Cardiovascular Disease Prediction with Explainable AI

Hailemicael Lulseged Yimer

October 13, 2024

Contents

1 Project Objective

This project aims to develop a model to predict cardiovascular disease (CVD) and integrate explainable AI techniques (SHAP and LIME) to interpret the model's predictions. The goal is to verify if the features identified in the exploratory analysis align with the model's outputs.

2 Background

Cardiovascular disease (CVD) is one of the leading causes of death worldwide. Identifying individuals at risk of developing CVD early is crucial for preventing severe outcomes. Several risk factors, such as age, blood pressure, cholesterol levels, and lifestyle choices like smoking, are known to contribute to the likelihood of CVD.

In this project, we leverage advanced machine learning techniques to build a predictive model for cardiovascular disease. The dataset includes key clinical features, such as age, creatinine phosphokinase levels, ejection fraction, serum creatinine, serum sodium, and more. These features provide a comprehensive view of the patient's health status and are critical for the model to make accurate predictions.

To enhance the transparency of the model's predictions, we incorporate explainable AI techniques such as SHAP and LIME. This allows us to understand which features most significantly impact the model's decisions. The goal is to ensure that the machine learning model's outputs align with known medical risk factors, ensuring the results are both accurate and interpretable.

3 Exploratory Data Analysis (EDA)

In this section, an exploratory data analysis (EDA) was performed to gain a deeper understanding of the dataset and uncover patterns useful for model development. This process involved visualizing the relationships between features, analyzing distributions, and identifying potential outliers and correlations. Below are the key insights obtained from the EDA.

3.1 Descriptive Statistics

The dataset includes 299 samples with 13 features, covering clinical and demographic factors such as age, creatinine phosphokinase, ejection fraction, serum creatinine, serum sodium, and the target variable `DEATH_EVENT`. A descriptive statistics summary shows the range and spread of the data, highlighting important factors like:

- **Average Age**: The average age of patients is 60.83 years, with a standard deviation of 11.89 years, indicating a wide age range.
- **Creatinine Phosphokinase Levels**: These levels vary significantly, with extreme outliers reaching values as high as 7861, which could indicate muscle damage.
- **Serum Creatinine and Ejection Fraction**: Both critical indicators of cardiovascular health display significant variation across the dataset.

3.2 Correlation Analysis

A correlation heatmap was generated to examine the relationships between the continuous variables in the dataset. This analysis helps identify highly correlated features, which could influence the model's predictions. Key findings include:

- **Serum Creatinine and Age**: A moderate positive correlation ($r = 0.16$) suggests a potential relationship between age and kidney function.
- **Sex and Smoking**: A notable correlation ($r = 0.45$) indicates that male patients are more likely to be smokers, potentially affecting heart disease outcomes.
- **Serum Sodium and Ejection Fraction**: A weak negative correlation ($r = -0.18$) indicates possible electrolyte imbalances in patients with poor heart function.

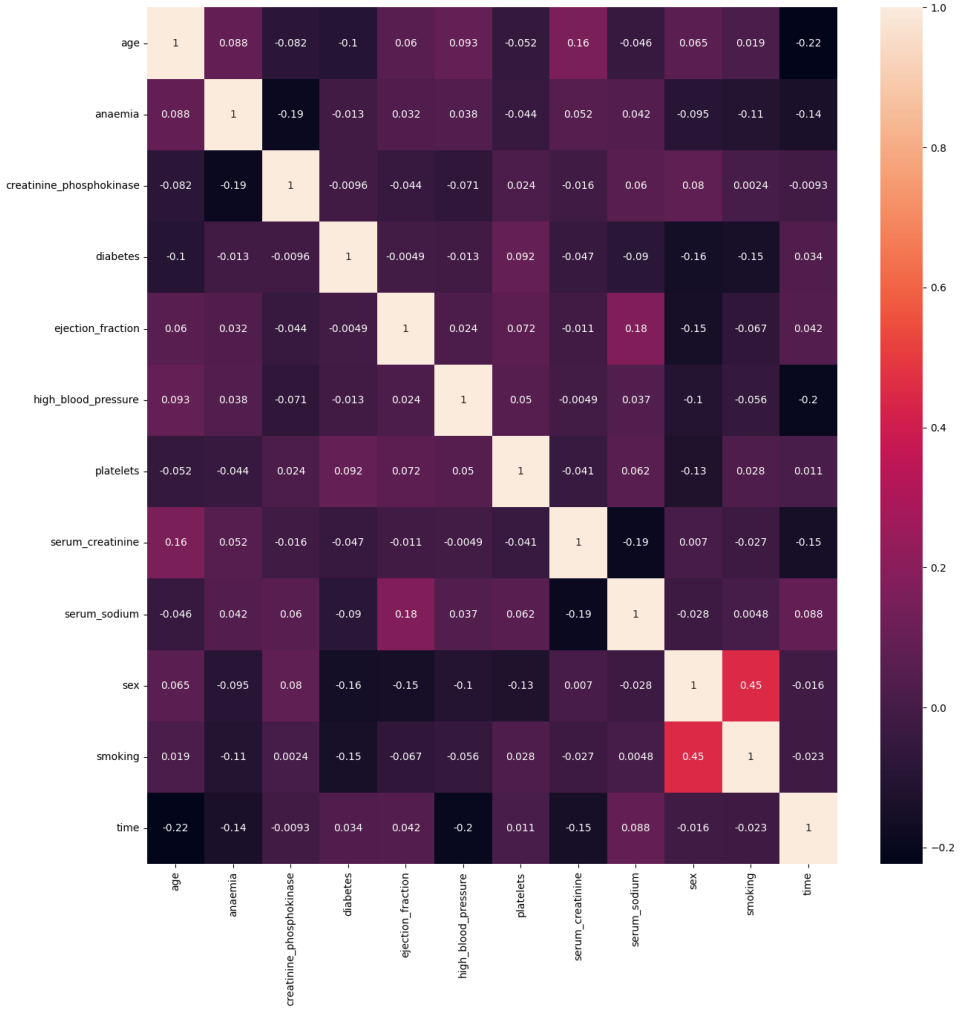

Figure 1: Correlation Heatmap of Continuous Features

3.3 Pairwise Feature Relationships

A pairplot was created to visualize the relationships between key continuous features and the target variable, `DEATH_EVENT`. This plot helps identify patterns, distributions, and

potential clusters that differentiate patients who experienced a death event from those who did not.

- **Age Distribution**: Higher concentration of death events among older patients.
- **Ejection Fraction**: Clear differences between patients who experienced heart failure and those who did not.
- **Serum Creatinine Levels**: Notably higher among patients who experienced a death event, suggesting a link between renal function and mortality risk.

