Engineered Science
DOI: https://dx.doi.org/10.30919/es1597

Hybrid Fully Connected Neural-Bidirectional Long Short-Term Memory Networks for Diabetic Complication Risk Prediction

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

Abstract

Accurate estimation of the risk level of chronic diseases such as heart disease, kidney disease, and retinopathy is necessary for early detection and appropriate treatment planning for diabetic patients. The traditional machine learning models, *i.e.*, Bidirectional long short-term memory (BiLSTM) and fully connected neural networks (FCNN), have been extensively used for disease prediction but are typically burdened with high computational complexity and redundant feature dependencies. In this research work, a new prediction model is proposed based on structured text data from CSV files to determine the levels of disease risk. The proposed method outperforms BiLSTM and FCNN in classification performance, with improved performance metrics. In addition, we also performed feature reduction using random forest (RF) and explainable artificial intelligence (XAI) techniques, such as SHAP (SHapley Additive exPlanations), with the objective of obtaining the most informative features. Regardless of feature reduction, the proposed system still maintains the best performance, confirming its efficacy and resilience in risk prediction. The outcomes reveal the potential for combining advanced deep learning models with feature selection techniques to improve diabetic complication disease risk assessment and prediction.

Keywords: Fully connected neural networks; Shapley additive explanations; Bidirectional long short-term memory network; Feature reduction; Random Forest.
Received: 02 April 2025; Revised: 19 May 2025; Accepted: 12 June 2025.
Article type: Research article.

1. Introduction

Various studies have studied the use of deep learning (DL) models for disease prediction, specifically cardiovascular and renal diseases. Bidirectional long short-term memory (BiLSTM) has found extensive use in sequence-based medical data due to its capability to learn long-term dependencies.[1] It has been shown to successfully predict heart disease progression based on patient history and diagnostic reports. BiLSTM models are, however, computationally intensive and sensitive to the need for a large amount of labelled data. Fully connected neural network (FCNN) has also been used in other disease classification problems, usually as a baseline.[2] FCNN models are well-suited for structured medical data, but are not capable of modeling sequential relationships and hence do not perform as well in time-series or progressive disease prediction.

Feature selection techniques have been commonly applied to enhance model efficiency and interpretability. Feature selection based on Random Forest has been applied in medical data to find significant risk factors and simplify the model.[3] Explanation techniques like SHapley Additive exPlanations (SHAP) have also been applied to provide interpretable details regarding feature importance such that clinicians can make sense of model decisions.[4] The union of these strategies in DL models is still an ongoing challenge because of feature reduction at the cost of predictive power. The proposed system builds on these previous methods by designing a hybrid prediction model that balances between model complexity, interpretability, and accuracy. Through the use of DL and feature selection, the proposed system provides a robust and scalable framework for disease risk level classification.

Recent studies have explored a wide range of machine

[1] *Department of Computer Engineering, Vishwakarma Institute of Technology, Savitribai Phule Pune University, Maharashtra, Pune, 411037, India*
[2] *Department of Computer Science & Engineering (Artificial Intelligence & Machine Learning), Vishwakarma Institute of Technology, Savitribai Phule Pune University, Maharashtra, Pune, 411037, India*
**Email:* bhanuseanita@gmail.com (A. B. Dombale)

learning (ML) and DL techniques to improve prediction accuracy for diabetic complications. Ma *et al*.[5] combined ML and DL approaches for diabetic nephropathy (DN) risk prediction by employing data preprocessing and feature selection techniques, building ten ML models using the most relevant variables. Mesquita *et al*.[6] validated the predictive power of ML models for DN using temporally embedded clinical data and emphasized the potential of DL and explainable artificial intelligence (XAI) techniques such as SHAP for future studies. Du *et al*.[7] proposed a hybrid model for diabetes prediction, focusing on balancing precision and computational cost, which can be adapted for DN with added feature selection mechanisms. Sudha *et al*.[8] and Gomaa Elshafie *et al*.[9] explored long short-term memory-convolutional neural network (CNN-LSTM) and BiLSTM-1D CNN hybrids for heart disease prediction, suggesting future improvements through attention mechanisms and SHAP-based interpretability. Zhao *et al*.[10] introduced a CNN-BiLSTM framework using a sliding time-window and improved activation functions, with potential applications in heart disease and feature-efficient modeling. Haq *et al*.[11] highlighted the need for DL models in diabetic retinopathy that optimize both accuracy and computational efficiency. Similarly, Atwany *et al*.[12] reviewed CNN and vision Transformer (ViT)-based models, pointing out challenges like interpretability and data limitations. Das *et al*.[13] suggested an attention-based BiLSTM-CNN architecture on electronic health records (EHRs), which can be adapted for diabetic retinopathy using feature selection for better performance. Ljubic *et al*.[14] demonstrated that gated recurrent unit (GRU)-based models outperform others in predicting multiple diabetic complications from EHRs. Alfian *et al*.[15] used recursive feature elimination (RFE) with deep neural networks (DNN) for early DR detection, achieving high accuracy and generalization. Artificial neural network (ANN) techniques were also found to be applicable.[16] Aliyu *et al*.[17] introduced an approach using particle swarm optimization (PSO) for data balancing and genetic algorithms for model tuning, showing significant improvements in predictive performance. XGBoost emerged as the most efficient model with minimal training time.[18] These findings collectively suggest that hybrid models, attention mechanisms, and feature selection techniques can significantly enhance the accuracy, interpretability, and efficiency of diabetic complication prediction systems.

2. Proposed system

The proposed system in Fig. 1 is implemented for nephropathy, heart disease, and retinopathy risk level classification for diabetic patients from text-based data. The data was split into training and test sets based on an 80:20 ratio, 80% of the data being utilized for model training and the rest 20% being kept for testing. United Kingdom Prospective Diabetes Study(UKPDS),[19] Action in Diabetes and Vascular Disease: Preterax and Diamicron MR Controlled Evaluation(ADVANCE),[20] and Diabetes Control and Complications Trial / Epidemiology of Diabetes Interventions and Complications (DCCT/EDIC),[21] these studies revealed that: Microvascular complications such as retinopathy enhance the risk of macrovascular complications (*e.g.*, coronary artery disease). Diabetic nephropathy patients tend to also present with retinopathy and increased cardiovascular mortality.[22] The severity of retinopathy correlates directly with: Decrease in eGFR (renal function), left ventricular dysfunction (cardiac) predictive models are studied.[23] It uses a combination of CNN, BiLSTM, and attention mechanisms in a hybrid DL model for feature extraction and sequence learning. The dataset has structured categorical features and text description based on progression for disease, which is preprocessed and tokenized to be used with DL.

The category columns are encoded with label encoding, and the text descriptions are tokenized and transformed into padded sequences for consistent input length. A hybrid DL structure is utilized, where an embedding layer transforms the input text into dense vector representations. A conv1D layer learns high-level features, followed by a MaxPooling layer for reducing dimensionality. The BiLSTM layer learns long-range dependencies in the sequence. To improve the model's power of attention over major features, an Attention mechanism is used for the BiLSTM outputs. The extracted features are then output through a fully connected layer and a Softmax output layer for classifying the process into three levels of risk.

The model is trained with categorical cross-entropy loss and Adam optimizer, and the performance is monitored using accuracy, confusion matrix, and classification reports. For explainability, SHAP is utilized, which identifies the most significant features impacting classification. The system makes sure that even after feature reduction (via Random Forest (RF) and SHAP-XAI), the model has optimal performance, proving to be robust in prediction. The steps of the mathematical model are as follows.

Step 1: Text embedding representation

Each input text sequence $\boldsymbol{X}$ is converted into an embedding representation using a word embedding layer in Eq. (1):

$$\boldsymbol{E}(\boldsymbol{X}) = \boldsymbol{w_e}.\boldsymbol{X} \tag{1}$$

where $\boldsymbol{w_e} \in \boldsymbol{R}^{v \times d}$ is the embedding matrix (with vocabulary

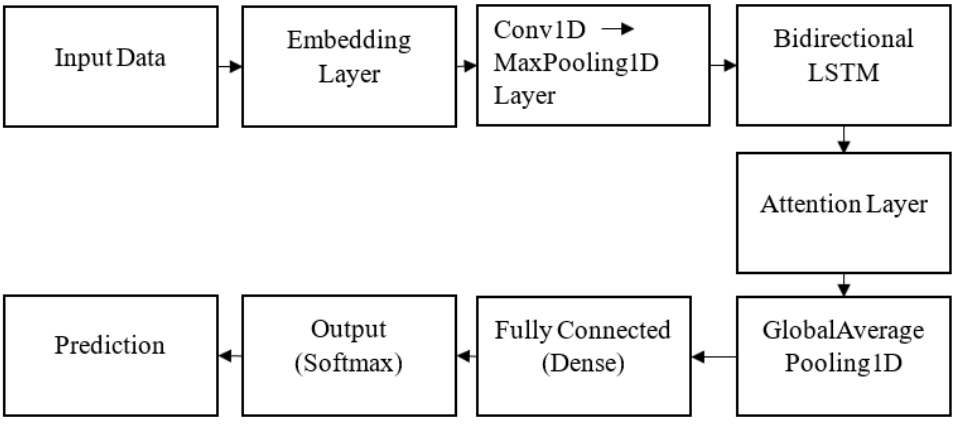

Fig. 1: Proposed system for prediction.

size $\boldsymbol{v}$ and embedding dimension $\boldsymbol{d}$), $\boldsymbol{X}$ represents the tokenized input text sequence. The embedded representation $\boldsymbol{E(X)}$ has dimensions $\boldsymbol{R^{v\times d}}$, where T is the sequence length.

Step 2: Convolutional feature extraction

A 1D Convolutional layer extracts local features in Eq. (2):

$$\boldsymbol{h_{c=f(w_c*E(X)+b_c)}} \tag{2}$$

where $\boldsymbol{w_c} \in \boldsymbol{R^{k\times d}}$ is the convolutional kernel of size $\boldsymbol{k}$, $*$ denotes the convolution operation, $\boldsymbol{b_c}$ is the bias term, $\boldsymbol{f}$ is the activation function (ReLU).

After convolution, max-pooling is applied to reduce dimensionality in Eq. (3):

$$\boldsymbol{h_{p=^{max}_i h_c}} \tag{3}$$

Step 3: BiLSTM for context learning

Bidirectional LSTM (BiLSTM) is applied to capture forward and backward dependencies in the sequence:

Forward LSTM Eq. (4):[24]

$$h_{t=}\sigma(W_f h_{t-1} + U_f X_t + b_f) \tag{4}$$

Backward LSTM Eq. (5):[25]

$$h_{t=}\sigma(W_b h_{t+1} + U_b X_t + b_b) \tag{5}$$

The final BiLSTM hidden representation is obtained by concatenating both directions:

Step 4: Attention mechanism

The attention mechanism assigns weights to different time steps to enhance interpretability and focus on important features. Attention score computation is in Eq. (6):

$$e_t = h_t^T W_\alpha h_t \tag{6}$$

where W_α is the attention weight matrix.

