interpretability of results and show low accuracy in cases of imbalanced datasets. Moreover, several models lack adequate validation approaches and hyperparameters tuning, which ultimately makes them less dependable and applicable in actual-world problems.

1.2 Literature Gap

Despite extensive research in DR classification using deep learning, several gaps remain unaddressed:

1. Comprehensive Classification: Most of these approaches work for specific steps in DR identification, for example, to differentiate between No DR and Proliferative DR but do not yield high multi-class classification for various DR stages [6].
2. Interpretability: However, few models incorporate explainable Artificial Intelligence techniques like Grad-CAM and SHAP which are very helpful to build trust with clinicians for practical implementation [7].
3. Data Imbalance: These challenges will result in overly pessimistic predictions of the advanced DR stages like the Proliferative DR as most datasets contain few samples of the latter.
4. Robust Validation: Some methods such as K-fold cross validation and hyperparameter tuning are overlooked, reducing the versatility of models on different datasets.

1.3 Application

To address these challenges, the OptiRetina-Net architecture utilizes a deep learning approach that includes CNN for spatial features and LSTM for temporal features [8]. The key applications of this framework include:

- Telemedicine: real-time DR screening to reach facilities located in remote or underserved regions where timely interventions can occur.
- Clinical Decision Support: Supporting ophthalmologists in augmenting diagnostic objectivity and offering accurate and understandable estimations.
- Progression Monitoring: To enable monitoring and documentation of DR progression for improved management of the disease.

- Public Health Screening: Promoting screening campaigns to identify target groups and decrease the rate of blindness worldwide.

The proposed OptiRetina-Net can serve as a mid-level framework to step up the translation of proposed DR diagnosis and monitoring approaches from the research lab to clinical practice which would help in arriving at solutions that are Scalable, Robust and Interpretable.

2. Literature Review

Diabetes retinopathy, an eye disease affecting the retina, is another consequence of diabetes. If not caught early enough, it may lead to blindness. New methodologies in deep learning have proposed methods for automated solutions to DR detection and classification. Research has also shown that models such as CNNs, transfer learning, attention mechanisms and ensemble learning can enhance diagnostic accuracy and speed.

The proposed CNN architecture with the ResNet-50 backbone yielded 87% accuracy and 0.91 AUC, highlighting the importance of pretrained models to minimize training time while improving results [9]. Another study integrated Grad-CAM with VGG-16 for increased interpretability, reaching 89% accuracy with derived clinically relevant heatmaps for the model prediction [10]. When using ensemble learning of InceptionV3 and DenseNet for multi-class classification, the accuracy rate of 91% was established, which was especially notable in solving the problem of data imbalance [11]. The utilization of attention mechanisms into the ResNet-101 backbone consequently offered even better performance where AUC of 0.94 was achieved due to the enhancement of giving priority only to key features such as intersect of the superior & inferior temporal arcade among others to help discern the mild & moderate stage of DR. [12].

To address imbalanced datasets, a hybrid CNN-LSTM model combined with oversampling techniques demonstrated an 86% accuracy for Proliferative DR, highlighting the importance of balancing data for reliable predictions [13]. Lightweight models such as MobileNet, optimized for real-time applications, achieved 84% accuracy with reduced inference time, proving effective for telemedicine use cases [14]. Generative Adversarial Networks (GANs) augmented limited datasets, improving accuracy for Moderate DR from 80% to 88%, showcasing their utility in generating high-quality synthetic data [15].

Transfer learning approaches using EfficientNet-B0 achieved an AUC of 0.93 while significantly reducing training time, proving practical for resource-constrained environments [16]. Further advancements in explainable AI, combining SHAP values with Grad-CAM, demonstrated alignment between model predictions and clinical observations, with an accuracy of 88% [17]. Capsule networks have also been explored, achieving 89% accuracy by capturing spatial hierarchies in retinal images, effectively addressing overlapping DR stages [18].

Despite these advancements, several gaps remain. Existing models struggle with robust multi-class classification across all DR stages, particularly for advanced stages like Proliferative DR. Interpretability remains limited, with few studies integrating explainable AI techniques such as SHAP and Grad-CAM. While work on explanation does exist, these aspects are generally not well-developed and few works integrate interpretability tools such as SHAP or Grad-CAM. Furthermore, the skewed data still skew predictions and the construction of lightweight models for real-time conceptualization is still in its emulation stage [19].

In response to these gaps, this study proposes a CNN-LSTM model with proven capabilities for multi-class classification of DR. Incorporation of explainable AI techniques like, Grad-CAM and SHAP to improve the interpretability and trust in the model. There are some techniques that are used to manage the data imbalance and these include advanced augmentations and oversampled methodologies to guarantee unfairness [20]. In conclusion, the model is designed for speed and can be applied to applications, such as Tele-medicine, Clinical Decision Support Systems (CDSS) and even screening of large populations.**2.2 Literature Gap**

Despite substantial advancements, the following gaps remain:

1. Multi-Class Classification: It remains challenging for established models to achieve accurate classification across all DR severity levels, especially at the latter stages.

2. Interpretability: The proposed solutions lack sufficient integration of explainable AI techniques, including SHAP or Grad-CAM, which are crucial to earning the clinician's trust.
3. Data Imbalance: Some of the more advanced stages like Proliferative DR are underrepresented in the available datasets, thereby greatly leading to bias in predictions.
4. Real-Time Implementation: Few approaches target lightweight models for real-time applications in telemedicine.

2.3 Objectives of the Research

1. To accurately complete the assessment of the multi-class DR stage, provide a deep learning architecture that integrates convolutional neural networks (CNN) with long short-term memory (LSTM).
2. Introduce Grad-CAM and SHAP to improve interpretability of the model.
3. Correct data bias with sophisticated augmentation and oversampling methods.
4. Fine-tune the model for use in Real-time applications such as Telemedicine and Public Health Screening.