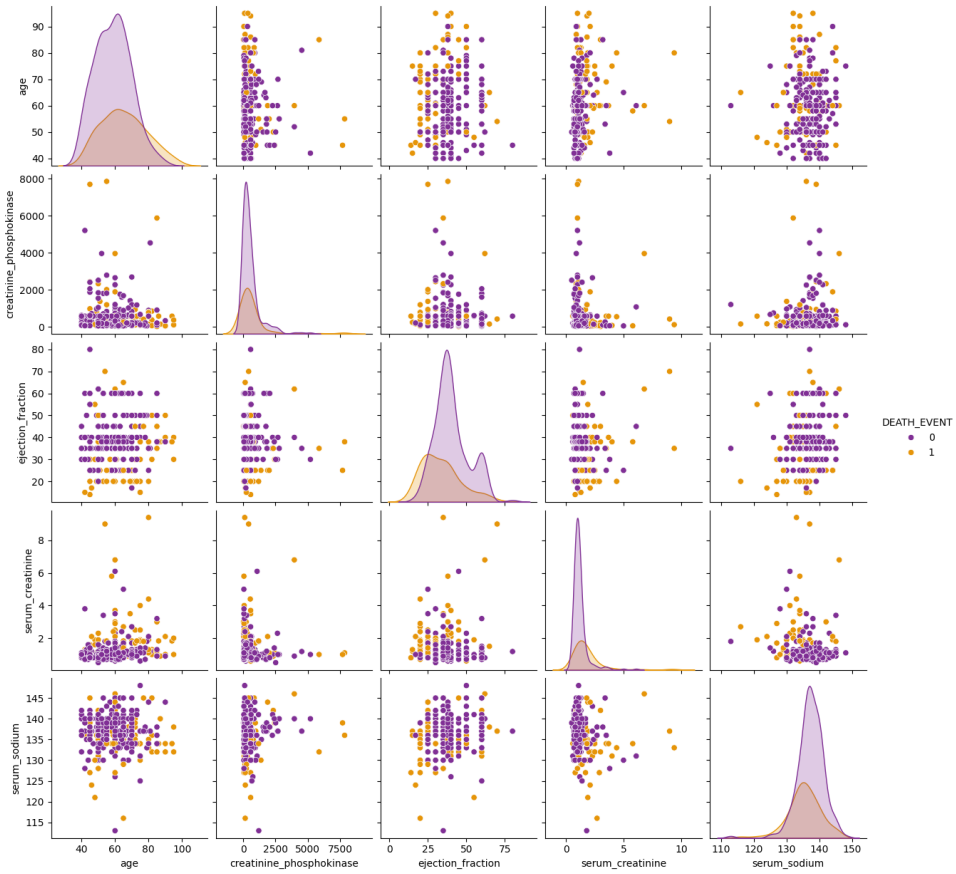

Figure 2: Pairplot Visualizing Feature Relationships with `DEATH_EVENT`

3.4 Feature Distribution and Outliers

To further explore the distribution of continuous features and identify potential outliers, boxenplots were generated. These plots examine how features such as age, ejection fraction, serum creatinine, and serum sodium vary between patients who experienced a death event and those who did not.

- **Age**: Patients who experienced a death event tend to be older, with most events occurring in patients above 60 years.
- **Ejection Fraction**: Lower ejection fractions are associated with higher mortality, confirming its significance in cardiovascular health.

- **Serum Creatinine**: Higher levels observed in patients who experienced a death event, highlighting the role of kidney function in heart failure outcomes.
- **Creatinine Phosphokinase**: Significant variability with extreme outliers present, which may require special handling before model development.