Softmax normalization to obtain attention weights is in Eq. (7):

$$\alpha_t = \frac{\exp(e_t)}{\sum_{t'} \exp(e_{t'})} \tag{7}$$

The context vector (weighted sum of hidden states) is in Eq. (8):

$$c = \sum_t \alpha_t \, h_t \tag{8}$$

Step 5: Fully connected layer

The context vector is passed through a fully connected (dense) layer is in Eq. (9):[26]

$$h_{fc} = ReLU(W_{fc} \, . \, c + b_{fc}) \tag{9}$$

where W_{fc} and b_{fc} are the weight and bias of the dense layer.

Step 6: Softmax classifier for risk prediction

The final classification probabilities are computed using a softmax function,[27] as shown in Eq. (10):

$$y_{pred} = softmax(W_o + h_{fc} + b_o) \tag{10}$$

where W_o and b_o are the weight and bias for the output layer. The predicted class is given by Eq. (11):

$$\hat{y} = argmax(y_{pred}) \tag{11}$$

Step 7: Loss function (Categorical cross-entropy)

Since the model performs multiclass classification, categorical cross-entropy loss is used, which is shown in Eq. (12):

$$L = -\sum_{i=1}^{c} \log y_i \, log \, \hat{y}_i \tag{12}$$

where C is the number of classes (3 in this case), y_i is the true one-hot encoded label, and $\hat{y}_i$ is the predicted probability.

3. Results and discussion

The classification results from the use of the chosen features for heart disease, nephropathy, and diabetic retinopathy risk level prediction show an excellent degree of accuracy, with

precision, recall, and F1-score all scoring a perfect 1.00 for all classes. This shows that the model has been able to pick up patterns in the data, enabling it to classify cases with absolute accuracy. The characteristics used in this categorization are age, gender, bmi, duration of diabetes, family history, level of hba1c, fasting blood glucose, postprandial blood glucose, use of insulin, use of oral medications, blood pressure (systolic and diastolic), cholesterol, triglycerides, resting heart rate, history of cardiac disease, ecg abnormality, creatinine level, egfr, urine albumin-creatinine ratio, stage of kidney disease, retinal thickness, optic cup-to-disc ratio, diabetic retinopathy, macular edema, progression of heart disease, progression of nephropathy, deterioration of kidney function, and progression of eye condition. Such features encompass characteristics broadly encompassing clinical, biochemical, and imaging-related factors with a properly balanced perspective towards classifying disease.

The optimal classification accuracy implies that the model can clearly come up with decisive observations based on these characteristics and make extremely realistic predictions. The accuracy level indicates that the data is set in the correct manner, and there are distinctive patterns that enable the model to determine various risk levels with clarity. The robust performance can also be a reflection of the efficacy of the selected machine learning or DL method. Additional confidence in the model's stability can also be sought with independent test dataset validation, cross-validation, and interpretability methods like SHAP analysis. These protocols will assist in ensuring that the model's predictions are based on significant patterns and validate its reliability for wider use in disease risk estimation and disease progression tracking.

3.1 For nephropathy

Fig. 2 presents the confusion matrix of diabetic nephropathy progression of a suggested system. The model achieved a perfect classification with 100% accuracy and no false positives or false negatives. This shows that the system works exceedingly well in discriminating between the three risk levels without misclassification.

Table 1 gives the classification report of diabetic nephropathy progression of a proposed system. The classification report demonstrates 100% model performance in terms of accuracy, precision, recall, and F1-score for all three classes (0, 1, and 2). The model accurately classified all 200 test samples with no misclassifications. This reflects an optimal and well-generalized risk level prediction system.

Fig. 3 gives the accuracy and loss graph of diabetic nephropathy progression of a proposed system. The plots represent the model's training in terms of loss and accuracy

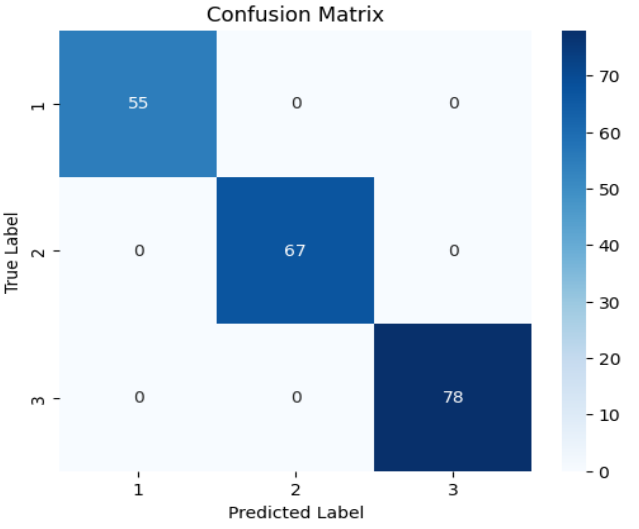

Fig. 2: Confusion matrix of diabetic nephropathy progression of the proposed system.

Table 1: Classification report of the diabetic nephropathy progression of the proposed system.

Classification report	Precision	Racall	F1-score	Support
0	1.00	1.00	1.00	66
1	1.00	1.00	1.00	65
2	1.00	1.00	1.00	69
Accuracy	-	-	1.00	200
Macro-avg	1.00	1.00	1.00	200
Weightes avg	1.00	1.00	1.00	200

over 32 epochs. The left plot depicts how the training loss and validation loss both drop sharply and reach close to zero in the initial few epochs, pointing towards rapid and stable learning. The right plot indicates that the training accuracy and validation accuracy both rise rapidly and are at 100% with no indication of overfitting or underfitting. The compact overlap of validation and training curves indicates that the model has high generalizability to unseen information. Generally, the results ensure a highly precise and effective model for predicting the risk level.

3.2 Results comparison

Similar to nephropathy, results were also obtained for retinopathy and heart disease. The model's performance was evaluated for all three conditions, *i.e.*, nephropathy, heart disease and retinopathy for diabetic patients, and the corresponding graphs for precision, recall and F1-score were presented in Fig. 4, Fig. 5, and Fig. 6, respectively., demonstrating the system's effectiveness in predicting disease risk levels. In Fig. 4, the chart compares the performance of three models, BiLSTM, FCNN, and a proposed system, across Classes 0, 1, and 2 using precision, recall, and F1-score

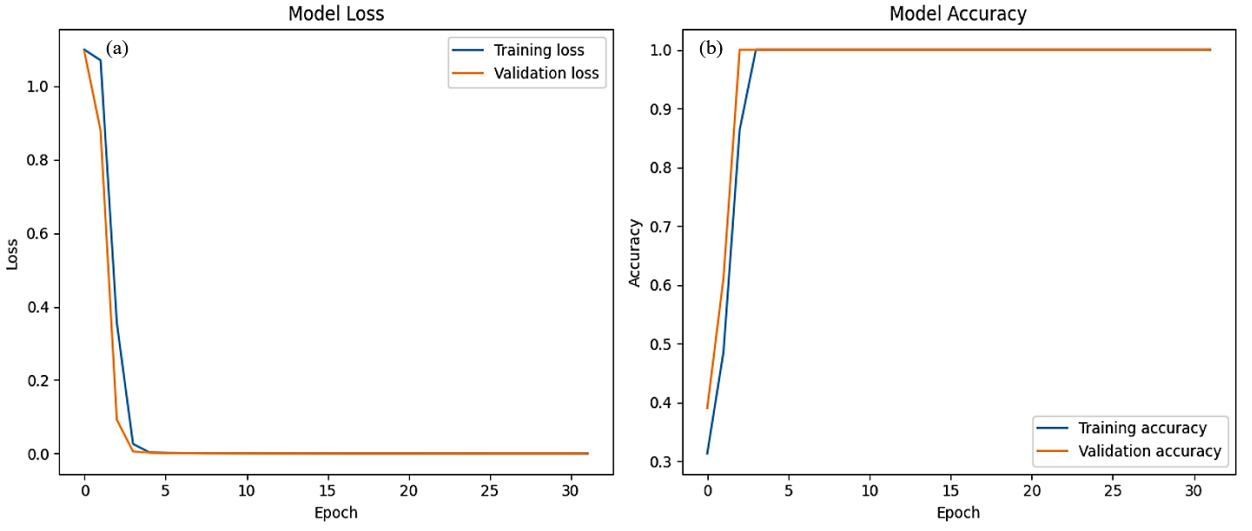

Fig. 3: The model loss (a) and accuracy (b) graphs of diabetic nephropathy progression of the proposed system.

for diabetic nephropathy. BiLSTM performs well on Class 2 with recall (0.61), while Class 0 shows moderate performance. Class 1 is showing low performance. FCNN performs strongly on Class 1 as compared to Class 0 and Class 1. The proposed system performs extremely well for all classes, with all metrics 1.

In Fig. 5, The chart depicts the same model's performance by three classes with precision, recall, and F1-score for Diabetic Heart Disease. for BiLSTM, On Class 0, the model has a precision of 0.55, recall of 0.28, and an F1-score of 0.38, reflecting poor capability to identify and retrieve correctly. In Class 1, the performance is marginally better with precision at 0.48, recall at 0.28, and F1-score at 0.35, also reflecting moderate precision but poor sensitivity. For Class 2, the model does significantly better with recall, registering at 0.71, with precision being 0.38 and F1-score being 0.50. This indicates that the model is better at identifying actual instances of Class 2 at the expense of precision. In general, the model does poorly in consistency for all classes, especially with recall for Classes 0 and 1, with the relatively better recall being recorded for Class 2. For FCNN, in Class 0, precision is about 0.51, recall is about 0.28, and F1-score is about 0.39. For Class 1,