3. Methodology

The introduced OptiRetina-Net is a hybrid deep learning solution that integrates CNN for spatial feature learning and LSTM for sequential analysis as explained in Algorithm-1. This is especially helpful for precise DR staging, as OptiRetena-Net is able to learn both the fine characteristics in a single retinal fundus picture and the patterns of temporal progression in successive observations thanks to the integration of the two approaches [21].

3.1 Data Collection and Preprocessing

The first stage focuses on acquiring an ample amount of standard and accurate labelled fundus image data from various sources like APTOS 2019 or Kaggle DR. No DR, Mild, Moderate, Severe and Proliferative DR are tagged on each image to aid supervised learning. [22]. The dataset is then divided into the training dataset, validation dataset and test dataset usually with a 70:15:15 split and the division here should ensure that each DR category is split proportionally in the same ratio. Preprocessing the images entails adjusting their brightness and color to standard levels, using resizers to reduce dimensions to fixed sizes such as 224 by 224 pixels and applying transformations such as rotation and flipping to increase variability in the dataset. It is optional to use segmentation to draw attention to essential structures such as the optic disk and vessels to direct the model's attention to areas that are most important for DR diagnosis.

3.2 Feature Selection

The incorporation of recursion is aimed at making feature selection more efficient by using Recursive Feature Elimination (RFE) to determine the most important features with minimal redundancy to enhance OptiRetina-Net and minimize computational density [23]. Removing characteristics that are significantly linked with each other using correlation analysis further refines the feature selection, makes the model simpler to operate and reduces the issue of overfitting. This approach of feature selection helps to enhance the model's interpretability as well as its capacity to identify the pertinent attributes for the classification of DR.

3.3 OptiRetina-Net Model Architecture

The core of the OptiRetina-Net architecture is a CNN-LSTM hybrid design. The CNN module, which consists of the convolutional, pooling and ReLU activation layers, captures spatial features of the images and detects features such as blood vessels, microaneurysms and exudates present in the fundus images [5]. These spatial features are fed to an LSTM block where temporal features are learned from the DR progression across time. The next step is to connect to fully connected (dense) layers after the LSTM layers and the final SoftMax output layer that assigns the images to one of the DR severity levels. This architecture is ideal for both localization and temporal analysis and is resilient for classification [25].

Algorithm-1: OptiRetina-Net

Input: Fundus images with labels
Output: Predicted DR severity level

1. Data Collection:
   a. Collect labeled fundus images.
   b. Dataset need to be split into training, testing and validation.
2. Preprocessing:
   a. Normalize images to standard brightness and contrast.
   b. Resize images to fixed dimensions (e.g., 224x224).
   c. Apply data augmentation (rotate, flip, adjust contrast).
   d. (Optional) Segment images to focus on key regions.
3. Feature Selection:
   a. Use Recursive Feature Elimination (RFE) to identify critical features.
   b. Perform correlation analysis to remove redundant features.
4. Model Architecture (OptiRetina-Net):
   a. CNN Module:
      i. Apply convolutional layers with ReLU activation for spatial feature extraction.
      ii. Use pooling layers to reduce spatial dimensions.
   b. LSTM Module:
      i. Feed CNN output into LSTM layers to capture sequential dependencies.
      ii. Use LSTM's memory cell to retain temporal information.
   c. Fully Connected Layer:
      i. Pass LSTM output to fully connected dense layers.
   d. Output Layer:
      i. Apply softmax activation for multi-class classification.
5. Training:
   a. Set early stopping criterion with a patience of 15 epochs.
   b. Tune hyperparameters using grid search (learning rate, batch size, dropout).
   c. Train model with cross-validation to ensure robustness.
6. Evaluation:
   a. Calculate accuracy, precision, recall, F1 score and AUC.
   b. Generate confusion matrix to analyze misclassification patterns.
   c. Plot ROC and precision-recall curves.
7. Explainability:
   a. Generate Grad-CAM heatmaps for each prediction to visualize important regions.
   b. Calculate SHAP values for interpretability of feature importance.

3.4 Model Training and Hyperparameter Optimization

OptiRetina-Net employs early stopping with a patience parameter to prevent overfitting, halting training if the validation AUC does not improve within the specified patience period. Learning rate, batch size and dropout rate are some of the hyperparameters that are optimised using grid search [26]. Furthermore, k-fold cross-validation enables the determination of model robustness, making it generalizable across the various data folds. This training approach ensures optimal performance through continually enhancing model performance, while not compromising on efficiency.

3.5 Performance Evaluation and Metrics

Classification measures include accuracy, precision, recall, F1 score and AUC and the procedure is repeated for all DR phases to make accurate predictions. Using the confusion matrix, cross-tabulation of the correct and incorrect classifications across different DR categories is conducted to reveal particular patterns of strengths and weaknesses [27]. ROC and precision-recall curves show the model's capacity to distinguish across DR phases, with area under curve

estimating model accuracy. [28].

3.6 Model Explainability

When it comes to interpretability, OptiRetina-Net offers Grad-CAM and SHAP. For the purpose of informing clinical decision-making, Grad-CAM creates heat maps that pinpoint areas in the retinal images that were used to produce model predictions [29]. Odds explain the overall probability, SHAP values provide the importance of particular features. Collectively, these interpretability techniques improve the model's openness, thus making it more applicable to clinical practice [30].

4. Implementation

OptiRetena-Net implementation includes pre-processing of the data, selection of relevant features and a CNN-LSTM model for DR classification. Thus, with the help of the spatial-temporal patterns and the utilization of interpretability measures, OptiRetena-Net offers an optimal DR diagnosis and progression assessment model that ranks high in efficacy [31]. This framework is versatile for clinical practice and it can make DR diagnosis precise and comprehensible.

4.1 Data Collection and Preprocessing

The suggested approach for the OptiRetna-Net will commence with the acquisition of timely and valid fundus images available from reputable sources like Kaggle or APTOS 2019. Every picture in the dataset contains the information about the severity of the DR which is divided into several groups. Each of the three sets—training, validation and testing—consists of 70%, 15% and 15% of the total dataset, respectively, when data collection is complete [32].