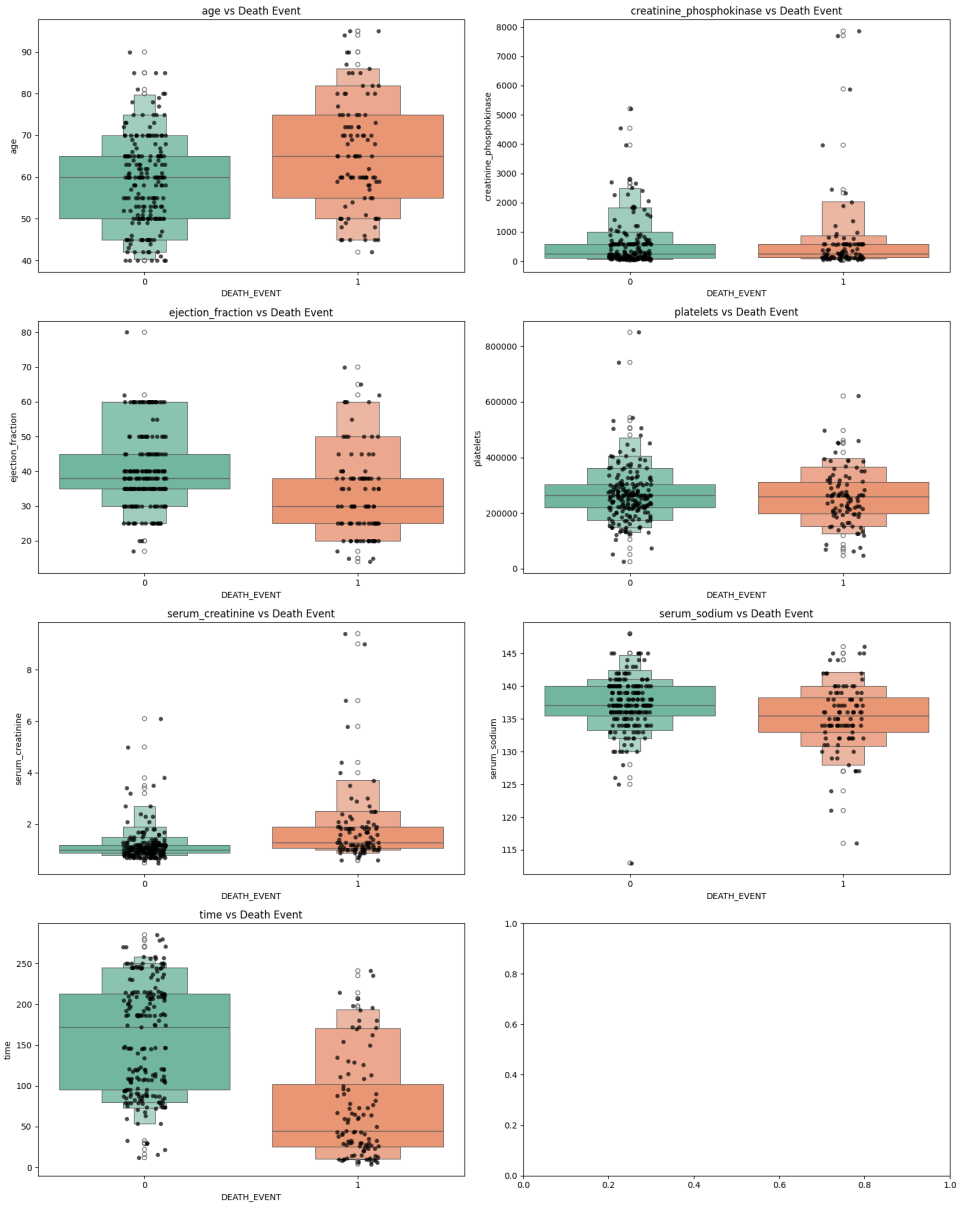

Figure 3: Boxenplots of Continuous Features vs `DEATH_EVENT`

3.5 Normality Testing with Q-Q Plots

Quantile-Quantile (Q-Q) plots were generated for key continuous features to assess whether the data follows a normal distribution. Deviations from normality were observed in several features, particularly those with outliers, such as creatinine phosphokinase and serum creatinine. Understanding the normality of the data is important for selecting appropriate machine learning models and determining if any feature transformations are needed.

- **Age**: Follows a reasonably normal distribution, although slight deviations are present at both extremes.
- **Creatinine Phosphokinase and Serum Creatinine**: Exhibit significant deviations from normality due to extreme values, suggesting that these features may require transformation.
- **Serum Sodium and Ejection Fraction**: Show slight deviations from normality but remain relatively well-distributed.

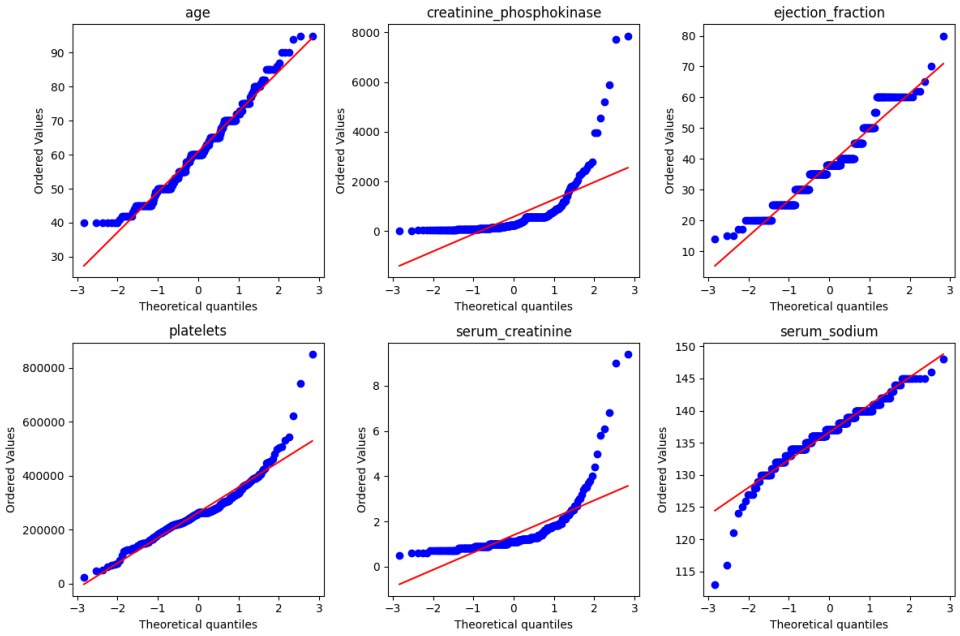

Figure 4: Q-Q Plots for Continuous Features

3.6 Average Feature Comparison: Heart Failure vs Non-Heart Failure

To better understand the differences between patients who experienced heart failure and those who did not, the average values of key continuous features were computed and visualized across these two groups. Notable differences include:

- **Age**: Patients with heart failure tend to be older on average.
- **Ejection Fraction**: Lower ejection fractions observed in the heart failure group, indicating poorer heart function.
- **Serum Creatinine**: Higher average serum creatinine levels associated with heart failure, suggesting kidney dysfunction as a possible contributing factor.

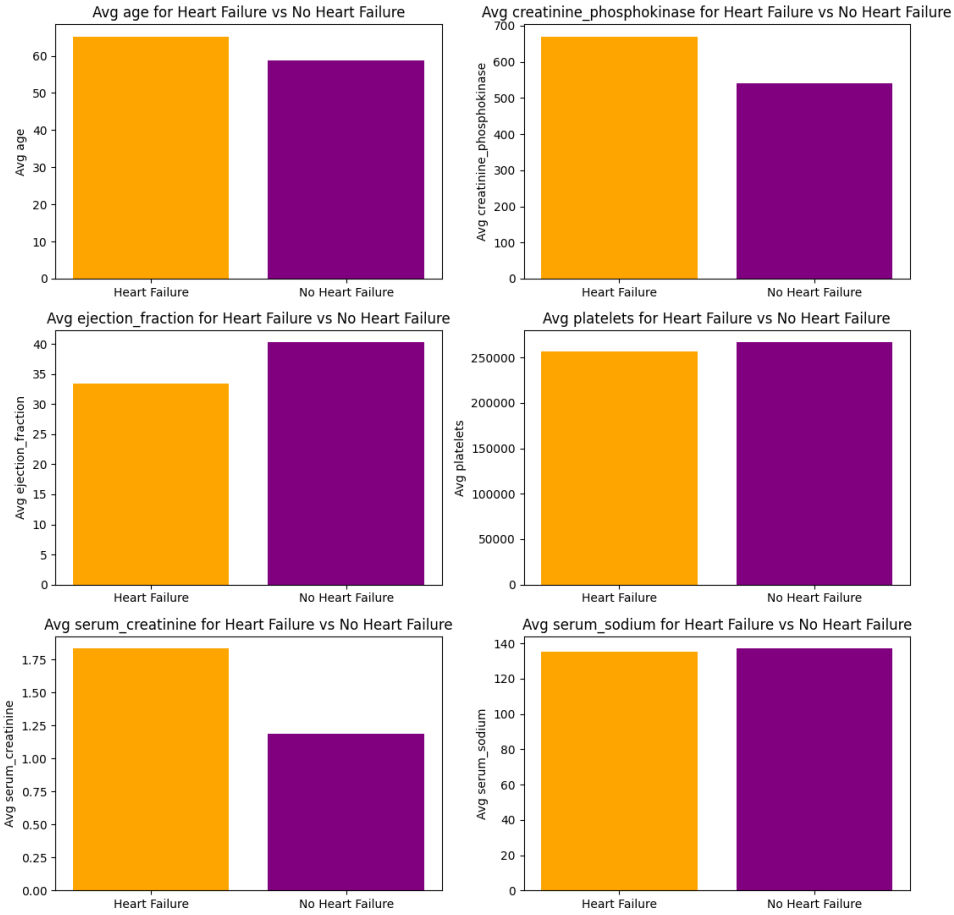

Figure 5: Average Feature Comparison for Heart Failure vs Non-Heart Failure

3.7 Feature Normalization

Since the dataset contains continuous features that vary significantly in scale, normalization was applied to ensure that all features contribute equally to the model. Continuous variables such as age, creatinine phosphokinase, ejection fraction, platelets, serum creatinine, serum sodium, and time were standardized using a standard scaler.

