ECG-based Heart Disease Detection Using Deep Learning Models

ANNBAR MEKOUARNew York, United States   
[annbar.mekouar@my.liu.edu](mailto:annbar.mekouar@my.liu.edu)

SHARIS STANLEY REBEIRONew York, United States  
Sharisstanley.rebeiro@my.liu.edu  
  
AMEER KHANNew York, United States

ameerkhan.jaffarkhankhaderkhan@my.liu.edu

This project presents a deep learning approach for ECG-based heart disease detection using the PTB-XL dataset. The methodology includes preprocessing 12-lead ECG signals, extracting QRS complex features, and implementing a hybrid Convolutional Neural Network (CNN) and Long Short-Term Memory (LSTM) model. Initially, the model was trained for binary classification, achieving an accuracy of 85.19% in differentiating between normal and abnormal ECG signals. The system was then extended to perform multiclass classification across five diagnostic categories: Normal, Myocardial Infarction, ST/T Changes, Conduction Disorders, and Hypertrophy. The multiclass ensemble model reached a test accuracy of 51.76%. These results demonstrate the potential of deep learning for scalable and accurate ECG diagnostics.

Keywords— ECG, QRS Complex, Deep Learning, CNN-LSTM, PTB-XL, Heart Disease Detection, Machine Learning in Cardiology, Binary ^Classification, Time Series Analysis

# Introduction

Cardiovascular disease (CVD) is one of the most prevalent causes of mortality worldwide. Electrocardiography (ECG) is a non-invasive diagnostic tool that records the electrical activity of the heart. Manual ECG interpretation is prone to inter-observer variability and requires expertise. This project proposes a deep learning-based model that automatically detects cardiac abnormalities using raw ECG signals from the PTB-XL dataset. We implemented a hybrid CNN-LSTM architecture trained for binary classification (Normal vs. Disease), improving diagnostic speed and consistency.

# Background: Heart Morphology and ECG Signals

The human heart is a muscular, four-chambered organ consisting of two atria (upper chambers) and two ventricles (lower chambers). [3] The right side of the heart receives deoxygenated blood from the body and pumps it to the lungs via the pulmonary artery for oxygenation. The left side receives oxygenated blood from the lungs and pumps it to the rest of the body through the aorta. This dual-circulation system ensures the continuous delivery of oxygen and removal of carbon dioxide.

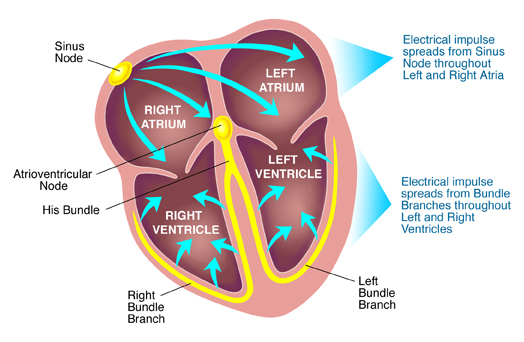


Figure 1. Electrical Conduction System Diagram

The heart’s activity is regulated by a specialized electrical conduction system. The sinoatrial (SA) node, located in the right atrium, acts as the natural pacemaker and initiates the heartbeat by generating an electrical impulse. This impulse spreads through the atria, causing them to contract and push blood into the ventricles. The signal then passes through the atrioventricular (AV) node, which delays the impulse slightly to allow complete ventricular filling. From there, it travels down the bundle of His, splits into right and left bundle branches, and reaches the Purkinje fibers that trigger synchronized ventricular contraction.

Electrocardiography (ECG)[4] is a non-invasive diagnostic method that records the electrical activity of the heart over time using electrodes placed on the skin. A standard 12-lead ECG provides views of the heart’s electrical behavior from multiple angles. It captures different phases of the cardiac cycle: atrial depolarization (P wave), ventricular depolarization (QRS complex), and ventricular repolarization (T wave). Each waveform component carries diagnostic information. For example, prolonged PR intervals suggest conduction delays, while ST-segment elevation is a hallmark of acute myocardial infarction. The QRS duration can indicate bundle branch blocks or ventricular hypertrophy.

Une image contenant texte, dessin humoristique, diagramme, clipart

Le contenu généré par l’IA peut être incorrect.

Figure 2. P, QRS, and T waves with corresponding cardiac activity.[5]

In digital signal processing and machine learning, the ECG is treated as a multi-channel time-series signal. This makes it suitable for automated analysis, where computational models can identify patterns across time and channels (leads) to detect cardiac diseases. The morphology of the heart directly influences ECG waveform patterns, making physiological understanding essential to developing interpretable and accurate diagnostic models.