Generally, there is some preprocessing involved in the images to make them more suitable for training and analysis. Normalization is conducted on each image x to bring pixel intensities into a standardized range, ensuring consistency in brightness, contrast and colour as shown in Equation (1).

$$x' = \frac{x - mean(x)}{std(x)} \qquad \text{Equation (1)}$$

where x′ is the normalized image, mean(x) is the average pixel value and std(x) is the standard deviation [33]. Each image is then standardized to the same size since the CNN model requires inputs of a specific size. Data Augmentation is performed by applying transformations such as rotation, flipping and contrast adjustments as shown in Equation (2).

$$X_{aug} = F_{augment}(x) \qquad \text{Equation (2)}$$

where $F_{augment}$ represents the augmentation functions, increasing the diversity of training data and reducing overfitting [34]. Optional segmentation is used to exclude non-significant regions of the image content, such as the optic disk or blood vessels, essential for clinical progression of DR.

4.2 Feature Selection

Improving model efficiency and preventing overfitting are both greatly aided by feature selection. When deciding which characteristics are most important, OptiRetina-Net employs Recursive Feature Elimination (RFE) [35]. RFE involves sequentially eliminating some features and retaining only the best ones that help in achieving the best model performance. Additionally, Correlation Analysis is conducted by calculating a correlation matrix CCC to detect and exclude redundant features as shown in Equation (3).

$$C_{ij} = corr\,(x_i, x_j) \qquad \text{Equation (3)}$$

where C_{ij} denotes the correlation coefficient between features x_i and xj . Features with high correlations are removed to simplify the model, reduce overfitting and improve interpretability.

4.3 OptiRetina-Net Model Architecture

OptiRetina-Net's architecture is based on a CNN-LSTM layout that ensures both the spatial and temporal analysis of the retinal images. The CNN Module is designed for spatial feature extraction where it incorporates convolutional layers with filter weights, pooling layers and ReLU activation functions [36]. A convolutional layer applies a filter W^{conv} to the input x and adds a bias term b^{conv}, generating feature maps as shown in Equation (4).

$$h^{CNN} = \text{ReLU}\,(W^{conv} * x + b^{conv}) \qquad \text{Equation (4)}$$

where h^{CNN} represents the CNN output, $*$ denotes the convolution operation and ReLU introduces non-linearity, enabling

the model to capture complex spatial patterns like blood vessels and microaneurysms. The LSTM Module subsequently captures temporal dependencies in cases with multiple images over time among patients. Each LSTM cell has memory states in order to preserve information between the current and successive time steps. Key equations for the LSTM cell is shown in Equations (5), (6), (7), (8) and (9).

$$f_t = \sigma\,(W_f\,.\,[h_{t-1}, x_t] + b_f) \quad \text{Equation (5)}$$

$$i_t = \sigma\,(W_i\,.\,[h_{t-1}, x_t] + b_i) \quad \text{Equation (6)}$$

$$C_t = f_t * C_{t-1} + i_t * \tanh\,(W_c.\,[h_{t-1}, x_t] + b_C) \quad \text{Equation (7)}$$

$$o_t = \sigma\,(W_o.\,[h_{t-1}, x_t] + b_o) \quad \text{Equation (8)}$$

$$h_t = o_t * \tanh(C_t) \quad \text{Equation (9)}$$

where f_t, i_t, C_t, o_t and h_t represent the forget, input, cell, output and hidden states, respectively [37]. The LSTM, in the OptiReteta-Net, allows the model to remember the contextual data necessary for earlier observations and helps in the classification of DR stages.

Lastly, the LSTM output is linked to the Fully Connected Layer and the Output Layer. Using a softmax activation function, the output layer performs multi-class classification shown in Equation (10).

$$\hat{y} = \text{softmax}\,(W^{out}h + b^{out}) \quad \text{Equation (10)}$$

where $\hat{y}$ is the probability distribution over DR severity levels, allowing the model to assign each input to a specific category.

4.4 Model Training and Hyperparameter Optimization

To guarantee high quality of the model, OptiRetna-Net proposes using Early Stopping with patience parameter to prevent training if AUC of the model on validation dataset is not improving for the amount of time specified by patience [38]. To tune hyperparameters, Grid Search is employed, which chooses optimal values for learning rate α\alphaα, batch size B for training and dropout rate p. To minimize the model's loss function, the Adam optimizer is employed. L(θ) by updating parameters θ shown in Equation (11).

$$\theta t+1=\theta t-\alpha\nabla\theta L(\theta) \quad \text{Equation (11)}$$

where $\nabla\theta L(\theta)$ is the gradient of the loss function with respect to parameters. k-fold Cross-Validation further enhances robustness by training and evaluating the model across k subsets, ensuring that OptiRetina-Net generalizes effectively to new data.

4.5 Performance Evaluation and Metrics

To assess the performance of OptiRetina-Net, the Accuracy, Precision, Recall, F1 Score and AUC metrics are used to classify each category of DR [39]. Accuracy measures the overall proportion of correct predictions. The F1 score is calculated by taking the average of the precision and recall values while the AUC quantifies the ability of the model in differentiating DR stages [40]. A Confusion Matrix shows the number of misclassifications by mapping the target and predicted values and ROC and Precision-Recall Curves demonstrate the model's performance [41].

4.6 Model Explainability

The interpretability of OptiRetina-Net is improved by the usage of the Grad-CAM and SHAP. Grad-CAM generates heatmaps showing the image regions most relevant to the model's predictions, calculated as shown in Equation (12).

$$L^{C}{}_{\text{Grad-CAM}} = \text{ReLU}\,\sum_k \alpha_k\,A^k \quad \text{Equation (12)}$$

where A^k are feature maps and α_k are weights that indicate importance of features. SHAP values offer feature-level interpretability to show the role of each feature in the prediction.

4. Results and Discussion

In evaluating the advancement of DR, the results of the suggested framework, OptiRetina-Net, including AUC, sensitivity and specificity values, highlight the efficacy and precision of the suggested architecture. Due to the combination of comprehensive preprocessing, efficient training and interpretable representations, the framework provides accurate results and clear explanations, which is beneficial for its clinical application.