Normalization ensures that each feature has a mean of 0 and a standard deviation of 1. This is particularly important when features are measured in different units (e.g., age in years, creatinine phosphokinase in mcg/L), as it prevents certain features from dominating the model due to their larger scale. This step helps the model converge faster and perform better during training.

3.8 Class Imbalance Handling with SMOTE

The target variable, `DEATH_EVENT`, was imbalanced, with 203 patients not experiencing heart failure (class 0) and only 96 patients experiencing heart failure (class 1). Such an imbalance can bias machine learning models to favor the majority class, leading to poor performance in detecting heart failure cases.

To address this, the Synthetic Minority Over-sampling Technique (SMOTE) was applied. SMOTE generates synthetic examples for the minority class, balancing the dataset. After applying SMOTE, the dataset had 203 instances of heart failure and 203 instances

of non-heart failure, providing a balanced dataset for training.

3.9 Dataset Splitting

After balancing the classes, the dataset was split into training, validation, and test sets:

- **Training Set (70%)**: Used for model training.
- **Validation Set (15%)**: Used for hyperparameter tuning and to prevent overfitting.
- **Test Set (15%)**: Used to evaluate the final model's performance.

The balanced dataset and the clear separation of training, validation, and test sets ensure that the machine learning model will be trained and evaluated fairly, with the potential to generalize well to new data.

3.10 Conclusion of EDA

In conclusion, the exploratory data analysis revealed several key patterns in the data. Age, ejection fraction, serum creatinine, and creatinine phosphokinase were identified as important clinical indicators associated with heart failure outcomes. The normalization of continuous variables and the application of SMOTE to address class imbalance were essential steps in preparing the data for machine learning model development. These steps ensure that the model will be trained on a balanced and standardized dataset, improving its ability to make accurate predictions.

4 Model Development and Evaluation

Several machine learning models were developed and evaluated to predict cardiovascular disease (CVD). These models include Logistic Regression, Support Vector Machines (SVM), Random Forest, Fully Connected Neural Network (FCNN), and Convolutional Neural Network (CNN). The performance of each model was evaluated using key metrics such as accuracy, precision, recall, and F1 score. Below are the details of each model, their evaluation results, and a comparison of their performance.

4.1 Logistic Regression

Logistic regression is commonly used for binary classification problems and is well-suited for predicting heart failure outcomes.

4.1.1 Training the Model

The logistic regression model was trained using the balanced dataset. The model was fitted to the training data using 1000 maximum iterations to ensure convergence.

4.1.2 Evaluation Metrics

The logistic regression model achieved the following performance metrics:

- **Accuracy**: 0.89
- **Precision**: 0.89
- **Recall**: 0.86
- **F1 Score**: 0.87

These results show that the logistic regression model performed well with a good balance between precision and recall.

4.1.3 Confusion Matrix

Table 1: Confusion Matrix for Logistic Regression

	Predicted No CVD	**Predicted CVD**
Actual No CVD	30	3
Actual CVD	4	24

The confusion matrix for the logistic regression model shows that the model correctly predicted 30 cases of no CVD and 24 cases of CVD, with 3 false positives and 4 false negatives. The false negatives suggest that while the model was overall effective, some actual CVD cases were missed.

4.2 Support Vector Machines (SVM)

Two variations of Support Vector Machines (SVM) were implemented: one using a linear kernel and the other using an RBF (Radial Basis Function) kernel. SVMs are powerful algorithms for classification tasks, capable of finding the optimal hyperplane separating classes.

4.2.1 SVM (Linear Kernel)

Evaluation Metrics:

- **Accuracy**: 0.85
- **Precision**: 0.85
- **Recall**: 0.82
- **F1 Score**: 0.83

Confusion Matrix: The SVM with a linear kernel correctly predicted 29 no-CVD cases and 23 CVD cases, with 4 false positives and 5 false negatives.

4.2.2 SVM (RBF Kernel)

Evaluation Metrics:

- **Accuracy**: 0.82
- **Precision**: 0.83
- **Recall**: 0.81
- **F1 Score**: 0.82

Confusion Matrix: The RBF kernel model correctly predicted 23 CVD cases but misclassified 5, with 4 false positives.

4.3 Random Forest Classifier

The Random Forest Classifier, an ensemble learning method that aggregates multiple decision trees, was implemented to enhance prediction accuracy.

4.3.1 Training and Hyperparameters

The Random Forest model was trained using 100 decision trees with default parameters. This ensemble method is particularly effective at reducing overfitting by averaging multiple decision trees.

4.3.2 Evaluation Metrics

The Random Forest model outperformed the other models with the following metrics:

- **Accuracy**: 0.95
- **Precision**: 0.93
- **Recall**: 0.96
- **F1 Score**: 0.95

4.3.3 Confusion Matrix

Table 2: Confusion Matrix for Random Forest

	Predicted No CVD	**Predicted CVD**
Actual No CVD	31	2
Actual CVD	1	27

The Random Forest model correctly predicted 31 no-CVD cases and 27 CVD cases, with only 1 false negative and 2 false positives. This performance demonstrates the strength of ensemble methods, which can capture complex patterns in the data while maintaining robustness.

4.4 Fully Connected Neural Network (FCNN)

A Fully Connected Neural Network (FCNN) was developed to explore the potential of deep learning models for this task.

4.4.1 Model Architecture

The FCNN consisted of the following layers:

- **Input Layer**: Matching the number of features in the dataset.
- **Dense Layers**: Two dense layers with 64 and 32 units, using ReLU activation.
- **Dropout Layers**: Dropout layers with a rate of 0.3 to prevent overfitting.
- **Output Layer**: 1 unit with a sigmoid activation function for binary classification.

4.4.2 Hyperparameter Tuning

A grid search was performed over the following hyperparameters:

- **Learning Rate**: 0.01, 0.001, 0.0005, and 0.0001.
- **Dropout Rate**: 0.3, 0.5, and 0.7.
- **Batch Size**: 32 and 64.
- **Epochs**: 50 and 100.