# PTB-XL Dataset and Patient Data Relevance

The PTB-XL dataset includes over 21,000 12-lead ECG recordings from nearly 19,000 patients[6]. Each ECG lasts 10 seconds, sampled at both 500 Hz and 100 Hz. The dataset covers five diagnostic classes: Normal (NORM), Myocardial Infarction (MI), ST/T changes (STTC), Conduction Disorders (CD), and Hypertrophy (HYP). Patient data is essential for training machine learning models. It allows recognition of diagnostic patterns but introduces privacy, standardization, and bias challenges. For this project, the original five diagnostic categories were consolidated into a binary classification problem: Normal vs. Disease.

To prepare the ECG signals for modeling:

* Signals were filtered with a bandpass filter (0.5–40 Hz) to remove baseline wander and high-frequency noise.
* Signals were normalized for amplitude consistency.
* We used a modified Pan-Tompkins algorithm to detect R-peaks.
* Feature extraction included calculating QRS duration, RR intervals, and signal amplitude statistics for each lead.

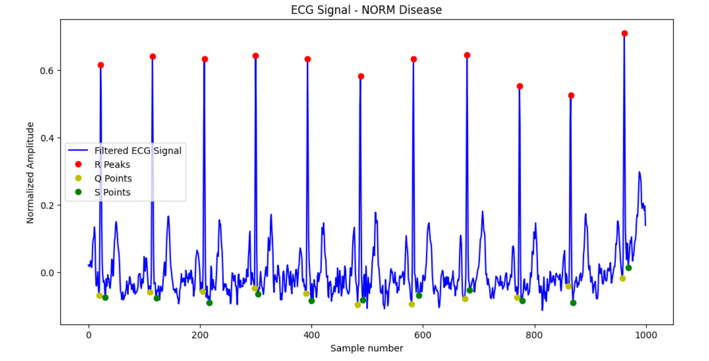


Figure 3. Normal ECG from the PTBL XL dataset.

Une image contenant Tracé, ligne, capture d’écran, diagramme

Le contenu généré par l’IA peut être incorrect.

Figure 4. Age distribution of patients by diagnostic category in PTB-XL.

This dataset provides a robust basis for training supervised machine learning models, but also poses challenges such as class imbalance, age-based variance, and potential noise in clinical ECGs.

Patient data is critical in machine learning, enabling models to learn from patterns across populations. While it allows for prediction and personalization, it also introduces challenges like privacy, standardization, and bias management.

# Machine Learning in Cardiology

Machine learning has become a critical tool in modern cardiology by enabling the analysis of high-dimensional physiological signals like ECG. These models can automatically identify hidden patterns in temporal data that may be difficult to interpret manually. In cardiac diagnostics, machine learning is commonly applied to tasks such as arrhythmia detection, myocardial infarction classification, hypertrophy diagnosis, and conduction disorder identification.

Supervised learning is central to these applications. With access to labeled ECG data like that provided in the PTB-XL dataset, models can be trained to classify heart rhythms based on waveform morphology and timing intervals.

Convolutional Neural Networks (CNNs) are widely used for extracting spatial features from ECG signals, such as QRS morphology, ST elevation, and T-wave changes. Recurrent Neural Networks (RNNs)—particularly Long Short-Term Memory (LSTM) networks—capture temporal dependencies across heartbeats by modeling relationships between R-R intervals and evolving wave patterns.

To combine the strengths of both models, hybrid CNN-LSTM architectures are often employed. In our final implementation, CNN layers extracted spatial features from raw 12-lead ECG signals, which were then passed into bidirectional LSTM layers to learn time-dependent dynamics. This approach improves classification performance by jointly modeling morphology and rhythm.

Recent work in interpretable AI has introduced techniques such as SHAP and LIME to enhance transparency in model predictions. These tools help clinicians trust automated outputs by visualizing which features contribute to each diagnosis.

Additionally, approaches like data augmentation, imbalance handling, and denoising are being incorporated to improve model robustness across diverse clinical ECG recordings.

# Methodology

The methodology is centered around building a complete machine learning pipeline for ECG-based heart disease classification. The initial step is signal preprocessing. Each 12-lead ECG signal is filtered using a bandpass filter between 0.5 Hz and 40 Hz. This removes baseline wander and high-frequency noise. Signals are then normalized to a common amplitude scale.