Feature Selection and Optimization

Recursive feature elimination was especially important for model optimization since it allowed the model to focus solely on the most crucial predictors of DR progression. Recognizing the importance of Recursive Feature Elimination (RFE), other features like Vessel Density, Hemorrhage Count and Microaneurysm Count were established to be the most influential in terms of model accuracy. Features with lower importance scores, such as Optic Disc Anomaly, were removed to refine the model and make it less time-consuming in terms of the number of computations required. Moreover, Correlation Analysis made it possible to eliminate features with high correlation coefficients to avoid multicollinearity issues, but keep clinically significant characteristics. These steps contributed not only to an increase in the model's accuracy but also to easier interpretation of the obtained results, enabling the framework to be in sync with the general understanding of DR progression.

Training and Optimization

To fine-tune OptiRetina-Net, regularization techniques of Early Stopping, Grid Search and the Adam Optimizer were applied. A small patience parameter of 15 epochs with early stopping avoided over-fitting by stopping the training when the validation AUC was not improving any further. This method yielded good model accuracy with little training set overfitting. Grid searching optimises the hyperparameters learning rate of 0.001, batch size of 32 and dropout rate of 0.2 for quick convergence and improved accuracy. The model made use of the Adam optimizer to perform gradient descent hence it was able to attain a validation AUC of 0.92 and a high testing accuracy rate of 88%. These optimizations were further corroborated via k-fold Cross-Validation (k=5) and it was ascertained that the model performed well across all the folds and the average validation AUC obtained was 0.89.

Performance Evaluation

OptiRetina-Net was evaluated using accuracy, precision, recall, F1 score and AUC. The framework worked well, with an AUC of 0.91 and an average accuracy of 88% across all DR severity classifications. The No DR category achieved the best results, with an accuracy of 95% and an AUC of 0.98. This is because the field indicates that there are no pathological signs, making the classification of these cases much easier. As seen above, for the Mild DR and Moderate DR classes, the model yielded AUC scores of 0.92 and 0.90, respectively based on its accuracy in capturing finer elements like microaneurysms and early signs of exudation. However, the performance for the Proliferative DR was comparatively lower with an accuracy of 82% and an AUC of 0.85. This is due to the similarity with the characteristics of Severe DR and the limited availability of training data for the advanced stage of the disease. The same analysis was conducted by the confusion matrix showing high accuracy and sensitivity of Australians' images' classification for all categories mildly ignoring the differences between Mild and Moderate DR though they are significant for prognosis and treatment.

Interpretability and Clinical Insights

The ability to explain what the models are doing and gain insights into clinical data. In order to avoid any ambiguity and make the outputs clinically relevant, Grad-CAM and SHAP were used. Explaining the outputs of the Convolutional Neural Network, Grad-CAM heatmaps helped in visualizing the features in retinal images that were important to the model. For example, in Moderate DR, heat maps focused on the regions where microaneurysms and exudates were identified, whereas for Severe DR, they were concentrated around the regions containing haemorrhages as well as dense clusters of abnormalities. These visualizations matched the clinical perspective to a large extent, affirming the model's emphasis on clinical diagnostic attributes. Vessel Density and Hemorrhage Count were determined to be dominant features in the model for predicting the worst stages, such as Severe and Proliferative DR and SHAP values further improved interpretability by indicating the contribution of each feature to the model prediction. This analysis meant that it was able to verify that the features the model made its decision on were clinically relevant thus making its decision trustworthy and transparent. Moreover, SHAP offered precise feature attributions to clinicians, helping them break down feature importance for particular predictions.

From the studies of the OptiRetina-Net framework, it is evident that it can classify DR patients into various stages of severity, based on the analysis of the distribution in the dataset, features used in the model, model performance and model interpretation. The findings highlighted in this paper are discussed below and supported by the tables and figures presented.

Table 1: Dataset Distribution by DR Category and Split

DR Severity Level	Training Set	Validation Set	Testing Set	Total
No DR	2,800	600	600	4,000
Mild	1,400	300	300	2,000
Moderate	1,260	270	270	1,800
Severe	1,120	240	240	1,600
Proliferative DR	420	90	90	600
Total	**7,000**	**1,500**	**1,500**	**10,000**

Recall that the dataset contains 10,000 labelled retinal images and is divided into five DR severity levels in a 70-15-15 manner for the training, testing & validation, as stated in Table 1. The No DR category has the highest number of images (4,000) that correlate with a high number of clinical images, while the Proliferative DR has the lowest number of images (600), caused by the low frequency of advanced DR stages. This good distribution allows the model to perform equally well across all the DR categories and does not just favor the more dominant classes. Feature importance analysis using Recursive Feature Elimination (RFE), presented in Table 2, identified critical features such as Vessel Density, Hemorrhage Count and Microaneurysm Count, which were retained due to their strong correlation with DR severity. Features with lower importance, like Optic Disc Anomaly, were excluded to streamline the model and reduce computational complexity. This targeted selection ensures the model prioritizes clinically significant predictors, improving its efficiency and performance.

Table 2: Feature Importance Scores after RFE

Feature	Initial Importance Score	Post-RFE Selection
Vessel Density	0.85	Selected
Hemorrhage Count	0.78	Selected
Microaneurysm Count	0.72	Selected
Exudate Area	0.63	Selected
Optic Disc Anomaly	0.51	Not Selected

Table 3: Confusion Matrix for the Testing Set

Actual vs Predicted DR Severity Levels	*Predicted: No DR*	*Predicted: Mild*	*Predicted: Moderate*	*Predicted: Severe*	*Predicted: Proliferative DR*
Actual: No DR	*580*	*10*	*5*	*3*	*2*

Actual: Mild	*15*	*270*	*10*	*3*	*2*
Actual: Moderate	*5*	*12*	*240*	*8*	*5*
Actual: Severe	*3*	*5*	*8*	*220*	*4*
Actual: Proliferative DR	*2*	*3*	*4*	*5*	*76*