The best model achieved an accuracy of 0.98 with the following parameters:

- **Learning Rate**: 0.001
- **Dropout Rate**: 0.3
- **Batch Size**: 32
- **Epochs**: 100

4.4.3 Evaluation Metrics

The FCNN model performed very well, with the following metrics:

- **Accuracy**: 0.98
- **Precision**: 0.93
- **Recall**: 0.96
- **F1 Score**: 0.95

4.5 Convolutional Neural Network (CNN)

A Convolutional Neural Network (CNN) was developed to capture complex, local patterns in data through convolutional layers.

4.5.1 Model Architecture

The CNN architecture consisted of:

- **Convolutional Layers**: Three 1D convolutional layers with 32, 64, and 128 filters, each using LeakyReLU activation and followed by batch normalization.
- **Pooling Layer**: A max pooling layer with pool size 3 and stride 2 to reduce spatial dimensions.
- **Dropout Layers**: Dropout layers to prevent overfitting.
- **Fully Connected Layers**: Two fully connected layers with 256 and 512 units, using LeakyReLU activation.
- **Output Layer**: 1 unit with a sigmoid activation for binary classification.

4.5.2 Training and Hyperparameters

The CNN was trained on the reshaped dataset, using 50 epochs and a batch size of 10. This model leveraged convolutional layers to capture relationships between the features.

4.5.3 Evaluation Metrics

The CNN achieved the following results:

- **Accuracy**: 0.87
- **Precision**: 0.79
- **Recall**: 0.96
- **F1 Score**: 0.87

4.5.4 Confusion Matrix

Table 3: Confusion Matrix for CNN

	Predicted No CVD	**Predicted CVD**
Actual No CVD	26	7
Actual CVD	1	27

The CNN correctly predicted 26 no-CVD cases and 27 CVD cases, with 7 false positives and 1 false negative. The model demonstrated a high recall, making it suitable for applications where minimizing false negatives is important.

4.6 Comparison of Models

The following table summarizes the performance of the machine learning models developed for cardiovascular disease prediction. The models were evaluated based on key metrics such as accuracy, precision, recall, and F1 score to provide a comprehensive view of their strengths and weaknesses.

Table 4: Comparison of Model Performance

Model	Accuracy	Precision	Recall	F1 Score
Logistic Regression	0.89	0.89	0.86	0.87
SVM (Linear Kernel)	0.85	0.85	0.82	0.83
SVM (RBF Kernel)	0.82	0.83	0.81	0.82
Random Forest	0.95	0.93	0.96	0.95
Fully Connected Neural Network (FCNN)	0.98	0.93	0.96	0.95
Convolutional Neural Network (CNN)	0.87	0.79	0.96	0.87

4.6.1 Summary of Model Performance

- The **Fully Connected Neural Network (FCNN)** achieved the highest overall performance, with an accuracy of 0.98 and an F1 score of 0.95, indicating its strong capability to balance precision and recall, making it highly suitable for predicting cardiovascular disease (CVD) outcomes. - The **Random Forest** model also demonstrated excellent performance, particularly excelling in terms of **recall** (0.96), which makes it effective in minimizing false negatives and detecting patients at risk of CVD. - The **Logistic Regression** and **SVM** models, although performing decently, lagged behind in terms of accuracy and recall compared to the FCNN and Random Forest models. These models still offer a good balance between precision and recall, with F1 scores in the range of 0.82–0.87. - The **Convolutional Neural Network (CNN)** performed well with a high recall (0.96), suggesting it is useful for identifying true positive CVD cases. However, its precision (0.79) and overall accuracy (0.87) were lower than other models, indicating a higher likelihood of false positives.

In summary, the **FCNN** and **Random Forest** models proved to be the most effective for this task, with the FCNN model achieving the best balance across all metrics. The CNN model showed high potential for identifying CVD cases due to its high recall, but its lower precision might limit its application in scenarios where false positives must be minimized. rengths for different scenarios.

5 Explainable AI

In addition to building predictive models for cardiovascular disease (CVD), Explainable AI (XAI) techniques were incorporated to interpret the models and understand which features contribute the most to the predictions. Two popular XAI methods were used: SHAP (SHapley Additive exPlanations) and LIME (Local Interpretable Model-agnostic Explanations). These techniques provide insights into the behavior of machine learning models, allowing for a deeper understanding of how individual features affect predictions.

5.1 SHAP for Logistic Regression

To understand the predictions made by the Logistic Regression model, SHAP KernelExplainer was used to calculate SHAP values for each feature in the dataset. SHAP values provide a measure of the impact that each feature has on the model's output, allowing for detailed interpretability.

Key Insights from the SHAP Summary Plot:

- **Time**: Had the largest impact on the model's predictions, with both positive and negative SHAP values, suggesting that the time variable is a crucial determinant in identifying patients at risk of cardiovascular disease.
- **Ejection Fraction**: Low ejection fraction values are strongly associated with cardiovascular disease, as shown by high SHAP values.
- **Age**: Older age contributes positively to the prediction of disease, meaning that older patients are more likely to be classified as having CVD.
- **Serum Creatinine and Serum Sodium**: These two features also have a notable influence on the model's predictions, suggesting that kidney function and electrolyte balance are significant factors in determining cardiovascular risk.

