Automated Abnormal ECG Detection for Early Heart Disease Diagnosis

PRESENTATION BY

K.SAIRAHUL_21BCE9121
Ch.KUSELA_21BCE9158

AGENDA

Problem statement

MIT-BIH Arrhythmia Dataset

ECG Signal Processing

Deep Learning Model Architecture

Model Training

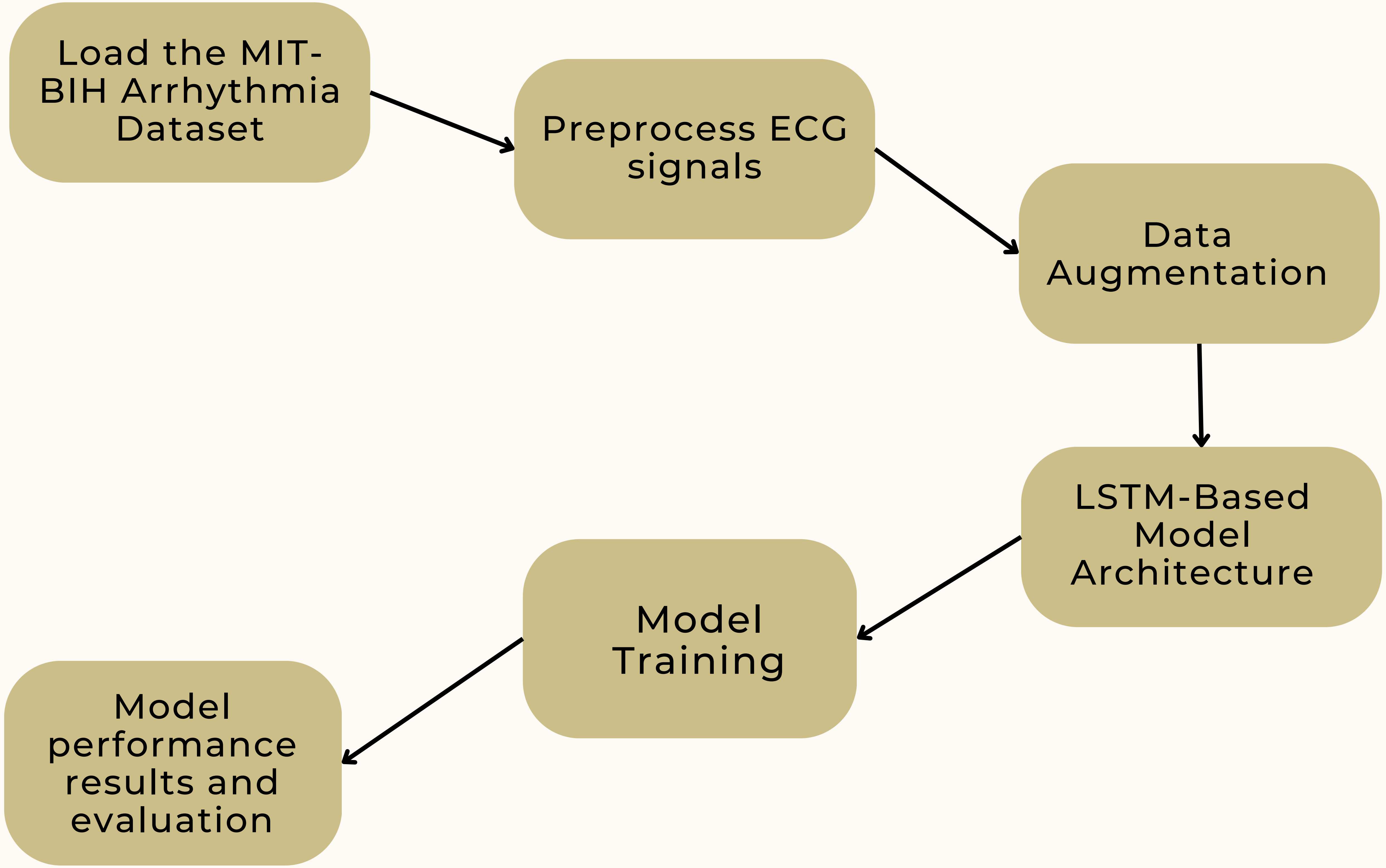
Our Research

Results

PROBLEM STATEMENT

Automated Abnormal ECG Detection for Early Heart Disease Diagnosis

The main aim of this project is to develop an intelligent system that can automatically detect abnormal patterns in ECG signals and classify the type of cardiac arrhythmias present, if any. The system should take an input of pre-recorded ECG signal data for analysis and provide a detailed output report of arrhythmia classification, ST segment and QT interval analysis, response to treatment evaluation, and risk stratification. It helps diagnose cardiac conditions, assess treatment response, and evaluate cardiac health status.


Load the MIT-BIH Arrhythmia Dataset
Preprocess ECG signals
Data Augmentation
LSTM-Based Model Architecture
Model Training
Model performance results and evaluation

MIT-BIH Arrhythmia Dataset

- The dataset consists of 48 half-hour ECG recordings, sampled at 360 Hz, collected from 47 different patients.
- Each ECG recording is digitized
- Format of the Dataset:Available in CSV and TXT files.
- CSV: ECG signal data in tabular format.
- TXT: Annotations for R-peaks and arrhythmias.
- Number of Samples and Classes: 109,446 ECG recordings.

Here are the five classes:

- **Class 'N': Normal rhythm**

Represents normal sinus rhythm, which is the standard regular rhythm of the heart.

- **Class 'S': Supraventricular ectopic beat**

Represents an ectopic (abnormal) heartbeat originating from above the ventricles.