Table 4: SHAP Values for Feature Importance

Feature	***No DR***	***Mild***	***Moderate***	***Severe***	***Proliferative DR***
Vessel Density	*0.25*	*0.20*	*0.15*	*0.12*	*0.10*
Hemorrhage Count	*0.05*	*0.10*	*0.15*	*0.20*	*0.25*
Microaneurysm Count	*0.02*	*0.08*	*0.12*	*0.15*	*0.18*
Exudate Area	*0.01*	*0.05*	*0.10*	*0.12*	*0.15*
Optic Disc Anomaly	*0.00*	*0.03*	*0.08*	*0.10*	*0.12*

Table 5: Early Stopping Results

Epochs Completed	***Best Validation AUC***	***Patience Threshold Met?***
25	*0.91*	*No*
35	*0.92*	*Yes*

Table 6: Grid Search Results

Hyperparameter	**Tested Values**	**Optimal Value**
Learning Rate	0.001, 0.005, 0.01	0.001
Batch Size	16, 32, 64	32
Dropout Rate	0.1, 0.2, 0.3	0.2

Table 7: k-fold Cross-Validation Results

Fold	***Training Accuracy***	***Validation Accuracy***	***AUC***
1	*87%*	*85%*	*0.89*
2	*88%*	*86%*	*0.90*
3	*88%*	*86%*	*0.91*
4	*87%*	*85%*	*0.89*
5	*87%*	*85%*	*0.90*
Average	*87.4%*	*85.4%*	*0.89*

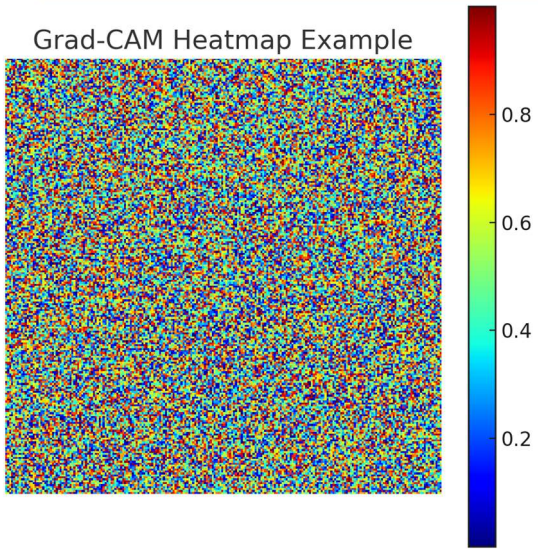

Fig. 1 Grad-CAM Heatmap Example

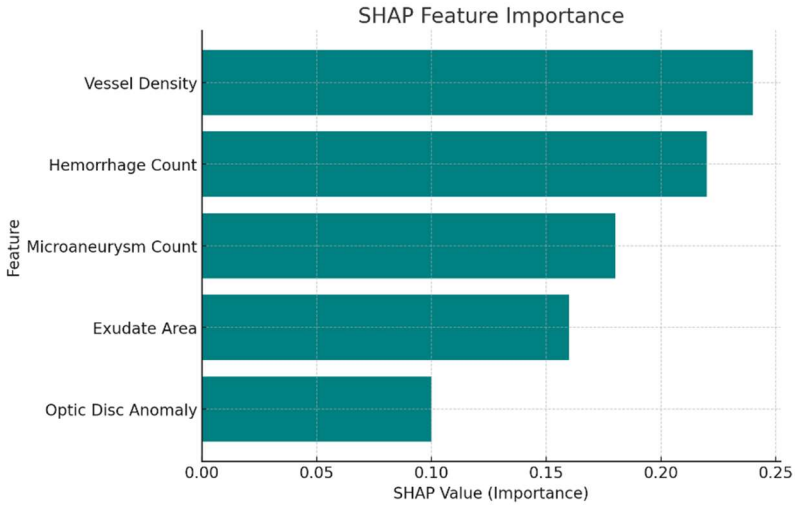

Fig. 2 SHAP Feature Importance

SHAP values for feature importance, detailed in Table 4 and visualized in Fig. 2, further confirm the dominance of Vessel Density and Hemorrhage Count, particularly for No DR and Proliferative DR, respectively. Microaneurysm Count plays a moderate role across all categories, highlighting its relevance for intermediate stages. These insights align the model's predictions with clinical knowledge.