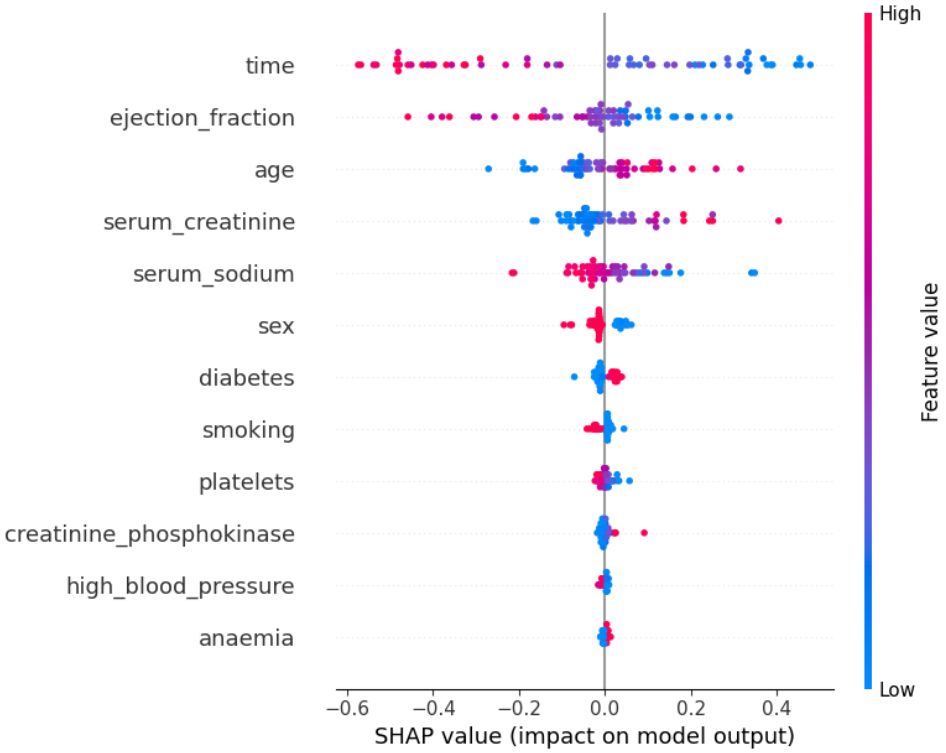

Figure 6: SHAP Summary Plot for Logistic Regression

5.2 SHAP for SVM Models

To gain further insights into the predictions of the Support Vector Machine (SVM) models, SHAP KernelExplainer was used to generate SHAP values for both the SVM with a linear kernel and the SVM with an RBF kernel. This allowed for a detailed interpretation of how features influenced the predictions made by these models.

5.2.1 SHAP for SVM (Linear Kernel)

Summary of SHAP Results for the Linear SVM Model:

- **Time**: Similar to the Logistic Regression model, 'time' had the largest impact on the model's predictions. High 'time' values contributed to predictions of no CVD, while lower values pushed predictions towards CVD.
- **Age**: Older age emerged as an important predictor, with high SHAP values driving predictions towards CVD.
- **Ejection Fraction and Serum Creatinine**: These features showed strong associations with CVD predictions. A lower ejection fraction and higher serum creatinine levels pushed the model towards predicting CVD.

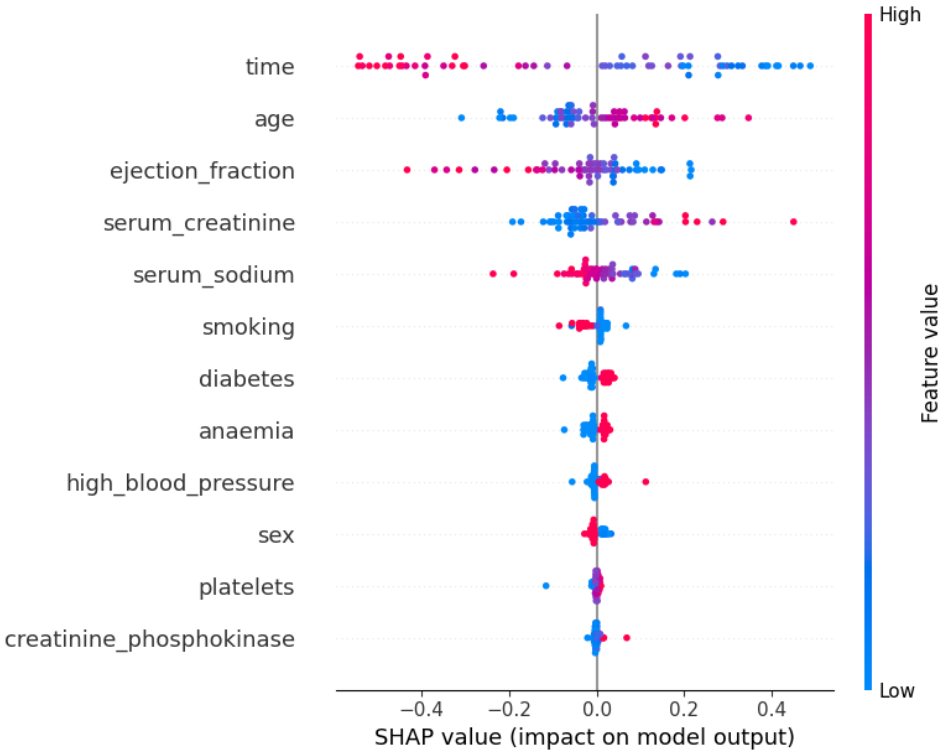

Figure 7: SHAP Summary Plot for SVM (Linear Kernel)

5.2.2 SHAP for SVM (RBF Kernel)

Summary of SHAP Results for the RBF SVM Model:

- **Time and Ejection Fraction**: These were the most important features driving the predictions. Lower ejection fraction values pushed the predictions toward CVD, while higher 'time' values indicated no CVD.
- **Serum Creatinine and Age**: These features continued to play a significant role, with high serum creatinine and older age contributing positively to the prediction of CVD.

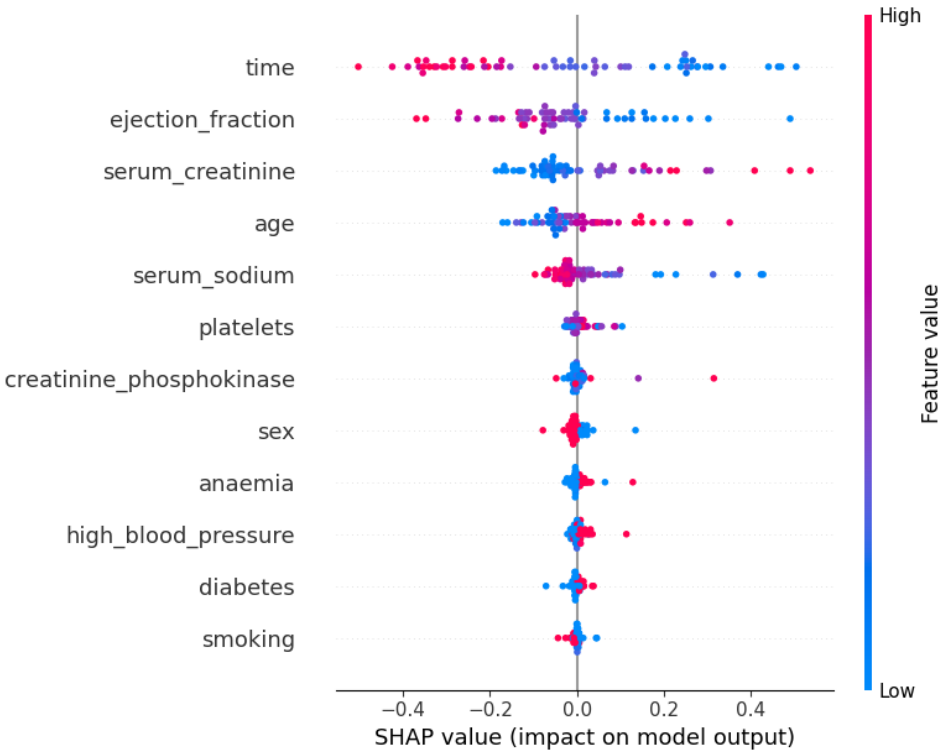

Figure 8: SHAP Summary Plot for SVM (RBF Kernel)

5.3 SHAP for Random Forest Classifier

For the Random Forest Classifier, SHAP KernelExplainer was applied to gain interpretability into the model's predictions. SHAP values were calculated for the test set, and the following insights were gathered from the SHAP summary plot.