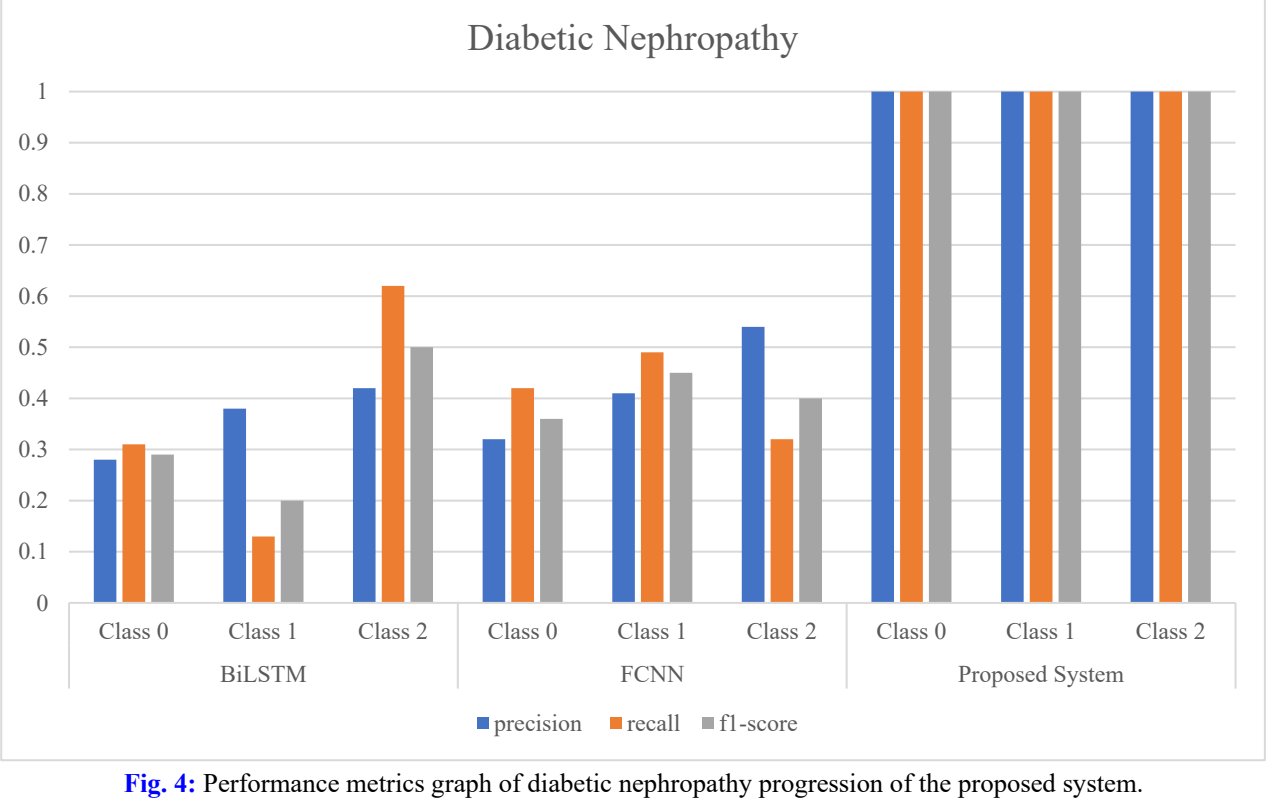

Fig. 4: Performance metrics graph of diabetic nephropathy progression of the proposed system.

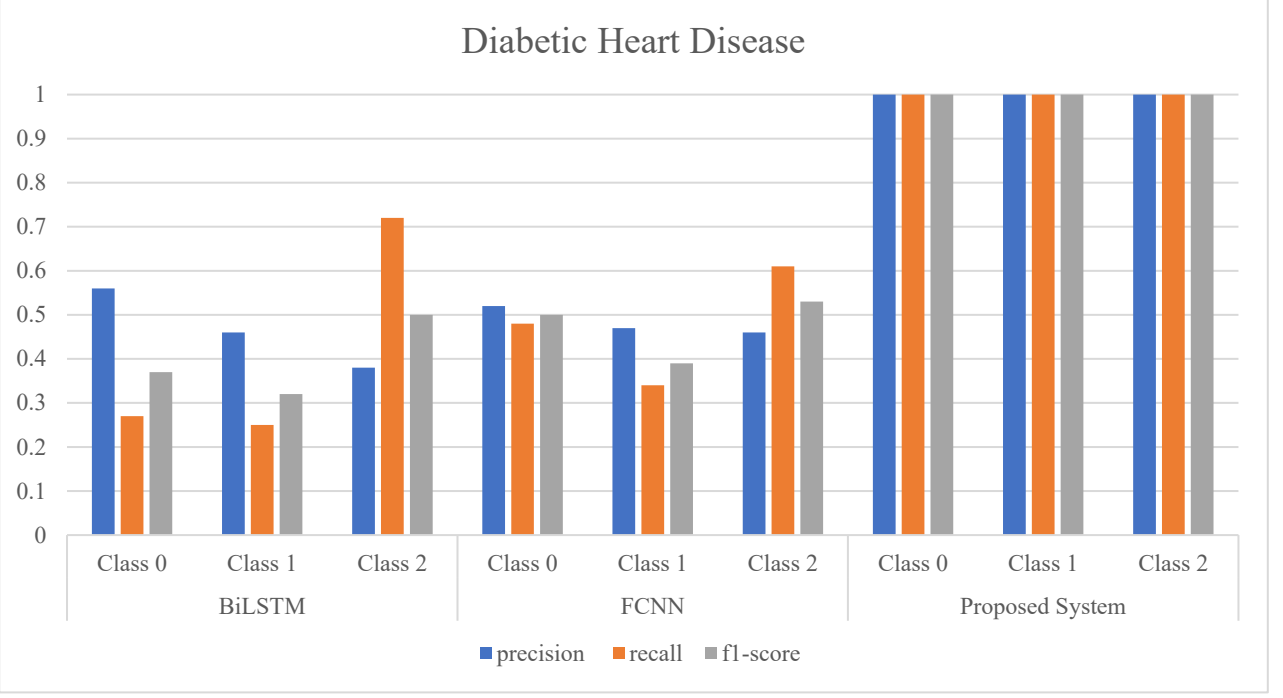

Fig. 5: Performance metrics graph of diabetic heart progression of the proposed system.

precision is about 0.48, recall is 0.29, and F1-score is about 0.37. In Class 2, recall is highest at about 0.72, F1-score is about 0.56, and precision is about 0.42. This is an improvement in the high rate of recall for Class 2 over the other classes. The proposed system has exemplary performance with 100% precision, recall, and F1-score on all three classes, showing that it detects every instance accurately with no false positives or false negatives.

In Fig. 6, for BiLSTM, Class 0 contains no visible bars, reflecting zero or minimal scores. For Class 1 and Class 2, recall is highest 0.60, implying the model identifies most positive instances. Precision is lower 0.39, reflecting more false positives. The F1 score, which balances both, is moderate 0.48. Overall, the model prefers recall over precision, with scope for improvement in prediction accuracy. For FCNN, Class 0, the recall is maximum 0.72, with precision and F1 measure approximating 0.38 and 0.5. Class 1 has a high precision value 0.58 but an extremely low recall value 0.15, resulting in a low F1 measure 0.25. Class 2 exhibits balanced performance with all values around 0.35. Again, the proposed

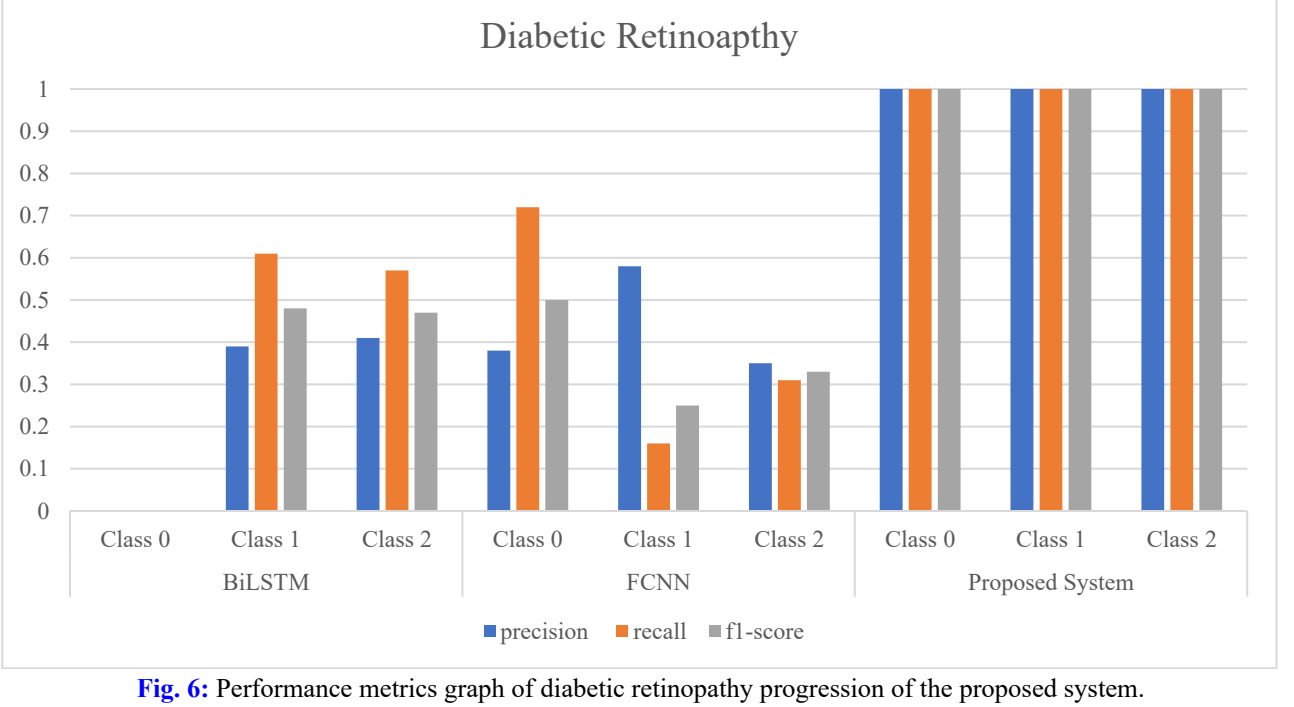

Fig. 6: Performance metrics graph of diabetic retinopathy progression of the proposed system.

system performs extremely well for all classes, with all metrics 1.

3.3 Feature reduction for nephropathy progression

This study uses an RF classifier and SHAP for feature selection to determine the most influential features for the prediction of heart disease progression, nephropathy progression, and eye condition progression. The minimized set of features is stored for reuse. Feature reduction enhances model efficiency and reliability by removing extraneous or redundant information.

To carry out feature selection, a function feature selection (target) is created, which accepts every target condition as an argument and returns the most significant features to predict it. Within this function, the dataset is divided into training and test sets based on an 80-20% ratio to facilitate effective model assessment. An RF classifier with 100 decision trees is trained on the dataset to learn patterns and calculate feature importance. The model computes an importance score for every feature, and features with a score above 0.01 are kept, thereby removing less important variables. A bar chart representing feature importance scores is plotted to visualize this process using Matplotlib.

To further understand how the features are impacting predictions, SHAP is used to explain the model's decision-making process. A SHAP summary plot is produced to illustrate the effect of every feature on the predictions. Also, SHAP dependence Plots are drawn for important features such as Age, BMI, and Gender, giving insights into their separate impact on the result. As SHAP interaction values are not available for Random Forest models, this step is omitted.

After identifying the most influential features for each target condition, they are kept in a dictionary for convenience. The script then loops through all three target conditions, running the feature selection function for each and building the chosen features. The feature set is then reduced and saved as a CSV file to be used for further model training and analysis.

This procedure is used to keep only the most significant features, enhancing the performance of disease risk prediction models with better efficiency and retaining interpretability. By incorporating RF feature selection and SHAP explainability, the methodology not only boosts model performance but also offers concise explanations of what drives disease development. From the total 29 available features, 19 features were chosen for heart disease risk prediction, 21 features were chosen for nephropathy risk prediction, and 21 features were chosen for retinopathy risk prediction.