- **Class 'F': Fusion of ventricular and normal beat**

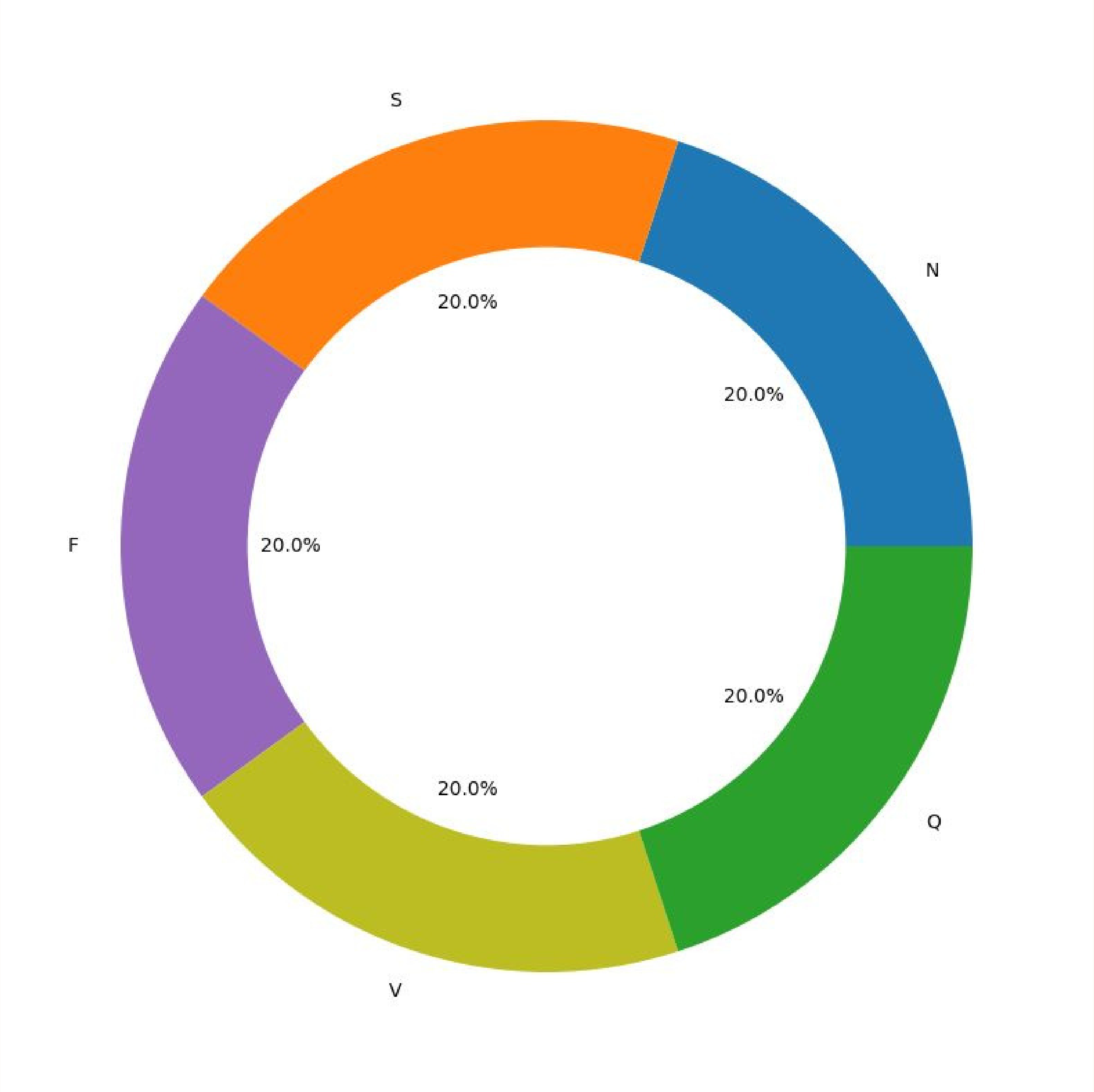
Represents a fusion beat that combines characteristics of both ventricular and normal beat.

- **Class 'V': Ventricular ectopic beat**

Represents an ectopic (abnormal) heartbeat originating from the ventricles.

- **Class 'Q': Paced beat**

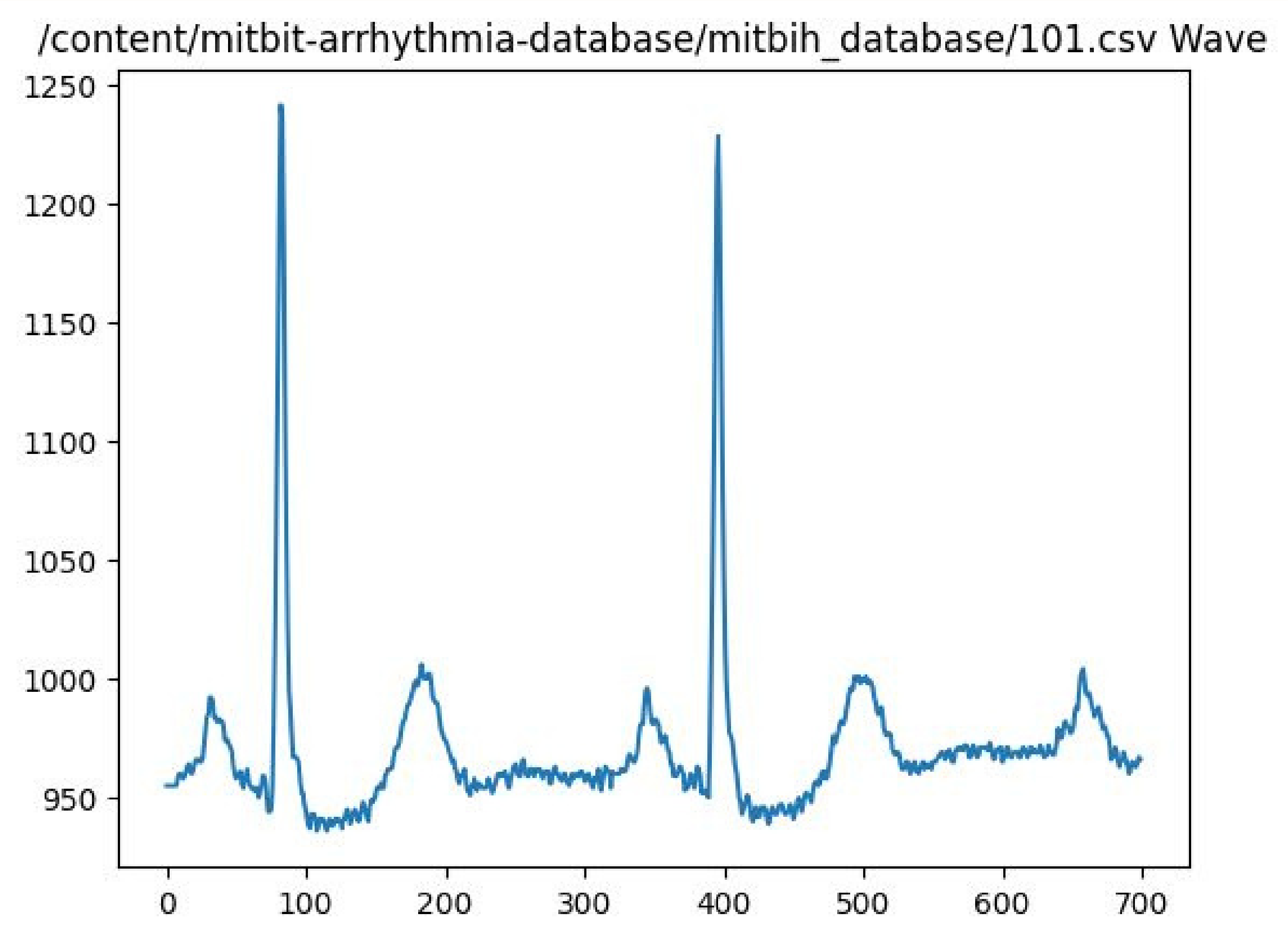
Represents a heartbeat that is artificially stimulated (paced) by an external pacemaker.