5.3.1 SHAP Summary Plot for Random Forest

The SHAP summary plot provides an overall view of the features that contributed most to the Random Forest model's predictions. Some key insights include:

- **Time**: Similar to other models, 'time' was the most significant feature affecting the predictions. High values of 'time' pushed the model towards predicting no CVD, while lower values contributed to predictions of CVD.
- **Serum Creatinine and Ejection Fraction**: High serum creatinine levels and low ejection fraction were strongly indicative of CVD, aligning with the medical understanding of kidney function and heart health as important risk factors.
- **Serum Sodium and Age**: These features also played a key role in predictions. Low serum sodium and older age were associated with a higher likelihood of CVD.

5.4 SHAP for Fully Connected Neural Network (FCNN)

To interpret the FCNN model, the SHAP unified explainer was used to generate SHAP values for the test set. The SHAP values provided valuable insights into how each feature contributed to the model's predictions.

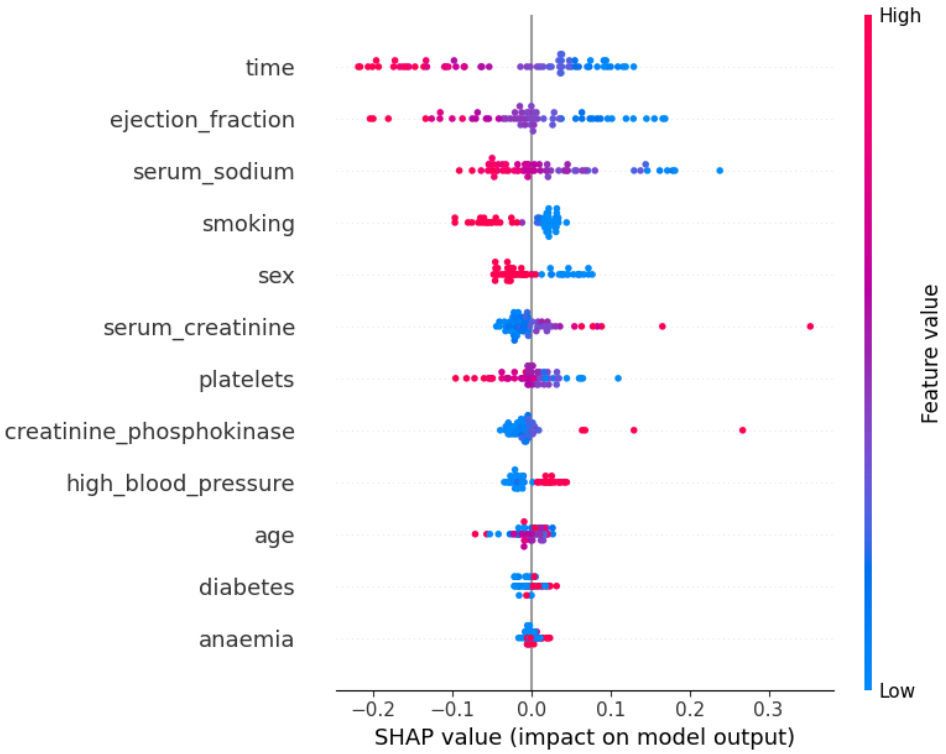

Figure 9: SHAP Summary Plot for FCNN

5.4.1 SHAP Summary Plot for FCNN

The SHAP summary plot for the FCNN model highlights the following key features:

- **Time**: Had the largest impact on the model's predictions, consistently pushing predictions towards or away from CVD depending on its value.
- **Ejection Fraction and Serum Sodium**: Lower values were strongly associated with CVD predictions, similar to other models.
- **Smoking and Serum Creatinine**: Smoking status and high serum creatinine levels significantly influenced the model's predictions. Smokers and individuals with high serum creatinine were more likely to be classified as having CVD.
- **Sex and Platelets**: These features had a moderate influence on the predictions, but their impact was generally less significant compared to the top features.

5.5 LIME for Logistic Regression

While SHAP provides a global explanation for the model's behavior, LIME (Local Interpretable Model-agnostic Explanations) helps explain individual predictions by creating locally interpretable surrogate models. LIME was used to analyze specific test set samples and understand how the model arrived at particular predictions.

5.5.1 LIME Explanation for an Individual Sample (Logistic Regression)

Using LIME, explanations for individual predictions made by the Logistic Regression model were generated. The LIME explainer provided a set of feature contributions for

each sample, showing how specific features influenced the model's decision to classify the patient as having cardiovascular disease (CVD) or not.

For one particular sample from the test set, LIME produced the following results:

- **Time**: Had the largest contribution, with a strong negative value pushing the prediction toward CVD.
- **Serum Creatinine**: High levels of serum creatinine (7.75) were a main contributor to predicting CVD, as impaired kidney function is a key indicator of cardiovascular risk.
- **Age**: Contributed positively to the prediction of CVD, as older age is associated with higher risk.
- **Serum Sodium and Ejection Fraction**: Both features, with lower values, pushed the prediction towards CVD, confirming their role in predicting cardiovascular health.

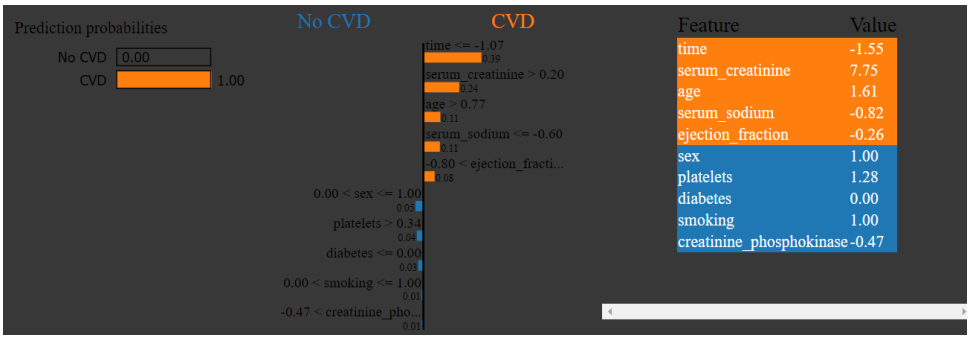

Figure 10: LIME Explanation for an Individual Sample (Logistic Regression)

5.6 LIME for Fully Connected Neural Network (FCNN)

LIME was also used to explain individual predictions made by the FCNN model. By analyzing specific samples, a deeper understanding of how the model arrived at its predictions was achieved.