Fig. 7 bar graphically displays the feature importance of predicting nephropathy progression with an RF classifier. The x-axis contains the features employed by the model, and the y-

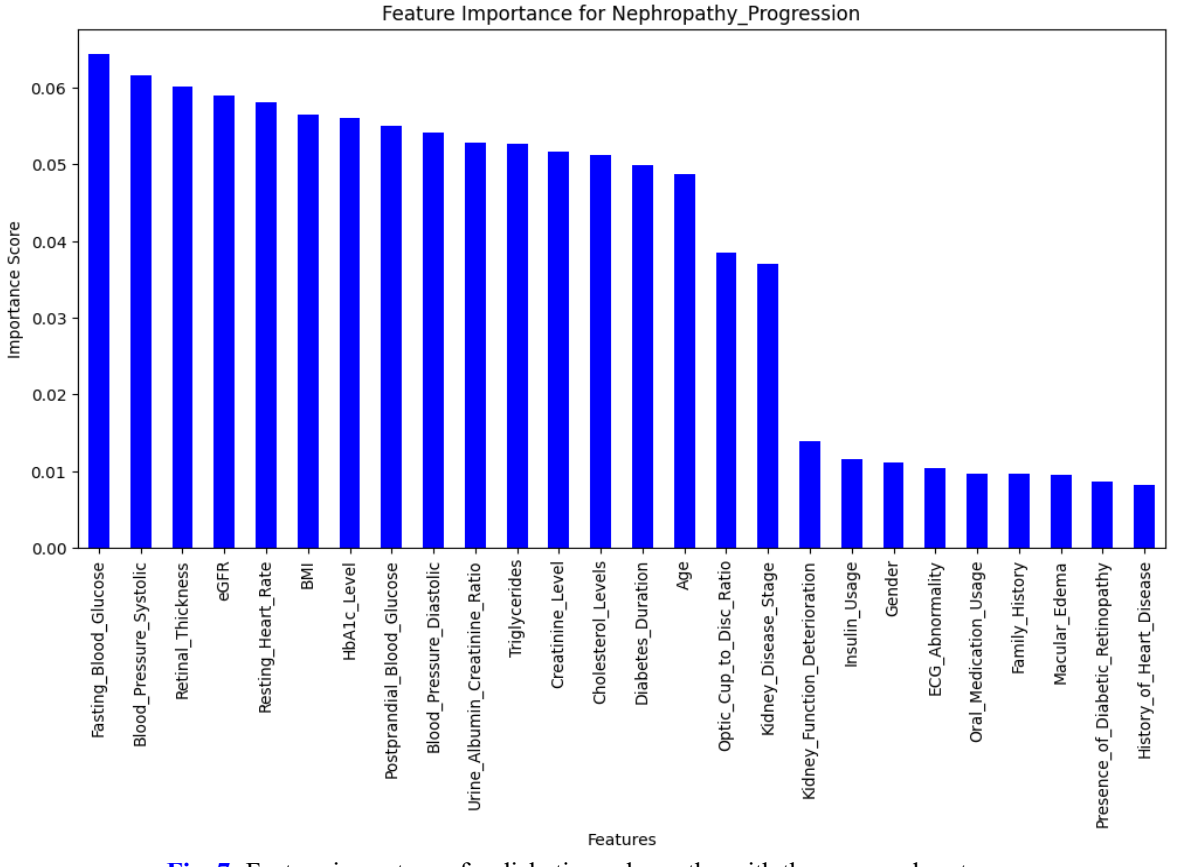

Fig. 7: Feature importance for diabetic nephropathy with the proposed system.

progression. The study corroborates that the most potent predictors of nephropathy progression are blood glucose levels, axis contains the importance score, representing how significantly each feature will help predict nephropathy blood pressure, renal function markers (eGFR, creatinine), and lipid parameters (triglycerides, cholesterol). This knowledge can be used to improve predictive models by concentrating on these most significant variables, increasing precision at the expense of decreased computational complexity.

The SHAP summary plot obtained in Fig. 8 gives a closer explanation of model prediction for nephropathy progression by showing the contribution of different features towards the output. The y-axis mentions the ranked features in order of descending importance, and systolic blood pressure, fasting blood glucose, diastolic blood pressure, eGFR, and retinal thickness are the top five most important features in predicting nephropathy risk. The x-axis is the SHAP values, which tell us the effect of each feature on the output of the model. Negative SHAP values nudge the prediction towards a lower risk, and positive SHAP values tell us that there is a higher risk. The plot also has a color gradient, with red being high feature values and blue being low feature values.

From this analysis, some important insights can be derived. Blood pressure and glucose levels are two of the most significant factors that affect the progression of nephropathy. Elevated systolic and diastolic blood pressure levels have a strong correlation with risk, indicating the impact of hypertension on kidney disease progression. In the same manner, elevated fasting and postprandial blood glucose levels play a substantial role, reaffirming that poor blood sugar control accelerates kidney damage. Furthermore, retinal markers like retinal thickness and optic cup-to-disc ratio indicate a relationship between nephropathy and retinal abnormalities, and this also supports the link between kidney disease and diabetic retinopathy. Metabolic markers like BMI, cholesterol levels, triglycerides, and HbA1c levels also reflect a significant influence, emphasizing the significance of total metabolic well-being in nephropathy risk.

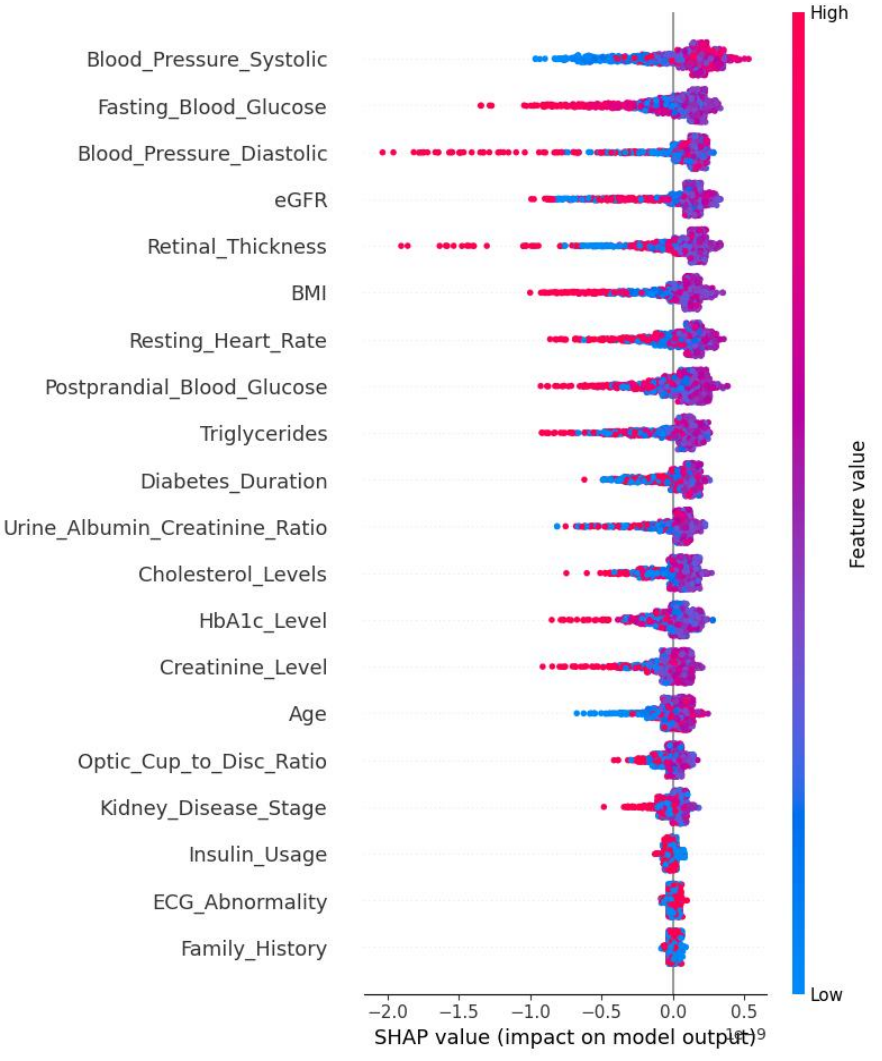

Fig. 8: SHAP feature importance for diabetic nephropathy with the proposed system.

In addition, markers of chronic kidney function, including eGFR, urine albumin-creatinine ratio, and creatinine level, are important in determining nephropathy severity. Decreasing values of eGFR (blue) reflect increasing risk of nephropathy, as would be anticipated, as worsening kidney function directly reflects increasing disease severity. The other characteristics, including duration of diabetes, use of insulin, ECG abnormality, and family history, have smaller but residual effects in predicting. In general, this SHAP analysis reinforces the fact that blood glucose control, hypertension, kidney function markers, and metabolic factors play a crucial role in increasing the risk of nephropathy. Such findings can prove useful for fine-tuning treatment approaches, allowing for targeted interventions to counteract the threat of disease progression.

The SHAP scatter plot, shown in Fig. 9, illustrates the effect of age on the model's prediction, with cholesterol levels encoded using color. The x-axis is plotted for age, and the y-axis plots the respective SHAP values, reflecting how age affects the model's prediction. A positive SHAP value indicates that age increases the predicted risk, whereas a negative value reflects a lower risk.

The younger groups (20-30 years old) have more negative SHAP values, implying that younger age has a lower predicted risk. Yet as age rises from approximately 30 to 50 years old, SHAP values increase, implying that age adds more positively to risk predictions. The trend levels off at 50-70, where SHAP values oscillate around zero, suggesting a lesser direct effect of age on risk at these ages. For those over 70, the SHAP values have a slight fluctuation but no clear trend of increasing or decreasing.

The color gradient shows cholesterol levels, with blue being lower and red being higher. It is apparent that the higher cholesterol (red) individuals are distributed throughout all ages, but appear more common in the mid-to-older age ranges. This indicates that cholesterol status might interact with age to affect the predictions of the model. Overall, this plot emphasizes that age is an important factor in the model's risk prediction, particularly in middle age, and its effect is mediated by cholesterol status.

This obtained SHAP summary plot in Fig. 10 illustrates the effect of BMI on the predictions of the model, where color is used to denote levels of systolic blood pressure. The x-axis is used to represent BMI values, and the y-axis is used to represent the SHAP values, which show the contribution of BMI to the model's predictions. The trend indicates that BMI has a nonlinear relationship, with SHAP values rising to a BMI of approximately 30, beyond which the contribution falls. Increased systolic blood pressure (red) occurs more often in subjects with high BMI. This shows that both blood pressure and BMI play a combined role in predicting the risk, perhaps for an illness like nephropathy or cardiovascular disease.

The SHAP scatter plot in Fig. 11 plots the effect of gender on the predictions of the model with the feature of having diabetic retinopathy as a color-coded feature. The x-axis is gender, presumably encoded as 0 for one category (*e.g.*, female) and 1 for the other (*e.g.*, male). The y-axis plots the SHAP values for gender, representing the contribution of this feature to the output of the model. A positive SHAP value indicates that gender adds to the predicted risk, whereas a negative value indicates a decrease in the prediction of risk.