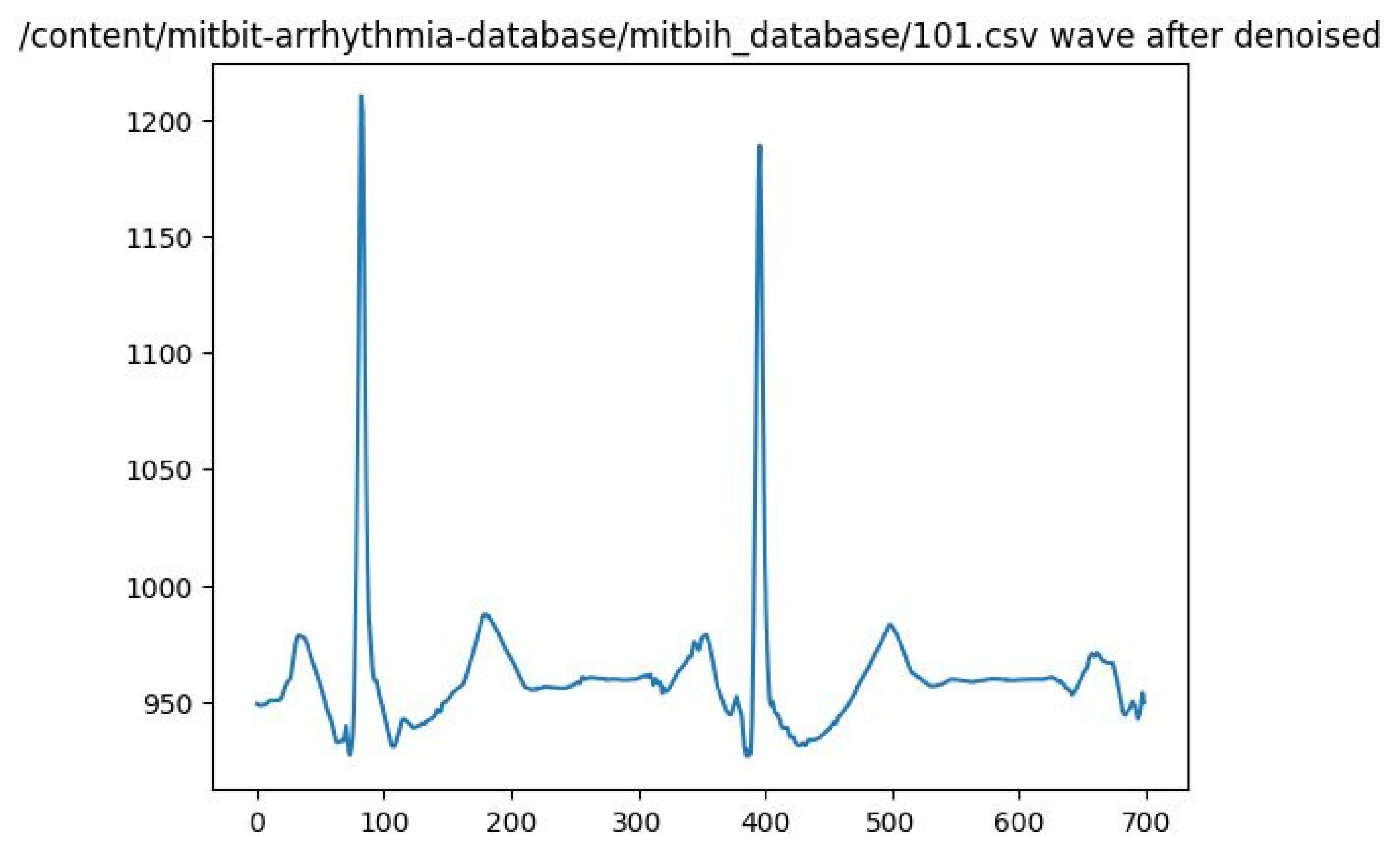


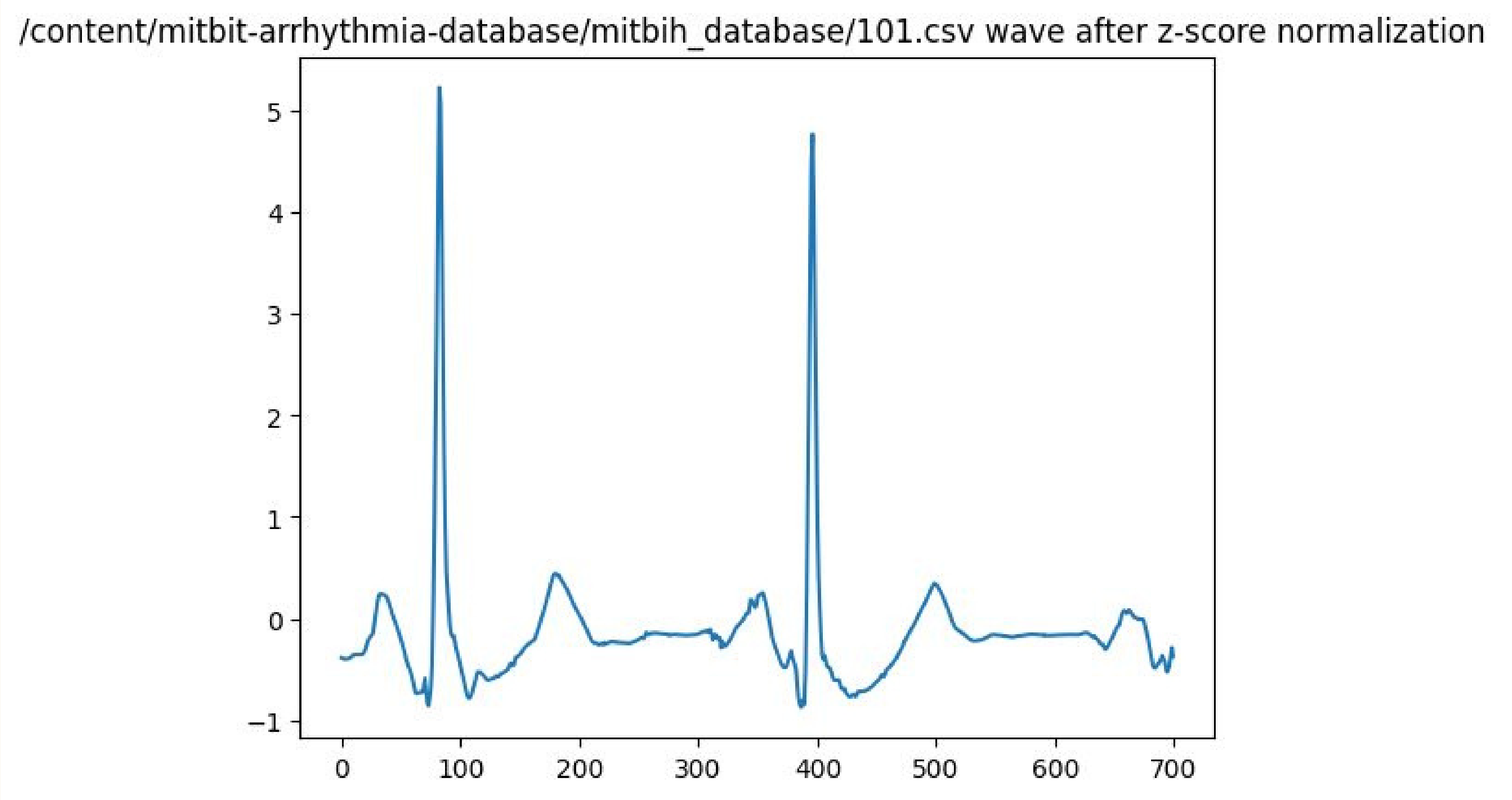
Preprocessing Steps for ECG Signals:

1. Removal of Baseline Wander
2. Apply Filtering Techniques (Low-Pass, High-Pass, Bandpass)
3. Resampling (Standardize Sampling Rate)
4. R-Peak Detection (Identify QRS Complex)
5. Normalization (Scale Amplitude)
6. Denoising (Wavelet Denoising)

These steps enhance accuracy, remove noise, and enable detection of abnormal patterns in ECG signals.


/content/mitbit-arrhythmia-database/mitbih_database/101.csv Wave
1250
1200
1150
1100
1050
1000
950
0
100
200
300
400
500
600
700


/content/mitbit-arrhythmia-database/mitbih_database/101.csv wave after denoised
1200
1150
1100
1050
1000
950
0
100
200
300
400
500
600
700


/content/mitbit-arrhythmia-database/mitbih_database/101.csv wave after z-score normalization
5
4
3
2
1
0
−1
0
100
200
300
400
500
600
700

LSTM-Based Model Architecture for ECG Analysis

- Suitable for sequential data like ECG signals, capturing complex patterns.
- Architecture includes an LSTM layer with 64 units, a dense layer with 128 units, a dropout layer with a dropout rate of 0.5, and a dense output layer with 5 units for classification.
- Enables learning long-term dependencies.
- Reduces manual feature engineering.
- Improved accuracy in detecting heart diseases.
- Aids in early diagnosis and better patient care

Arrhythmia Classification

The classification of arrhythmia into 5 classes is done using a pre-trained LSTM model.