5.6.1 LIME Explanation for an Individual Sample (FCNN)

For a specific sample from the test set, LIME provided the following insights:

- **Time**: Was the most significant feature pushing the prediction towards CVD.
- **Serum Sodium and Serum Creatinine**: Low levels of serum sodium and high serum creatinine were key factors influencing the model to predict CVD for this patient.
- **Ejection Fraction and Smoking**: Low ejection fraction and the patient's smoking status contributed to the model's prediction.
- **High Blood Pressure and Platelets**: These features had a moderate influence on the prediction, further pushing it towards CVD.

The LIME explanation for the FCNN model confirmed the importance of the core clinical features (time, serum sodium, serum creatinine, ejection fraction) in determining cardiovascular disease risk.

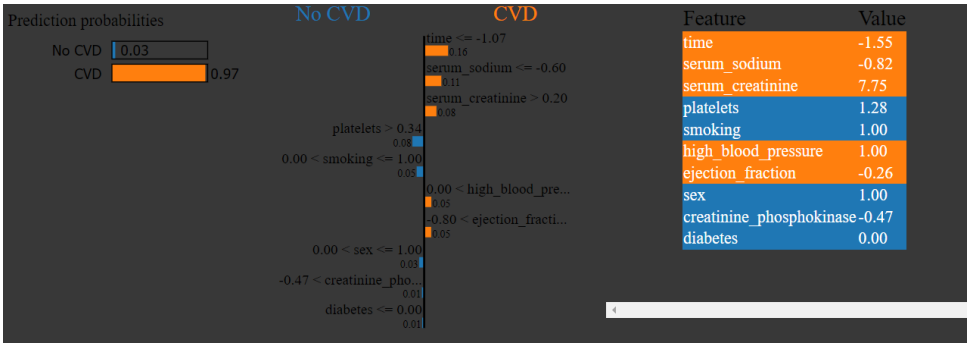

Figure 11: Lime for FCNN

5.7 Conclusion of Explainable AI with SHAP and LIME

By applying both SHAP and LIME, comprehensive explanations for the models' predictions were obtained. SHAP provided a global understanding of feature importance across all predictions, while LIME allowed for detailed local explanations of individual predictions. Both methods confirmed that clinically important features, such as ejection fraction, serum creatinine, and age, play a central role in predicting cardiovascular disease. These XAI techniques add transparency to the machine learning models, making them more interpretable and trustworthy for potential use in clinical settings.

6 Comparison and Validation

The SHAP results were compared with the findings from the exploratory analysis to validate the consistency and relevance of the features identified by both methods. The goal was to assess whether the features that were significant in the exploratory data analysis (EDA) aligned with those identified as important by the machine learning models.

6.1 Exploratory Analysis Findings

In the exploratory analysis, several features emerged as important predictors of cardiovascular disease (CVD). Specifically:

- **Ejection Fraction**: Lower ejection fraction values were associated with higher rates of CVD.
- **Serum Creatinine and Serum Sodium**: Imbalances in kidney function, as indicated by elevated serum creatinine and lower serum sodium levels, were strongly linked to CVD.
- **Age**: Older patients were at a higher risk of CVD, with age showing a clear correlation with heart disease.
- **Time**: Time, likely representing patient follow-up periods, appeared to be a critical factor in predicting outcomes.
- **Smoking and High Blood Pressure**: Smoking status and high blood pressure were additional risk factors identified during the exploratory analysis.

6.2 Comparison with SHAP Results

The SHAP analyses for the Logistic Regression, SVM, Random Forest, and Fully Connected Neural Network (FCNN) models consistently identified the same core features as crucial for predicting CVD:

- **Time**: Across all models, 'time' was the most important feature, closely aligning with the exploratory findings. Lower values for time were typically associated with higher CVD risk, reinforcing the importance of this factor.
- **Ejection Fraction**: The SHAP results consistently highlighted low ejection fraction values as a strong predictor of CVD, supporting the insights from the EDA.
- **Serum Creatinine and Serum Sodium**: Both SHAP and the exploratory analysis agreed that these features, particularly high serum creatinine and low serum sodium, are critical indicators of CVD risk.
- **Age**: SHAP analyses confirmed that older age is associated with higher CVD risk, in line with the findings from the EDA.
- **Smoking and High Blood Pressure**: These features were also identified by the SHAP analysis as contributing to CVD predictions, further validating the exploratory analysis results.

6.3 Validation and Consistency

The alignment between the SHAP results and the findings from the exploratory analysis demonstrates the robustness and validity of the feature importance identified by the machine learning models. The agreement across different models (Logistic Regression, SVM, Random Forest, FCNN) and explainability techniques (SHAP and LIME) adds credibility to the insights drawn from this study. Key clinical features such as time, ejection fraction, serum creatinine, serum sodium, age, and smoking consistently appeared as top predictors of CVD, confirming their relevance in both the statistical and machine learning contexts.

The integration of SHAP and LIME allowed for the validation of these findings at both the global (dataset-wide) and local (individual predictions) levels, ensuring that the model outputs were interpretable and aligned with medical understanding.

7 Conclusion

The developed models, particularly the Random Forest and Fully Connected Neural Network (FCNN), demonstrated high accuracy and reliability in predicting cardiovascular disease. The integration of explainable AI techniques such as SHAP and LIME provided critical insights into the models' decision-making processes, making it easier to interpret the influence of individual features. This added transparency is crucial in healthcare settings where model interpretability can enhance trust and facilitate informed decision-making.

7.1 Rationale for Not Using CNN, LSTM, and CNN-LSTM Models

The decision to not use more advanced architectures such as Convolutional Neural Networks (CNNs), Long Short-Term Memory networks (LSTMs), or hybrid CNN-LSTM models was based on the nature of the dataset. CNNs are particularly suited for image data or data with a spatial structure, and LSTMs excel in processing sequential data or time-series information. However, the dataset used in this study consists of structured tabular data with no inherent spatial or sequential relationships between the features.

For structured tabular datasets like this one, traditional machine learning models (e.g., Logistic Regression, SVM, Random Forest) and Fully Connected Neural Networks (FCNNs) are generally more appropriate as they are well-equipped to handle tabular data where each feature is independent and has no sequential order. CNNs and LSTMs would likely not offer any significant performance improvement and could lead to overfitting due to the limited amount of data and the absence of complex spatial or temporal dependencies.

Moreover, FCNNs are capable of capturing complex interactions between the features while remaining computationally efficient for tabular data. This made them the preferred deep learning model for this study. The combination of FCNNs with explainable AI techniques provided not only high accuracy but also interpretability, ensuring that the models' predictions were both effective and explainable.