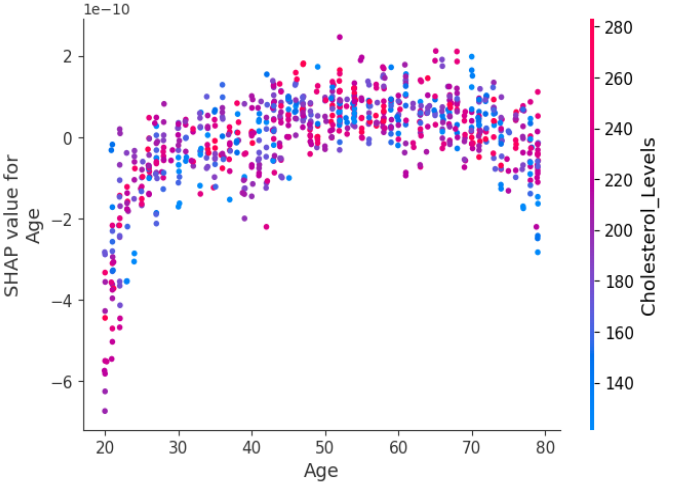

Fig. 9: SHAP feature importance of age with cholesterol level for diabetic nephropathy with the proposed system.

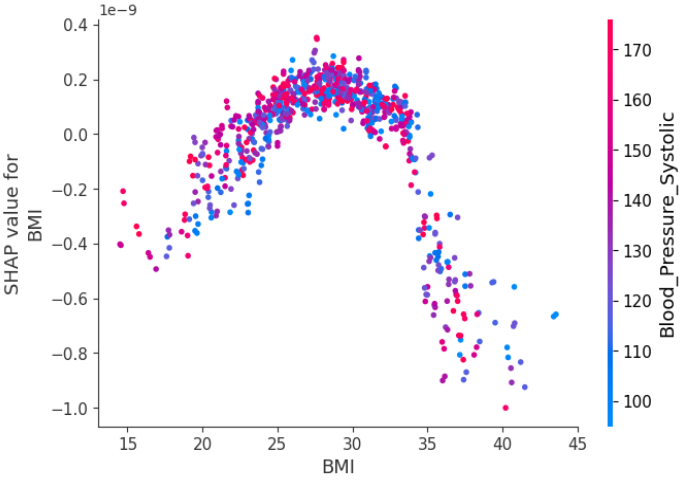

Fig. 10: SHAP feature importance of BMI with blood pressure systolic for diabetic nephropathy with the proposed system.

Therefore, SHAP values plot symmetrically around zero for both gender categories, which means that gender has a very small effect on predictions by the model. There is also no dominant trend in favor of either gender. The blue (0) and red (1) colors in the color gradient represent the presence or absence of diabetic retinopathy, which reinforces that there is

no specific distribution of color favoring a particular gender, making it certain that gender plays a negligible role in diabetic retinopathy prediction within the model. Generally, this visualization indicates that gender does not significantly contribute to the model's decision-making process for predicting diabetic retinopathy since SHAP values are near zero and do not display a clear trend between genders.

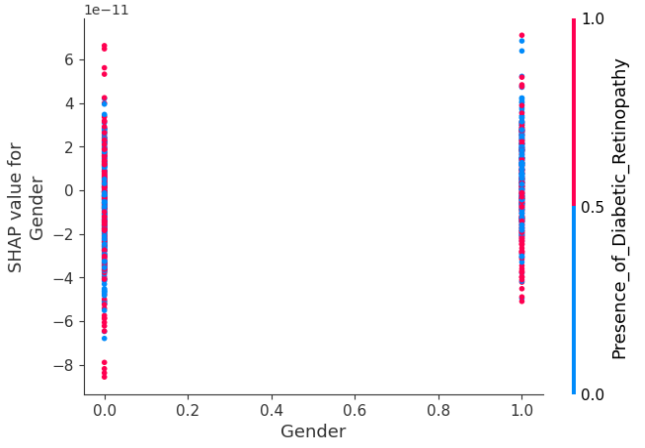

Fig. 11: SHAP feature importance of gender with presence of diabetic retinopathy for diabetic nephropathy with the proposed system.

Fig. 12 shows the feature importance for diabetic heart disease with the proposed system. The bar chart shows the importance scores of different clinical features employed in a predictive model, presumably for disease risk or progression assessment. Of the features, fasting blood glucose has the highest importance score, meaning it is the most significant factor in the decision-making process of the model. This is followed by HbA1c level and glomerular filtration rate (GFR), both of which also play important roles in the model's predictive power. Retinal thickness and urine albumin-creatinine ratio have moderate importance, indicating that they are significant in diseases like diabetic nephropathy or retinopathy. Cholesterol levels and age, on the other hand, have relatively lower importance values, yet they are also contributing to the overall performance of the model. This analysis of feature importance can assist in determining which clinical parameters are most vital for precise predictions and can influence model optimization and clinical decision-making.

Fig. 13 shows SHAP feature importance for diabetic heart disease with the proposed system. This is a SHAP summary plot that produces an interpretable visualization of feature contributions to the model's predictions. Each point on the plot is a SHAP value for an instance and feature, telling us how much each feature contributed to increasing or decreasing the prediction. The y-axis gives the features ordered by their global importance, and the x-axis indicates the magnitude of the SHAP value—how much the feature affects the model's output. The color of each point indicates the true value of the feature: red for high feature values and blue for low values.

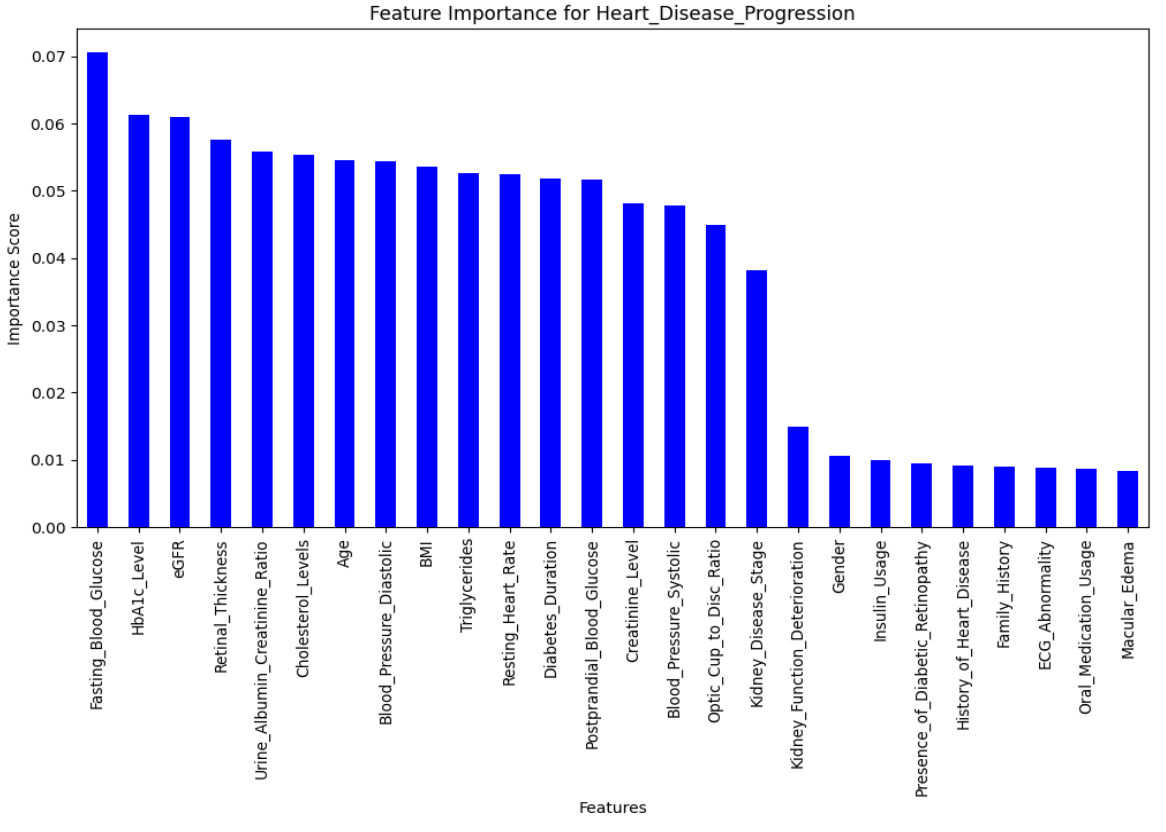

Fig. 12: Feature importance for diabetic heart disease with the proposed system.

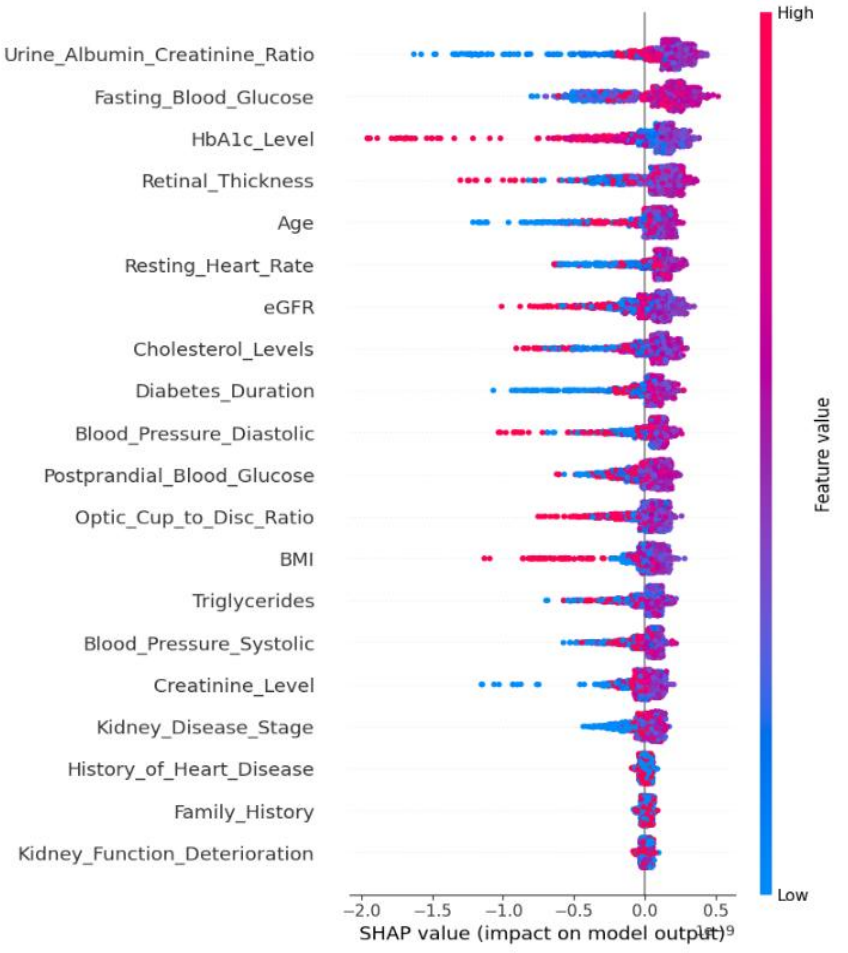

Fig. 13: SHAP feature importance for diabetic heart disease with the proposed system.