- **Load Pre-trained LSTM Model:** The model should have been trained on a labeled dataset containing ECG signals with their corresponding arrhythmia classes.
- **Preprocess ECG Data for Classification:** The same preprocessing steps mentioned earlier are applied to the ECG data to make it suitable for feeding into the LSTM model.
- **Predictions:** The preprocessed ECG data is fed into the LSTM model to make predictions. The **model outputs a probability distribution over the 5 arrhythmia classes for each beat. The class with the highest probability is selected as the predicted arrhythmia class for each beat.**

Model Training

- LSTM-based model trained on augmented data for improved performance.
- We have applied data augmentation techniques that includes signal scaling, shifting, noise addition, time warping, flipping, and cropping.
- Categorical Crossentropy used as the loss function.
- Optimized using Adam optimizer.
- Trained for 20 epochs with a batch size of 100.

Other main Features extracted:

```
R-peaks: [    154     740    1326 ... 1298970 1299468 1299982]
Q-Peaks: [0 0 0 ... 0 0 0]
S-Peaks: [    166     758    1338 ... 1298980 1299480 1299998]
J-Points: [    166     758    1338 ... 1298980 1299480 1299998]
T-Peaks: [    154     740    1326 ... 1298970 1299468 1299982]
ST-Segments: [array([], dtype=float64), array([], dtype=float64), array([], dt
QT Intervals: [    154     740    1326 ... 1298970 1299468 1299982]
RR Intervals: [586 586 568 ... 506 498 514]
PR Intervals: [    154     740    1326 ... 1298970 1299468 1299982]
```