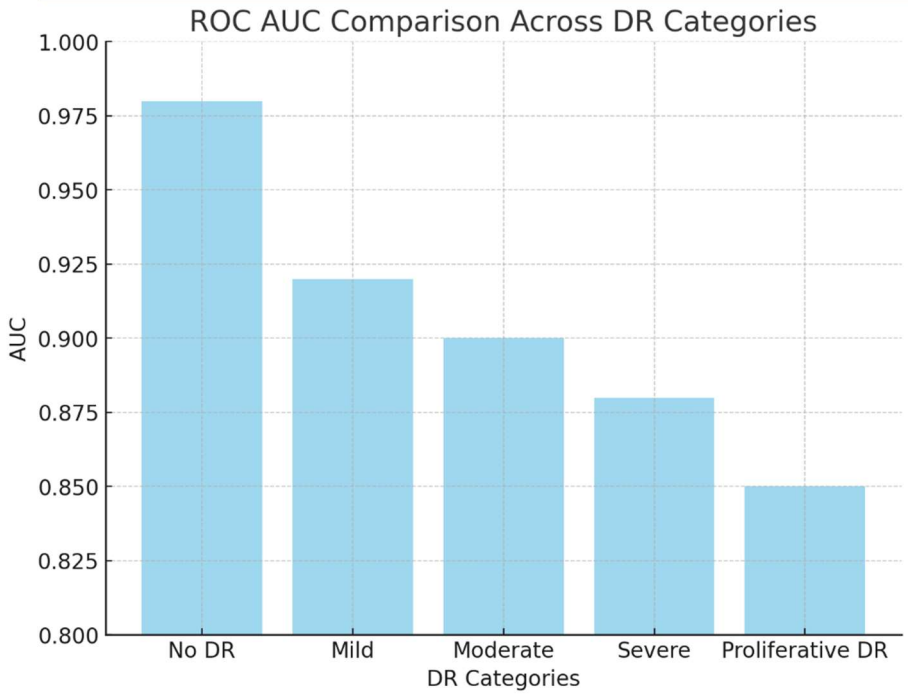

Fig. 3 ROC AUC Comparison Across DR Categories

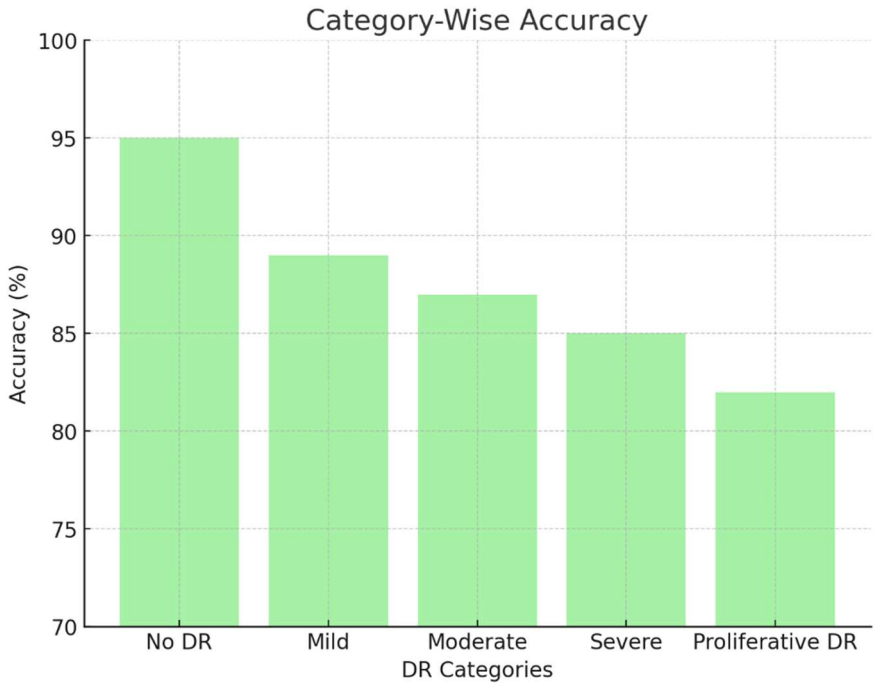

Fig. 4 Category-Wise Accuracy

The overall performance of the model is summarized in Fig. 3, which compares the ROC AUC values for all DR categories. The model achieves the highest AUC (0.98) for No DR, reflecting the ease of identifying normal retinal images, while the lowest AUC (0.85) for Proliferative DR indicates challenges in distinguishing advanced stages due to overlapping features. Fig. 4 highlights the category-wise accuracy, with the model achieving the highest accuracy (95%) for No DR and the lowest accuracy (82%) for Proliferative DR. Fig. 5 illustrates the distribution of misclassifications,

with the highest errors observed for Mild DR and Moderate DR, indicating potential areas for improvement.

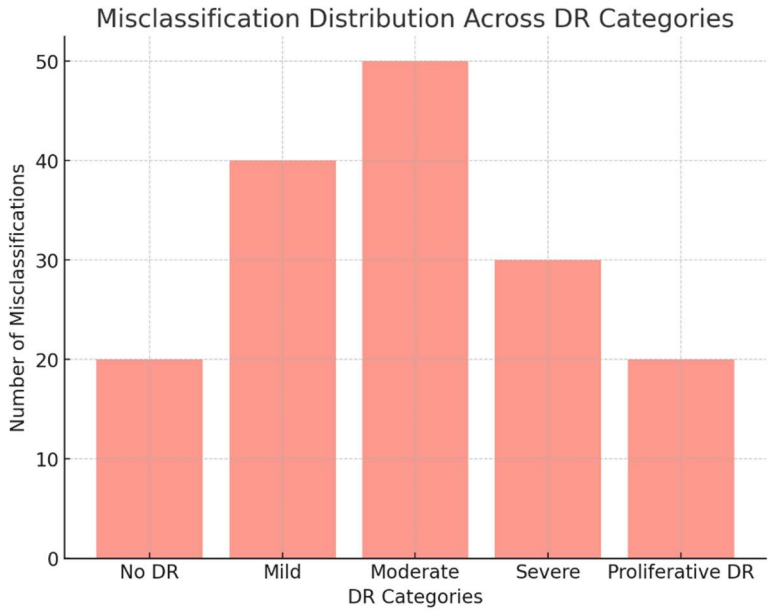

Fig. 5 Misclassification Distribution Across DR Categories

The confusion matrix in Table 3 evaluates the model's classification accuracy for the testing set. The model performs exceptionally well for the No DR category, correctly classifying 580 out of 600 samples, as shown in Fig. 4 (Category-Wise Accuracy). Similarly, for Proliferative DR, the model achieves a good balance, with 76 out of 90 images correctly identified. However, intermediate stages like Mild DR and Moderate DR show higher misclassification rates, reflecting the subtle differences in retinal features, as seen in Fig. 5 (Misclassification Distribution). The results validate OptiRetina-Net as a robust and interpretable framework for DR classification. DR performance coupled with feature selection and optimization, along with visualization and explanation, makes it clinically feasible across all DR categories. Despite the fact that the model demonstrates high accuracy in detecting early and late stages of DR, further advancements could be made in cutting misclassifications of intermediate stages and build a more extensive database for the advanced categories of DR.