For instance, high values of Urine_Albumin_Creatinine_Ratio and Fasting_Blood_Glucose (red points moved to the right) correlate with a higher predicted risk, which means these are driving up the model's output. Low values (blue) drive it in the other direction. This plot is useful in interpreting which attributes the model to make decisions and how a single attribute value affects the decision, hence important for explainability

Fig. 14 shows the bar chart showing the importance scores of different clinical features employed in a predictive model for diabetic retinopathy with the proposed System, indicating their relative significance to the performance of the model. Among them, age is the most prominent one, which indicates that age has a significant impact on the outcome and could be indicative of its strong correlation with disease progression or risk. Fasting blood glucose and BMI are also close behind, suggesting that metabolic health markers are similarly influential in the model's decision-making process. Other factors like urine albumin-creatinine ratio and retinal thickness are of moderate significance, further supporting their use in determining conditions such as diabetic complications. Cholesterol levels and diastolic blood Pressure, on the other hand, have relatively lower importance scores, yet they also still make meaningful contributions to the model. This feature importance analysis is useful for determining the most influential variables, informing both model interpretation and clinical attention.

Fig. 15 shows feature importance for diabetic retinopathy with the proposed system, with the SHAP summary plot graphically illustrating the effect of individual feature values on the model output for a prediction task related to health. Each feature on the y-axis is sorted by its overall contribution to the model's predictions, with age and BMI being the most dominant. The x-axis is the SHAP value, a number that measures the direction and size of each feature's contribution to the prediction: values to the right increase the prediction, and values to the left decrease the prediction. Each point is an individual, and the color of each point is the feature's actual value — red for high values, blue for low values. For instance, greater values of age and BMI (red dots right) will lift the model's prediction, maybe reflecting greater risk. Likewise, retinal thickness, blood pressure (systolic and diastolic), and fasting blood glucose also affect the prediction to a greater or lesser extent based on their values. This plot not only ranks

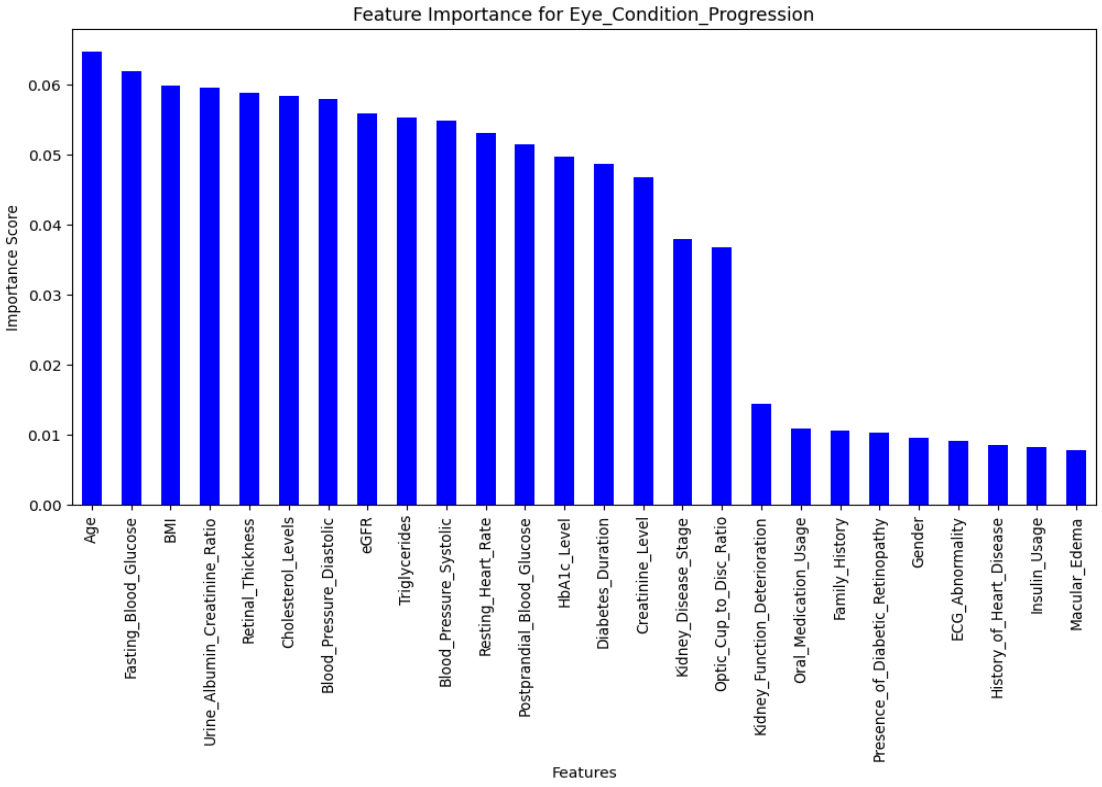

Fig. 14: Feature importance for diabetic retinopathy with the proposed system.

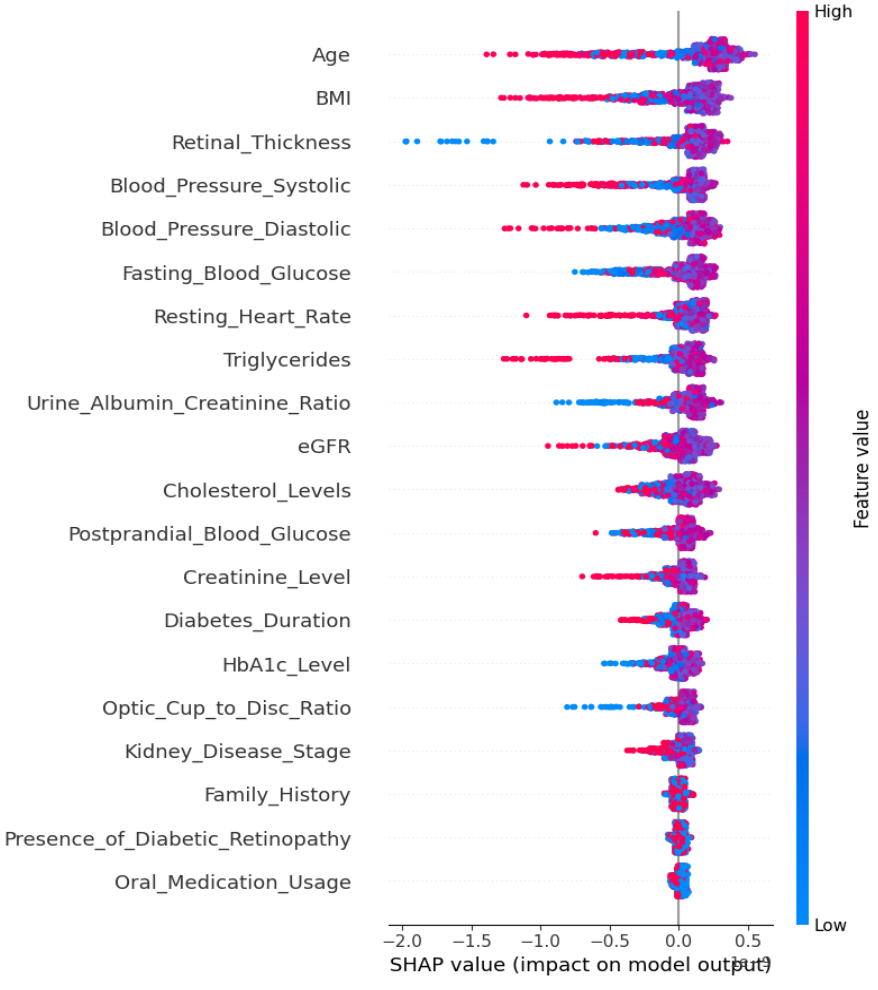

Fig. 15: Feature importance for diabetic retinopathy with the proposed system.

feature importance but also illustrates how various value ranges influence the model's output, providing clear, instance-level interpretability.

3.4 Results after feature reduction

3.4.1 For nephropathy

Fig. 16 gives the confusion matrix of diabetic nephropathy progression of the proposed system after feature reduction. The model got a complete classification with 100% accuracy and no false negatives or false positives. This indicates that the system performs very well in differentiating among the three risk levels without misclassification.

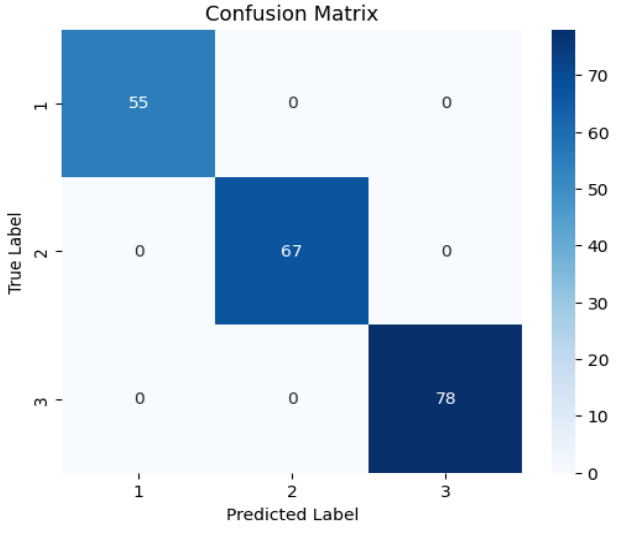

Fig. 16: Confusion matrix of diabetic nephropathy progression of the proposed system after feature reduction.

Table 2 presents the classification report of diabetic nephropathy progression for a proposed System after feature reduction. The classification report demonstrates 100% model performance in terms of accuracy, precision, recall, and F1-score for all three classes (0, 1, and 2). The model accurately classified all 200 test samples with no misclassifications. This reflects an optimal and well-generalized risk level prediction system.

Fig. 17(a) shows the obtained loss, which is 0 after a few epochs, and Fig.17(b) shows the accuracy graph of diabetic nephropathy progression of the proposed system after feature reduction, which illustrates the model's training and validation accuracy after 32 epochs. The initial accuracy is low, but quickly improves and reaches almost flawless levels (approaching 1.0) within the initial epochs. Training and validation curves then level off, suggesting stable, reliable performance. This indicates the model has learned the task quickly and is generalizing effectively. Nonetheless, a mere sharp increase in accuracy could indicate a trivial task or potential leakage of data, hence the need for further verification for reliability.

3.4.2 Comparision of results

Fig. 18 shows obtained performance metrics graph of diabetic nephropathy progression of a proposed system after feature reduction, the plot illustrates that even after feature reduction, the model's performance remains unaffected and maintain consistent high values throughout, indicating that the reduced feature set still captures the essential information required for accurate predictions and the model's ability to generalize is preserved.

Table 2: Classification report of diabetic nephropathy progression of the proposed system after feature reduction.