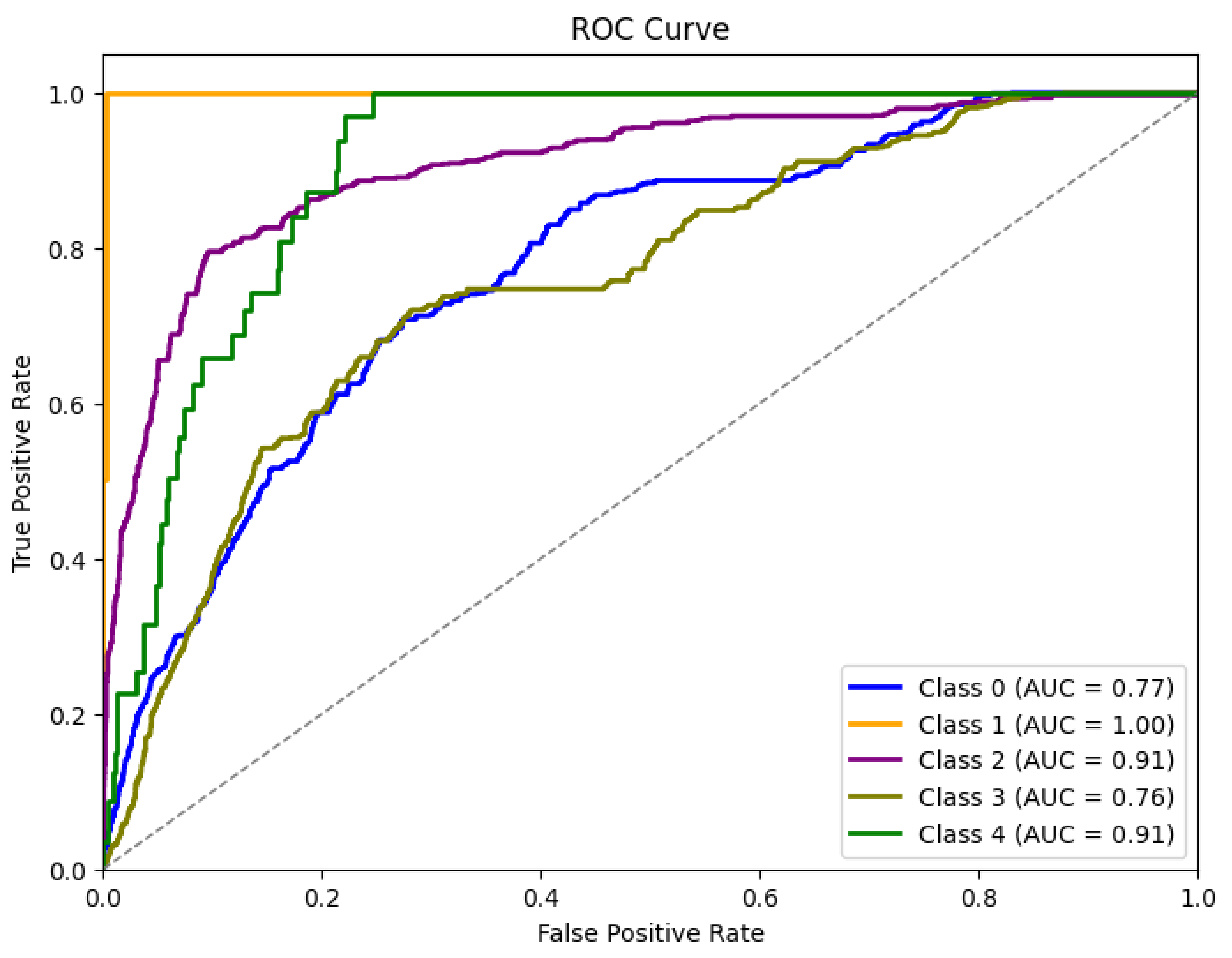
Our Research

- **ST Segment Analysis:** Measure the length of the ST segment for each beat in the ECG signal. The ST segment represents the interval between the end of the QRS complex and the beginning of the T wave. Analyzing the ST segment can help identify potential abnormalities in the heart's electrical activity.
- **QT Interval Analysis:** Calculate the duration of the QT interval for each beat in the ECG signal. The QT interval reflects the time it takes for the heart's ventricles to depolarize and repolarize. Prolonged or shortened QT intervals can indicate specific cardiac conditions.

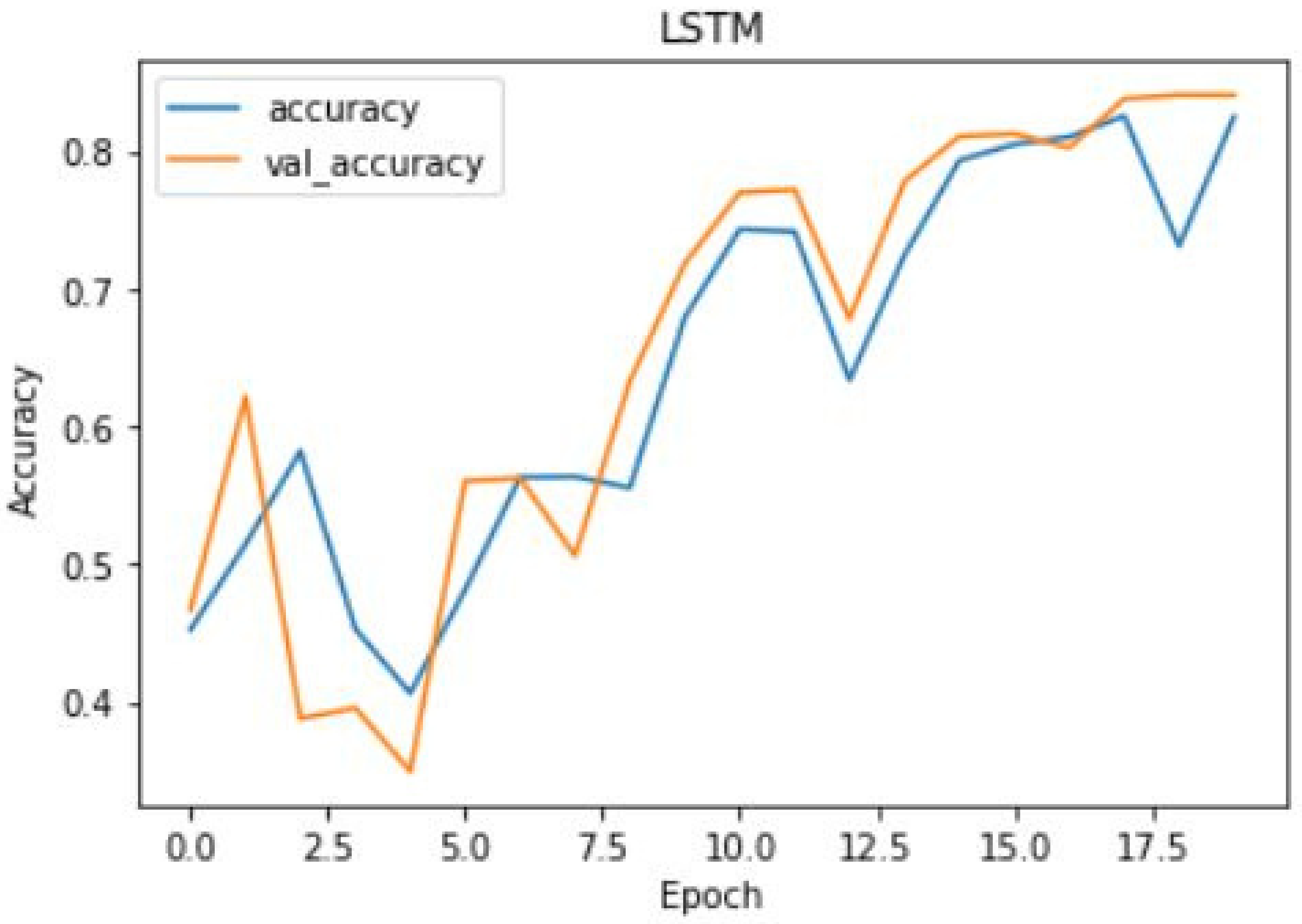
Our Research

- **Response to Treatment Analysis:** Based on the QT intervals, provide insights into how the patient may respond to specific treatments. This analysis is particularly relevant for patients with prolonged QT intervals, as it can help assess the potential risks associated with certain medications and treatments.
- **Risk Stratification Analysis:** Evaluate the patient's risk level based on the ECG signal, especially considering the average RR interval as an indicator of heart rate variability. Risk stratification can aid in assessing the likelihood of certain cardiac events or conditions