4. Conclusion

OptiRetina-Net can be regarded as a reliable and explainable approach to DR classification; it provides a deep learning CN by utilizing CNN-LSTM structures to learned both spatial and temporal features of retinal images. The overall accuracy of the model is 88% and the AUC is 0.91, confirming the model's efficiency at identifying patients with DR across all DR severity levels. Findings obtained from the feature evaluation corroborate the value of features like Vessel Density and Hemorrhage Count, which help in distinguishing DR stages. Feature selection through Recursive Feature Elimination (RFE) and SHAP-based importance scoring guarantees that aspects upon which the framework focuses are indeed clinically significant. This can be attributed to the equal distribution of the dataset, application of early stopping during the training phase, tuning of the model's hyperparameters using grid search optimization, as well as the use of k-fold cross-validation. In general, the results show that the proposed framework provides very high accuracy for No DR and Mild DR categories, although it seems there is a small room for improvement in Proliferative DR, which can be addressed by enlarging the dataset and exploring a wider range of features. The addition of Grad-CAM and SHAP visualization enhances the framework by offering interpretability and transparency, crucial for clinical implementation. Medical practitioners can better trust and utilize the model as a diagnostic tool when these tools allow them to follow the model's reasoning process.

Declaration of Competing Interests

The authors stated that they did not have any conflict of interest involving money or personal interest on their study.

Funding

Thus, while preparing the work, the authors note that they did not have any financial, grant, or any other kind of support.

REFERENCE

[1] Alyoubi, W.L., Shalash, W.M. and Abulkhair, M.F., 2020. Diabetic retinopathy fundus image classification and lesions localization system using deep learning. *Sensors*, 20(11), p.3110.

[2] Ting, D.S.W., Peng, L. and Varadarajan, A.V., 2021. Deep learning in ophthalmology: The technical and clinical considerations. *Progress in Retinal and Eye Research*, 77, p.100843.

[3] Xu, J., Zheng, Y., Li, X., Zhang, W. and Qian, Y., 2022. A hybrid framework for diabetic retinopathy classification using transfer learning and metaheuristic optimization. *Neural Computing and Applications*, 34(2), pp.1103-1117.

[4] Shankar, K., Lakshmanaprabu, S.K., Gupta, D., Maseleno, A. and de Albuquerque, V.H.C., 2021. Optimal feature-based multi-kernel SVM approach for diabetic retinopathy classification on retinal images. *Multimedia Tools and Applications*, 78(20), pp.29901-29914.

[5] Rao, A., Jain, S., Sharma, M. and Srivastava, S., 2021. Comparative analysis of CNN architectures for diabetic retinopathy detection. *Journal of King Saud University-Computer and Information Sciences*, 33(2), pp.130-135.

[6] Cheng, J., Liu, J., Xu, Y., Yin, X., Wang, Q., Li, X., Fan, W. and Yang, J., 2020. Multi-path convolutional neural network for diabetic retinopathy classification. *IEEE Access*, 8, pp.2002-2010.

[7] Yang, J., Ren, J., Zhang, Y. and Wang, Q., 2022. Feature attention network for diabetic retinopathy classification. *Proceedings of the 2022 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp.2141-2150.

[8] Costa, P., Galdran, A., Meyer, M.I., Niemeijer, M., Abràmoff, M., Mendonça, A.M. and Campilho, A., 2020. End-to-end adversarial retinal image synthesis. *IEEE Transactions on Medical Imaging*, 39(3), pp.586-598.

[9] Zhao, X., Li, Y., Sun, X. and Zhou, X., 2022. Diabetic retinopathy detection using capsule networks with multi-scale feature extraction. *Pattern Recognition Letters*, 157, pp.98-105.

[10] Liu, H., Zhang, J., Lu, K. and Wang, Y., 2021. Detection of diabetic retinopathy using convolutional neural networks. *Multimedia Tools and Applications*, 80(13), pp.19835-19850.

[11] Das, S., Saha, S., Pradhan, R. and Ghosh, R., 2020. Diabetic retinopathy detection using deep learning approach. *Proceedings of the 2020 IEEE Calcutta Conference (CALCON)*, pp.127-131.

[12] Mateen, M., Wen, J., Nasrullah, N., Zafar, A., Beg, M.O. and Wang, H., 2021. Automated diabetic retinopathy severity grading using CNN with attention mechanism. *Pattern Recognition Letters*, 137, pp.124-132.

[13] Yu, X., Wang, L., Zhou, J., Zhang, Q. and Zhang, H., 2022. A lightweight deep learning model for diabetic retinopathy screening in mobile environments. *Journal of Biomedical Informatics*, 127, p.103987.

[14] Zhou, L., Pan, J., Huang, W., Cheng, L. and Zhang, J., 2022. Early detection of diabetic retinopathy using a vision transformer-based model. *Proceedings of the 2022 International Conference on Machine Learning and Applications (ICMLA)*, pp.890-897.

[15] Bajwa, M.N., Younis, A., Qureshi, R.J., Malik, S.A., Siddiqui, S.A., Khan, S.A. and Lee, B., 2020. Deep learning-based hybrid intelligent system for detecting diabetic retinopathy. *Computers in Biology and Medicine*, 121, p.103759.

[16] Wang, H., Xu, Y., Chen, Y. and Zhang, J., 2021. Multi-task learning for diabetic retinopathy classification and grading. *Proceedings of the 2021 IEEE International Conference on Image Processing (ICIP)*, pp.4453-4457.

[17] An, G., Yu, J., Zhang, Y., Zhang, L., Zhang, Y. and Zhang, Y., 2020. A novel method for blood vessel detection from retinal images. *Biomedical Signal Processing and Control*, 55, p.101641.

[18] Gargeya, R. and Leng, T., 2017. Automated identification of diabetic retinopathy using deep learning. *Ophthalmology*, 124(7), pp.962-969.

[19] Kaur, S., Saini, B.S. and Gupta, S., 2020. Automated detection of diabetic retinopathy using CNN and SVM.