Classification report	Precision	Racall	F1-score	Support
0	1.00	1.00	1.00	66
1	1.00	1.00	1.00	65
2	1.00	1.00	1.00	69
Accuracy	-	-	1.00	200
Macro-avg	1.00	1.00	1.00	200
Weightes avg	1.00	1.00	1.00	200

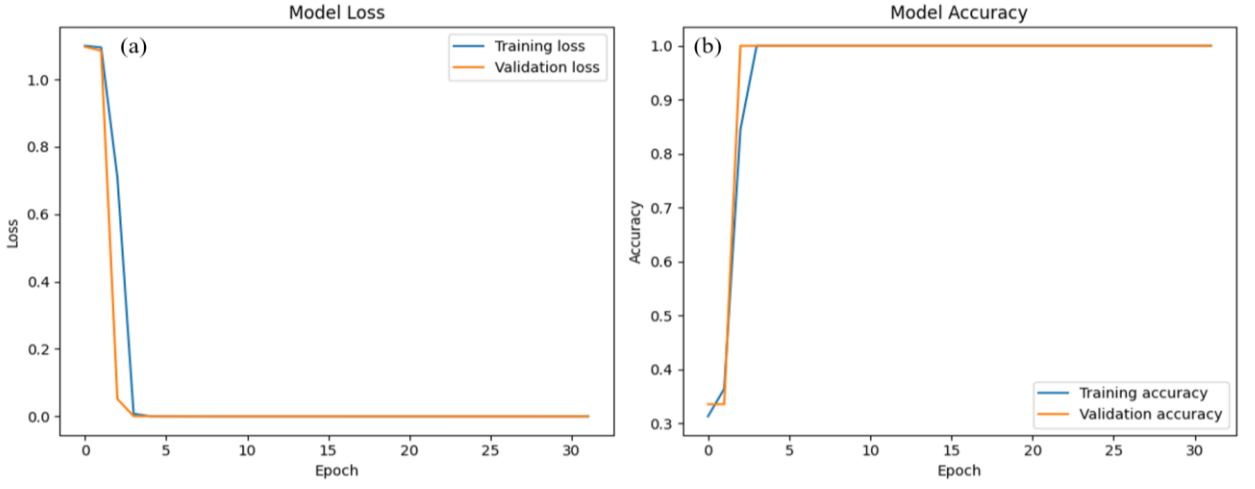

Fig. 17: The loss graph (a) and accuracy graph (b) of diabetic nephropathy progression of the proposed system after feature reduction.

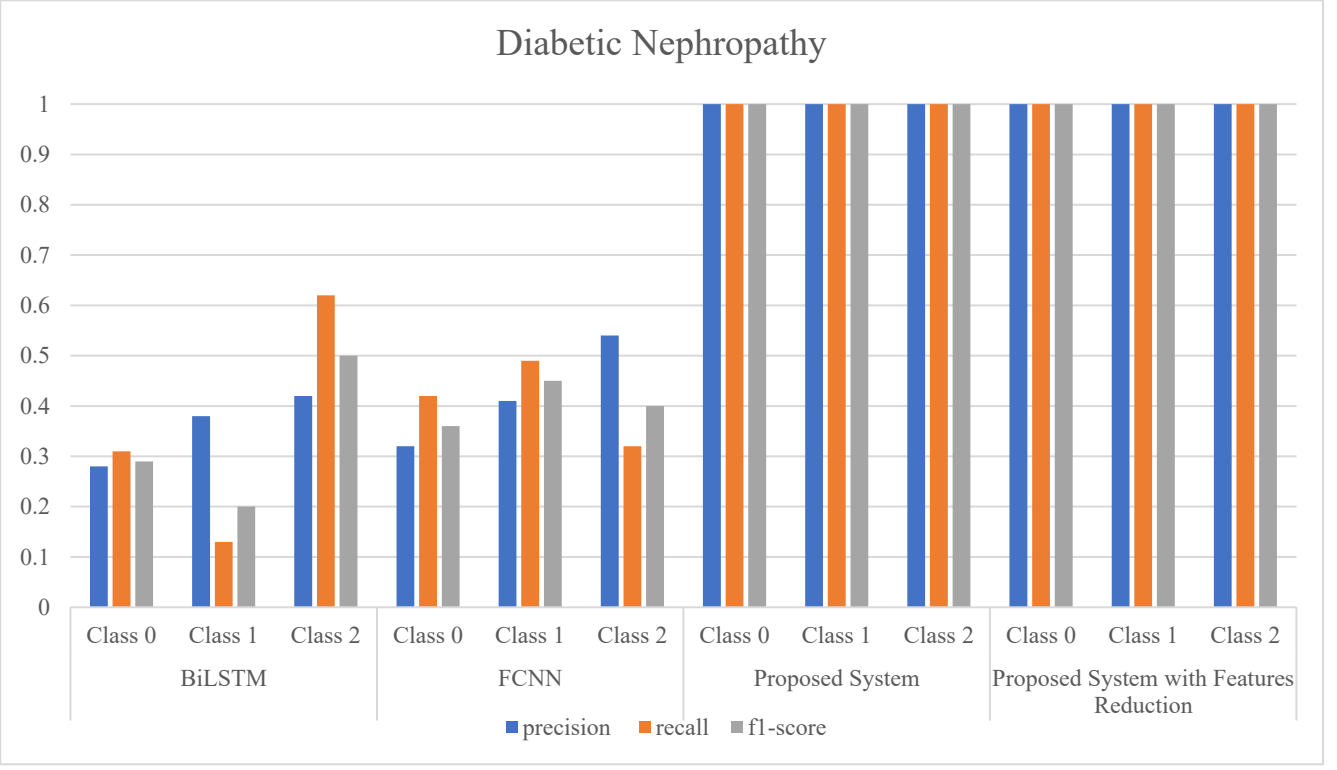

Fig. 18: Performance metrics graph of diabetic nephropathy progression of a proposed system after feature reduction

Fig. 19 shows the performance metrics graph of diabetic heart disease progression of a proposed system after feature reduction. Here, after feature reductions, the results are not hampered. The bar plot illustrates that, even in feature-reduced models, the most significant features —such as age, fasting blood glucose, BMI, and Retinal Thickness—still play a major role in the model's predictions. The same importance scores and similar performance of the models validate that feature reduction did not impair predictive power. This emphasizes the effectiveness of the feature selection process and the robustness of the proposed model in sustaining high accuracy with a reduced input set.

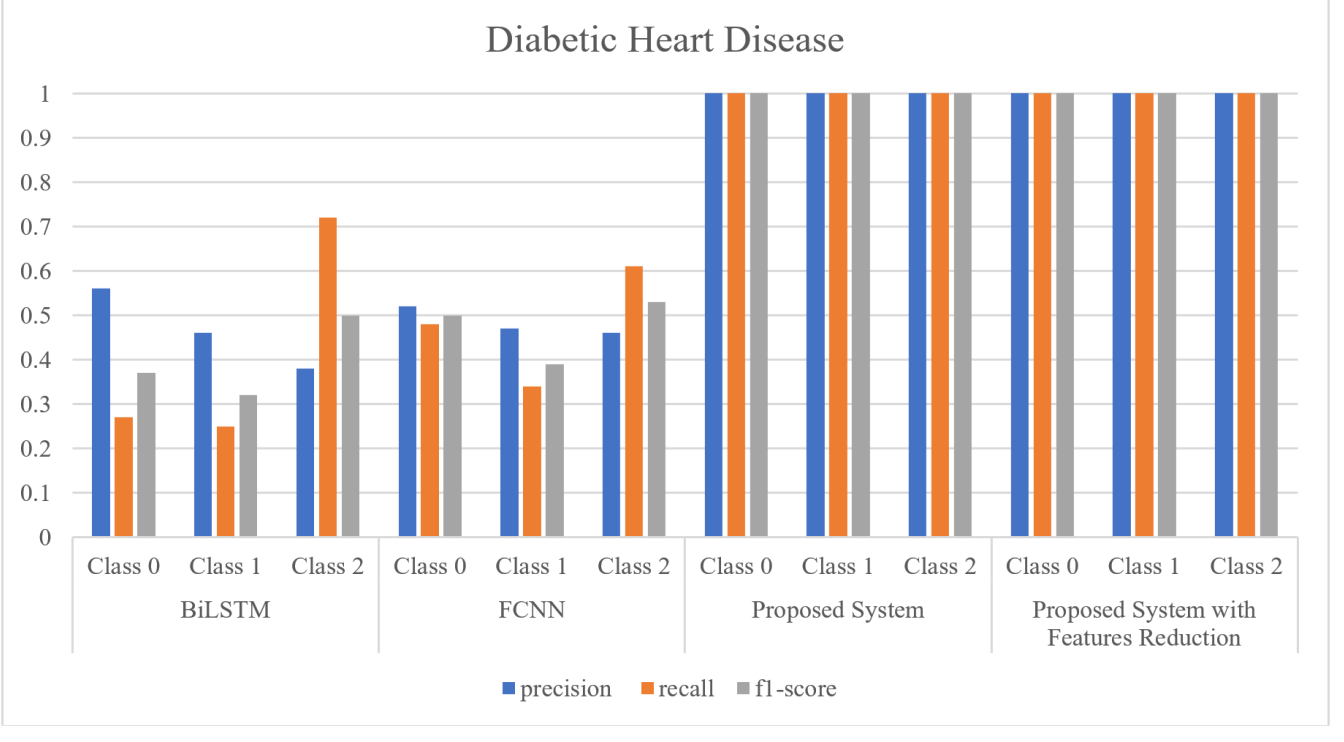

Fig. 19: Performance metrics graph of diabetic heart disease progression of the proposed system after feature reduction.