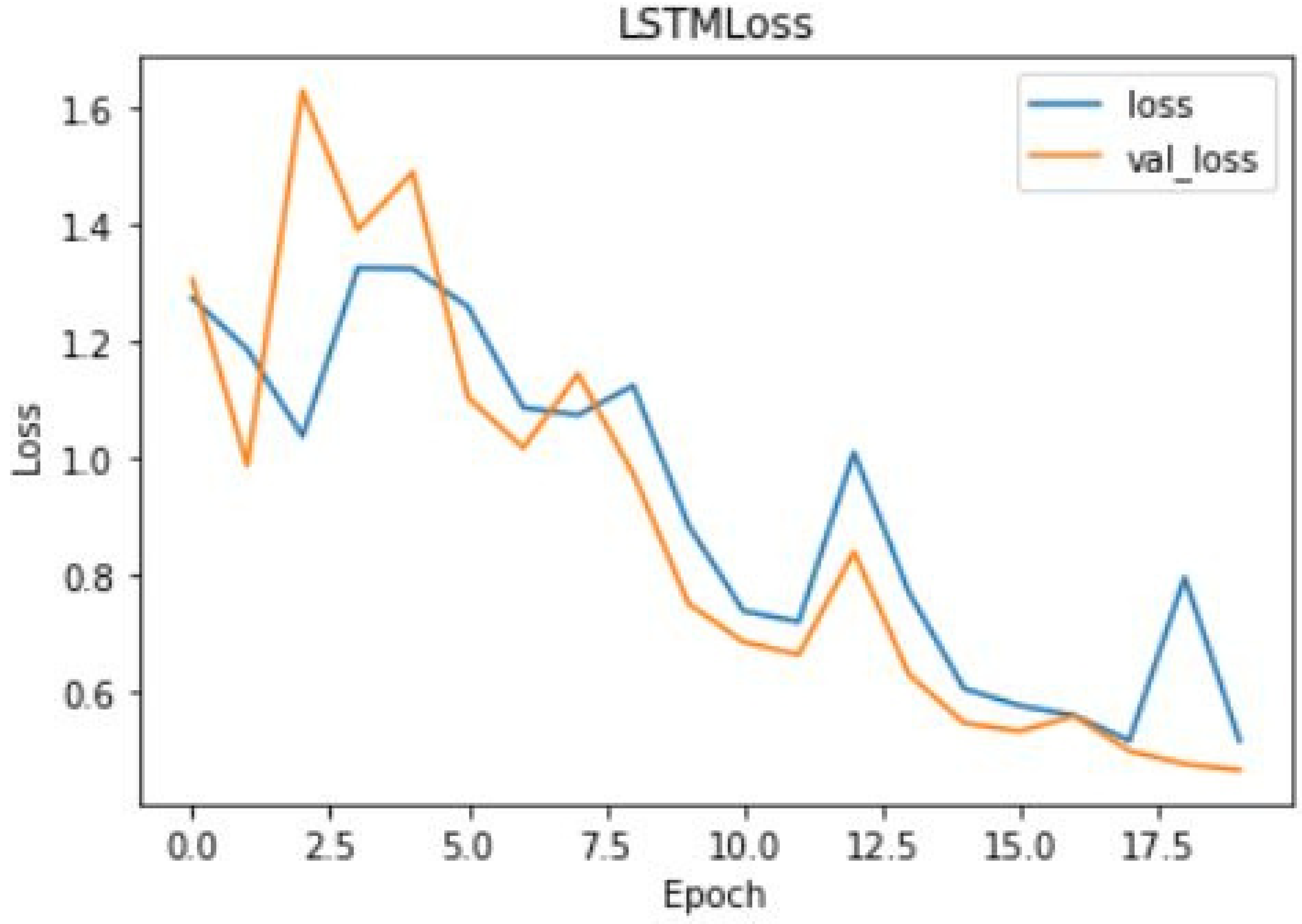
The results of these analyses can provide valuable information for medical decision-making and patient care.