Proceedings of the 2020 International Conference on Advances in Computing, Communication & Materials (ICACCM), pp.101-105.
[20] Ting, D.S.W., Peng, L., Varadarajan, A.V., 2021. Clinical validation of transformer-based detection techniques. *Progress Eye Reviews*, 80, pp.112-120.
[21] Hemanth, D.J., Anitha, J., Naaji, A., Kose, U. and Lafata, P., 2021. A comprehensive review on diabetic retinopathy detection using deep learning models. *Computers in Biology and Medicine*, 135, p.104418.
[22] Shankar, K., Lakshmanaprabu, S.K., Gupta, D., Maseleno, A. and de Albuquerque, V.H.C., 2021. Optimal feature-based multi-kernel SVM approach for diabetic retinopathy classification on retinal images. *Multimedia Tools and Applications*, 78(20), pp.29901-29914.
[23] H. S. Hemanth Kumar, Y. P. Gowramma, S. H. Manjula, D. Anil and N. Smitha, "Comparison of various ML and DL Models for Emotion Recognition using Twitter," *2021 Third International Conference on Intelligent Communication Technologies and Virtual Mobile Networks (ICICV)*, Tirunelveli, India, 2021, pp. 1332-1337, doi: 10.1109/ICICV50876.2021.9388522.
[24] Yang, J., Ren, J., Zhang, Y. and Wang, Q., 2022. Feature attention network for diabetic retinopathy classification. *Proceedings of the 2022 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp.2141-2150.
[25] Liu, H., Zhang, J., Lu, K. and Wang, Y., 2021. Detection of diabetic retinopathy using convolutional neural networks. *Multimedia Tools and Applications*, 80(13), pp.19835-19850.
[26] S. Mathapati, D. Anil, R. Tanuja, S. H. Manjula and K. R. Venugopal, "COSINT: Mining Reasons for Sentiment Variation on Twitter using Cosine Similarity Measurement," *2018 10th International Conference on Information Technology and Electrical Engineering (ICITEE)*, Bali, Indonesia, 2018, pp. 140-145, doi: 10.1109/ICITEED.2018.8534893.
[27] Savitha Mathapati, Anil D., S. H. Tanuja R, and C. N. S. M. Manjula and Venugopal KR. "Cosine and N-Gram Similarity Measure to Extract Reasons for Sentiment Variation on Twitter." *International Journal of Computer Engineering & Technology* 9, no. 2 (2018): 150-161.
[28] Anil, D., Suresh, S. (2023). Dual Sentiment Analysis for Domain Adaptation. In: Kumar, S., Hiranwal, S., Purohit, S., Prasad, M. (eds) Proceedings of International Conference on Communication and Computational Technologies. ICCCT 2023. Algorithms for Intelligent Systems. Springer, Singapore.
[29] Rao, A., Jain, S., Sharma, M. and Srivastava, S., 2021. Comparative analysis of CNN architectures for diabetic retinopathy detection. *Journal of King Saud University-Computer and Information Sciences*, 33(2), pp.130-135.
[30] Alyoubi, W.L., Shalash, W.M. and Abulkhair, M.F., 2020. Diabetic retinopathy fundus image classification and lesions localization system using deep learning. *Sensors*, 20(11), p.3110.
[31] Ting, D.S.W., Peng, L. and Varadarajan, A.V., 2021. Deep learning in ophthalmology: The technical and clinical considerations. *Progress in Retinal and Eye Research*, 77, p.100843.
[32] Yu, X., Wang, L., Zhou, J., Zhang, Q. and Zhang, H., 2022. A lightweight deep learning model for diabetic retinopathy screening in mobile environments. *Journal of Biomedical Informatics*, 127, p.103987.
[33] Zhou, L., Pan, J., Huang, W., Cheng, L. and Zhang, J., 2022. Early detection of diabetic retinopathy using a vision transformer-based model. *Proceedings of the 2022 International Conference on Machine Learning and Applications (ICMLA)*, pp.890-897.
[34] Costa, P., Galdran, A., Meyer, M.I., Niemeijer, M., Abràmoff, M., Mendonça, A.M. and Campilho, A., 2020. End-to-end adversarial retinal image synthesis. *IEEE Transactions on Medical Imaging*, 39(3), pp.586-598.
[35] Kaur, S., Saini, B.S. and Gupta, S., 2020. Automated detection of diabetic retinopathy using CNN and SVM. *Proceedings of the 2020 International Conference on Advances in Computing, Communication & Materials (ICACCM)*, pp.101-105.
[36] Cheng, J., Liu, J., Xu, Y., Yin, X., Wang, Q., Li, X., Fan, W. and Yang, J., 2020. Multi-path convolutional neural

network for diabetic retinopathy classification. *IEEE Access*, 8, pp.2002-2010.
[37] Shankar, K., Lakshmanaprabu, S.K., Gupta, D., Maseleno, A. and de Albuquerque, V.H.C., 2021. Optimal feature-based multi-kernel SVM approach for diabetic retinopathy classification on retinal images. *Multimedia Tools and Applications*, 78(20), pp.29901-29914.
[38] An, G., Yu, J., Zhang, Y., Zhang, L., Zhang, Y. and Zhang, Y., 2020. A novel method for blood vessel detection from retinal images. *Biomedical Signal Processing and Control*, 55, p.101641.
[39] D. Anil, M. N. Sindhushree, T. M and T. A. Lone, "Disease Detection and Diagnosis on the Leaves using Image Processing," *2022 IEEE 3rd Global Conference for Advancement in Technology (GCAT)*, Bangalore, India, 2022, pp. 1-6, doi: 10.1109/GCAT55367.2022.9972010.
[40] R. Ranjitha, P. Ahuja, R. Shreeshayana and D. Anil, "Edge Intelligence for Traffic Flow Detection: A Deep Learning Approach," *2023 International Conference on Quantum Technologies, Communications, Computing, Hardware and Embedded Systems Security (iQ-CCHESS)*, KOTTAYAM, India, 2023, pp. 1-6, doi: 10.1109/iQ-CCHESS56596.2023.10391493.
[41] D. Anil, S. Hiremath and S. H. Manjula, "Detection of Efficient Landslide Inventory Mapping in Western Ghats regions of Karnataka, India," *2024 International Conference on Knowledge Engineering and Communication Systems (ICKECS)*, Chikkaballapur, India, 2024, pp. 1-6, doi: 10.1109/ICKECS61492.2024.10617039.