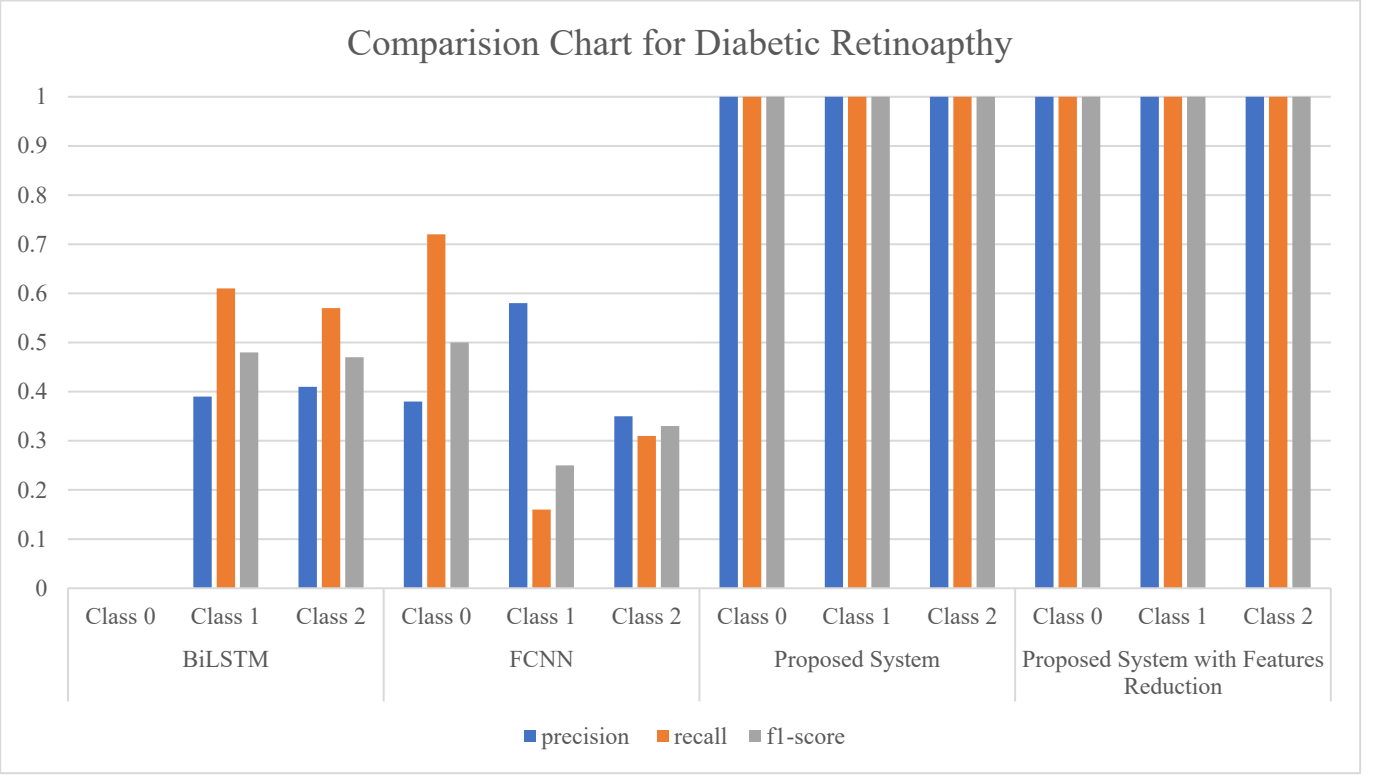

Fig. 20: Performance metrics graph of diabetic retinopathy progression of the proposed system after feature reduction.

Fig. 20 shows the performance metrics graph of diabetic retinopathy progression of a proposed system after feature reduction, in contrast to BiLSTM and FCNN architectures, which resulted in much lower accuracy and generalization. Not only does the proposed model surpass these baseline methods, but it is also effective after reducing features, ensuring that the chosen features are extremely informative and adequate to achieve trustworthy prediction. This proves the efficiency and robustness of the proposed method.

4. Conclusion

The system successfully predicts risk levels for heart disease, nephropathy, and retinopathy in patients with diabetes through a hybrid DL architecture incorporating CNN, BiLSTM, and attention. The model operates on text data of the patients, comprised of different clinical and demographic variables, to make precise predictions on risk levels. Feature selection for better interpretability and efficiency optimization of the model was performed through RF and SHAP. The findings proved that even with feature reduction, the model retained accuracy, validating that the chosen features were extremely pertinent and adequate to make precise predictions.

In addition, rigorous performance testing was performed through confusion matrices, classification reports, and accuracy measures, ensuring that the system is working well across varied risk categories. The application of XAI methods like SHAP not only increases confidence in the model's outputs but also offers greater insights into the key factors driving disease progression. Due to its high reliability and interpretability, this system is a useful tool for early risk stratification and clinical decision-making support, assisting healthcare professionals in making evidence-based decisions for diabetic patient care. The limitation of the study is that the dataset is not multimodal. The future scope of the proposed system is to test the same system for real-time data.

Acknowledgements

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

Conflict of Interest

There is no conflict of interest.

Supporting Information

Not applicable.

References

[1] Y. Fan, R. Kang, W. Huang, L. Li, Research on medical text parsing method based on BiGRU-BiLSTM multi-task learning, *Applied Sciences*, 2024, **14**, 10028, doi: 10.3390/app142110028.
[2] J. Yin, J. Jiang, L. Tong, P. Huang, FCNN: an improved fully connected neural network high-accuracy prediction model, *2023*

8th International Conference on Information Systems Engineering, June 23-25, 2023, Dalian, China, IEEE, 2023, 539-542, doi: 10.1109/ICISE60366.2023.00120.
[3] R. Iranzad, X. Liu, A review of random forest-based feature selection methods for data science education and applications, *International Journal of Data Science and Analytics*, 2024, 1-15, doi: 10.1007/s41060-024-00509-w.
[4] Y. Liu, Z. Liu, X. Luo, H. Zhao, Diagnosis of Parkinson's disease based on SHAP value feature selection, *Biocybernetics and Biomedical Engineering*, 2022, **42**, 856-869, doi: 10.1016/j.bbe.2022.06.007.
[5] J. Ma, S. An, M. Cao, L. Zhang, J. Lu, Integrated machine learning and deep learning for predicting diabetic nephropathy model construction, validation, and interpretability, *Endocrine*, 2024, **85**, 615-625, doi: 10.1007/s12020-024-03735-1.
[6] F. Mesquita, J. Bernardino, J. Henriques, J. F. Raposo, R. T. Ribeiro, S. Paredes, Machine learning techniques to predict the risk of developing diabetic nephropathy: a literature review, *Journal of Diabetes and Metabolic Disorders*, 2023, **23**, 825-839, doi: 10.1007/s40200-023-01357-4.
[7] X. Du, H. Liu, M. Huang, Y. Wu, Y. Zhang, A novel hybrid CNN-BiLSTM-attention model for diabetes prediction, *2023 IEEE 6th International Conference on Pattern Recognition and Artificial Intelligence (PRAI)*, August 18-20, 2023, Haikou, China, IEEE, 2023, 551-556, doi: 10.1109/PRAI59366.2023.10332052.
[8] V. K. Sudha, D. Kumar, Hybrid CNN and LSTM network for heart disease prediction, *SN Computer Science*, 2023, **4**, 172, doi: 10.1007/s42979-022-01598-9.
[9] M. G. El-Shafiey, A. Hagag, E. S. A. El-Dahshan, M. A. Ismail, A hybrid bidirectional LSTM and 1D CNN for heart disease prediction, *International Journal of Network Security*, 2021, **21**, 135-144, doi: 10.22937/IJCSNS.2021.21.10.18.
[10] C. Zhao, X. Huang, Y. Li, M. Yousaf Iqbal, A double-channel hybrid deep neural network based on CNN and BiLSTM for remaining useful life prediction, *Sensors*, 2020, **20**, 7109, doi: 10.3390/s20247109.
[11] N. U. Haq, T. Waheed, K. Ishaq, M. A. Hassan, N. Safie, N. F. Elias, M. Shoaib, Computationally efficient deep learning models for diabetic retinopathy detection: a systematic literature review, *Artificial Intelligence Review*, 2024, **57**, 309, doi: 10.1007/s10462-024-10942-9.
[12] M. Z. Atwany, A. H. Sahyoun, M. Yaqub, Deep learning techniques for diabetic retinopathy classification: A survey, *IEEE Access*, 2022, **10**, 28642-28655, doi: 10.1109/ACCESS.2022.3157632.
[13] P. P. Das, L. Wiese, M. Mast, J. Böhnke, A. Wulff, M. Marschollek, L. Bode, H. Rathert, T. Jack, S. Schamer, P. Beerbaum, N. Rübsamen, A. Karch, C. Groszweski-Anders, A. Haller, T. Frank, An attention-based bidirectional LSTM-CNN architecture for the early prediction of sepsis, *International Journal of Data Science and Analytics*, 2024, **568**, doi: 10.1007/s41060-024-00568-z.
[14] B. Ljubic, A. A. Hai, M. Stanojevic, W. Diaz, D. Polimac, M. Pavlovski, Z. Obradovic, Predicting complications of diabetes mellitus using advanced machine learning algorithms, *Journal of the American Medical Informatics Association*, 2020, **27**, 1343-1351, doi: 10.1093/jamia/ocaa120.
[15] G. Alfian, M. Syafrudin, N. L. Fitriyani, M. Anshari, P. Stasa, J. Svub, J. Rhee, Deep neural network for predicting diabetic retinopathy from risk factors, *Mathematics*, 2020, **8**, 1620, doi: 10.3390/math8091620.
[16] P. Vengsungnle, N. Naphon, S. Poojeera, A. Srichat, P. Naphon, Artificial neural network, experimental and numerical study on air cooling rubber mattresses for elderly and bedridden patients, *Engineered Science*, 2024, **32**, 1301, doi: 10.30919/es1301.
[17] H. A. Aliyu, I. O. Muritala, H. Bello-Salau, S. Mohammed, A. J. Onumanyi, O. O. Ajayi, Optimizing machine learning algorithms for diabetes data: a metaheuristic approach to balancing and tuning classifiers parameters, *Franklin Open*, 2024, **8**, 100153, doi: 10.1016/j.fraope.2024.100153.
[18] T. Fan, Y. Li, Emissivity prediction of multilayer film radiators by machine learning using an ultrasmall dataset, *ES Energy & Environment*, 2022, **18**, 122-130, doi: 10.30919/esee8c790.
[19] E. M. Kohner, Microvascular disease: what does the UKPDS tell us about diabetic retinopathy?, *Diabetic Medicine*, 2008, **25**, 20-24, doi: 10.1111/j.1464-5491.2008.02505.x.
[20] ADVANCE Collaborative Group, Intensive blood glucose control and vascular outcomes in patients with type 2 diabetes, *New England Journal of Medicine*, 2008, **358**, 2560-2572, doi: 10.1056/NEJMoa0802987.
[21] Diabetes Control and Complications Trial (DCCT)/Epidemiology of Diabetes Interventions and Complications (EDIC) Study Research Group, Intensive diabetes treatment and cardiovascular outcomes in type 1 diabetes: The DCCT/EDIC study 30-year follow-up, *Diabetes Care*, 2016, **39**, 686-93. doi: 10.2337/dc15-1990.
[22] W. Wang, H. Ou, M. Wen, P. Su, C. Yang, T. Kuo, M. Wang, W. Lin, Association of retinopathy severity with cardiovascular and renal outcomes in patients with type 1 diabetes: a multi-state modeling analysis, *Scientific Reports*, 2022, **12**, 4177, doi: 10.1038/s41598-022-08166-4.
[23] P. Lavanya, I. V. Subba Reddy, V. Selvakumar, Long range radio technology implementation on Internet of Things to detect particulate matter at the community level and prediction using machine learning based approach, *Engineered Science*, 2024, **29**, 1119, doi: 10.30919/es1119.
[24] S. Hochreiter, J. Schmidhuber, Long short-term memory, *Neural Computation*, 1997, **9**, 1735-1780, doi: 10.1162/neco.1997.9.8.1735
[25] A. Graves, J. Schmidhuber, Framewise phoneme classification with bidirectional LSTM and other neural network architectures, *Neural Networks*, 2005, **18**, 602-610, doi: 10.1016/j.neunet.2005.06.042.
[26] A. Krizhevsky, I. Sutskever, G. E. Hinton, ImageNet classification with deep convolutional neural networks. *Communications of the ACM*, 2017, **60**, 84-90, doi: 10.1145/3065386.

[27] V. Shatravin, D. Shashev, S. Shidlovskiy, Implementation of the SoftMax activation for reconfigurable neural network hardware accelerators, *Applied Sciences*, 2023, **13**, 12784, doi: 10.3390/app132312784.