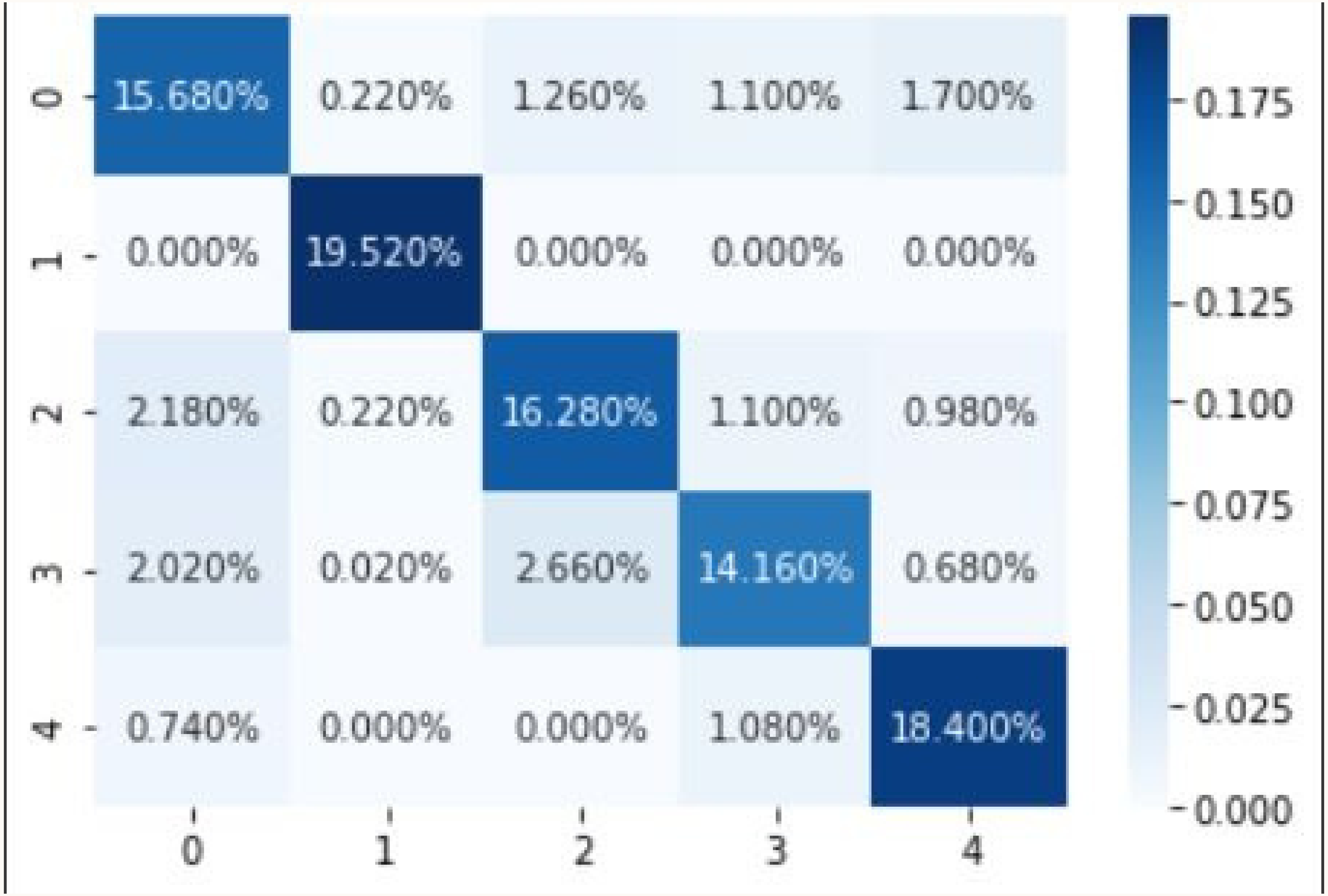
ROC Curve
True Positive Rate
False Positive Rate
0.0
0.2
0.4
0.6
0.8
1.0
Class 0 (AUC = 0.77)
Class 1 (AUC = 1.00)
Class 2 (AUC = 0.91)
Class 3 (AUC = 0.76)
Class 4 (AUC = 0.91)



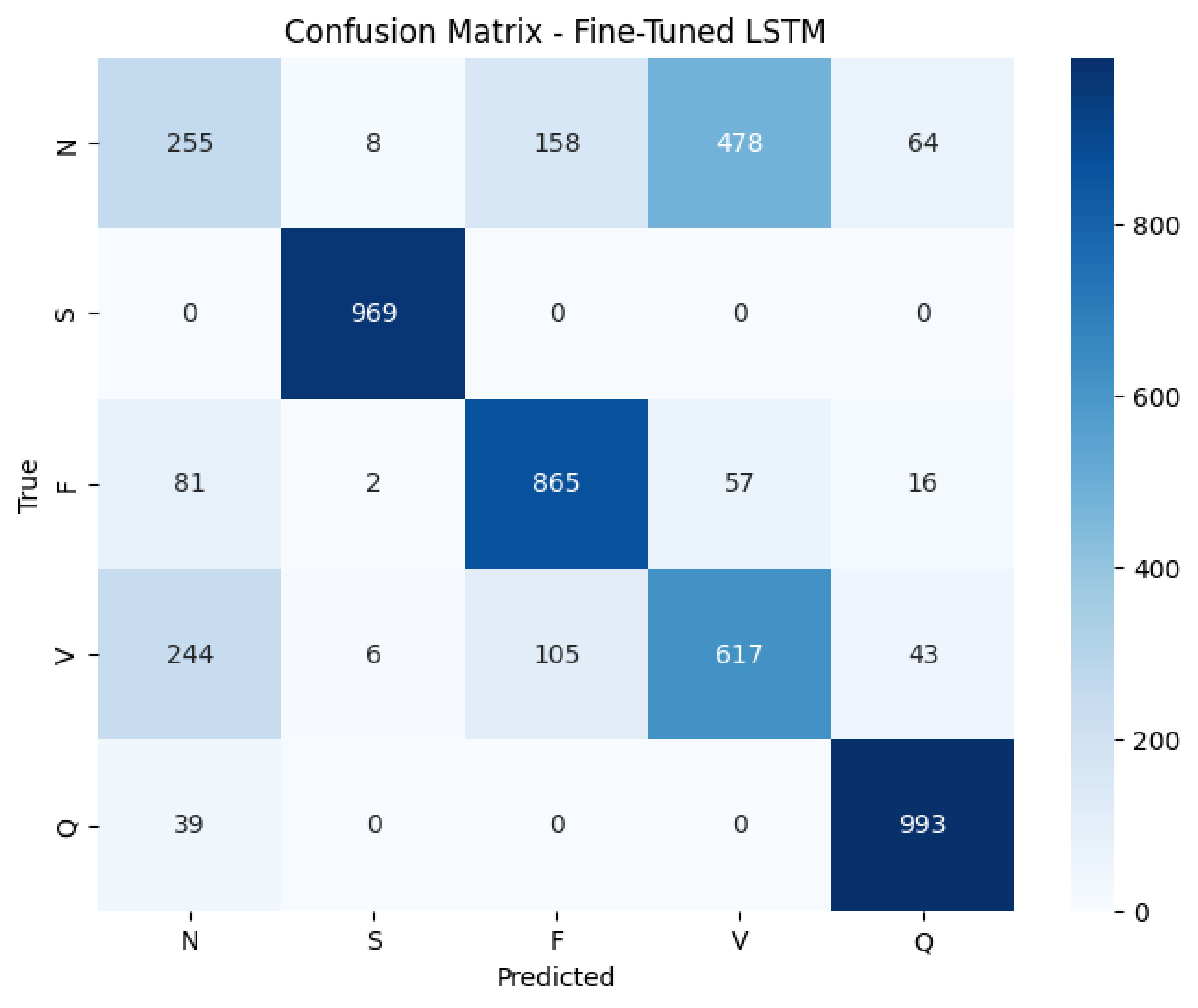
LSTM
accuracy
val_accuracy
Accuracy
0.4
0.5
0.6
0.7
0.8
0.0
2.5
5.0
7.5
10.0
12.5
15.0
17.5
Epoch



LSTMLoss
loss
val_loss
Loss
0.6
0.8
1.0
1.2
1.4
1.6
0.0
2.5
5.0
7.5
10.0
12.5
15.0
17.5
Epoch

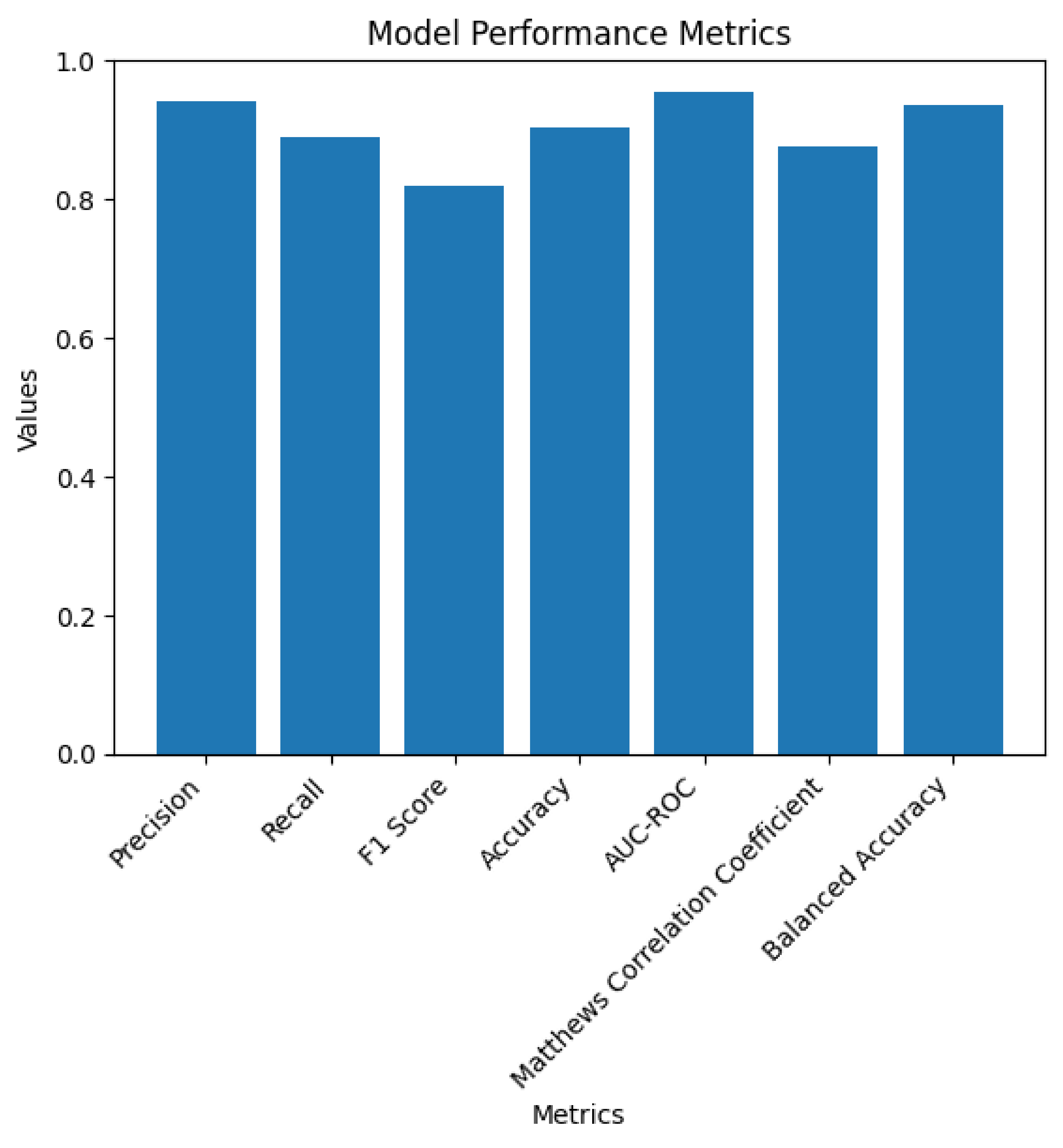

0
1
2
3
4
15.680%
0.220%
1.260%
1.100%
1.700%
0.000%
19.520%
0.000%
0.000%
0.000%
2.180%
0.220%
16.280%
1.100%
0.980%
2.020%
0.020%
2.660%
14.160%
0.680%
0.740%
0.000%
0.000%
1.080%
18.400%
0
1
2
3
4
0.175
0.150
0.125
0.100
0.075
0.050
0.025
0.000

The confusion matrix evaluates the arrhythmia classification model's performance, comparing predicted results to actual labels. The matrix displays true positives (correct predictions) along the diagonal and false positives/negatives off the diagonal. Metrics like accuracy, precision, recall, and F1-score provide insights into overall performance and per-class evaluation. Specificity measures the model's ability to identify true negatives. The balanced accuracy gives an overall balanced view. Understanding the confusion matrix aids healthcare professionals in making informed decisions and enhancing patient care based on the model's effectiveness in classifying arrhythmia types.



- Precision: 0.9398 (Low false positives, accurate positive predictions)
- Recall: 0.8898 (Captures a high proportion of actual positive samples)
- F1 Score: 0.8198 (Balance between precision and recall)
- Accuracy: 0.9017 (Overall correctness of predictions)
- AUC-ROC: 0.9537 (High ability to distinguish between classes)
- Matthews Correlation Coefficient: 0.8770 (Balanced measure of classification performance)
- Balanced Accuracy: 0.9364 (Suitable for imbalanced datasets)

Overall, the model demonstrates strong performance, effectively detecting abnormal ECG patterns for early heart disease diagnosis.



CONCLUSION

The developed LSTM-based model demonstrates promising results in automated ECG analysis for early heart disease diagnosis. Its high precision, recall, F1 score, and accuracy showcase its potential for reliable arrhythmia detection and classification. By significantly reducing the manual workload of medical professionals, the automated ECG analysis model enables quicker and more accurate heart disease diagnoses. Early detection of heart diseases can potentially save lives and improve patient outcomes, making the model a valuable addition to the medical field. With continuous improvements and validation, the model may become an indispensable tool in regular health monitoring and telemedicine applications.

THANK YOU

K.SAI RAHUL_21BCE9121
Ch.KUSELA_21BCE9158