A critical step is detecting the QRS complex. We apply a modified Pan-Tompkins algorithm, which uses a combination of signal squaring, differentiation, and moving window integration to locate R-peaks. From these peaks, we extract key temporal and morphological features:

* R-R Intervals: time between successive R-peaks
* QRS Width: duration between Q and S points
* QRS Amplitude: difference between R peak and baseline
* Heart Rate Variability: derived from R-R series
* Lead-specific variation: all features are extracted per lead

These extracted features are stored in structured arrays with labels from the PTB-XL dataset for supervised training.

We visualize key diagnostic segments, verify feature alignment, and perform data exploration to examine class balance, feature distributions, and signal consistency.

## QRS feature detection

Feature extraction is functional and consistent with known ECG physiology. Visual checks on randomly selected signals confirm the accurate detection of R, Q, and S points. Heart rate and QRS width distributions are within expected ranges for both normal and abnormal classes.

Signal plots confirm appropriate preprocessing, and class-specific trends (e.g., wide QRS in conduction disorders) are observable. A distribution of age and diagnosis classes has been analyzed to understand demographic variation.

Une image contenant texte, capture d’écran, Tracé, ligne

Le contenu généré par l’IA peut être incorrect.

Figure 3. Normal ECG with identified R, Q, and S points.

## Binary Classification

Following successful QRS feature extraction and validation, we implemented a binary classification model using a hybrid deep learning architecture. The model was designed to distinguish between normal and abnormal ECG recordings based on spatial and temporal waveform patterns.

The input to the model consists of preprocessed 12-lead ECG signals formatted as fixed-size time windows. A series of 1D convolutional layers were applied to capture spatial features across the ECG channels, identifying key waveform characteristics such as QRS shape, ST elevation, and signal amplitude variation. These convolutional outputs were passed to a bidirectional Long Short-Term Memory (LSTM) layer, which models sequential dependencies and beat-to-beat variability, particularly useful for analyzing RR intervals and rhythm disturbances.

The final part of the network includes fully connected dense layers and a sigmoid-activated output layer to produce a probability score indicating the likelihood of cardiac abnormality. The model was trained using the Adam optimizer and binary cross-entropy loss. Training was conducted over 100 epochs with a batch size of 64 and a 70/30 training-validation split.

The classification output was evaluated using test set accuracy, confusion matrix analysis, and metrics such as precision and recall. Throughout training, loss convergence and validation trends were monitored to prevent overfitting and ensure model generalization.

This hybrid CNN-LSTM approach effectively leverages both morphological and temporal aspects of ECG signals, enabling robust classification performance in the context of heart disease detection.

## Multiclass Classification

In the final phase, the binary classification model was expanded to support multiclass classification. The same CNN-LSTM architecture was adapted by replacing the binary output layer with a softmax-activated dense layer for five-category prediction. The training objective was changed to categorical cross-entropy loss, and the model was evaluated using accuracy, confusion matrix.

The model was trained on ECG segments labeled according to the five PTB-XL diagnostic classes: NORM, MI, STTC, CD, and HYP integrating also the QRS features of each signal. An ensemble method was also implemented to improve performance consistency. This phase demonstrates the model’s ability to distinguish between multiple clinically relevant conditions, despite the natural overlap in ECG waveforms between categories.

# Results

1. *Binary Classification*

This study implemented a binary classification model using a CNN-LSTM hybrid architecture trained on 12-lead ECG signals from the PTB-XL dataset. A key innovation in the design was the explicit inclusion of QRS complex features extracted from all leads. This addition provided rich physiological detail that significantly enhanced the model’s ability to detect subtle cardiac abnormalities.

The final model achieved a test accuracy of 85.19%, demonstrating strong generalization to unseen ECG signals. The QRS complex added vital diagnostic information, allowing the model to capture both waveform morphology and beat-to-beat dynamics. By doing so, it improved anomaly detection and ensured more accurate identification of diseased versus normal signals.

The confusion matrix (Figure 4) illustrates this performance, showing a true positive rate and a true negative rate both at 85.19%. This balance indicates reliable performance across both classes, with a misclassification rate of just 14.81% per class.

Training progress is shown in Figure 4, where both training and validation loss curves converge toward 0.4 over 35 epochs. The minimal divergence between the curves suggests effective learning and little overfitting.

Beyond metrics, the addition of the QRS complex reinforces the importance of incorporating structured cardiac features into deep learning pipelines. It supports the model’s architectural choices and highlights the clinical relevance of waveform-level detail in disease detection. This integration contributes not only to diagnostic accuracy but also to potential improvements in treatment planning and healthcare delivery.

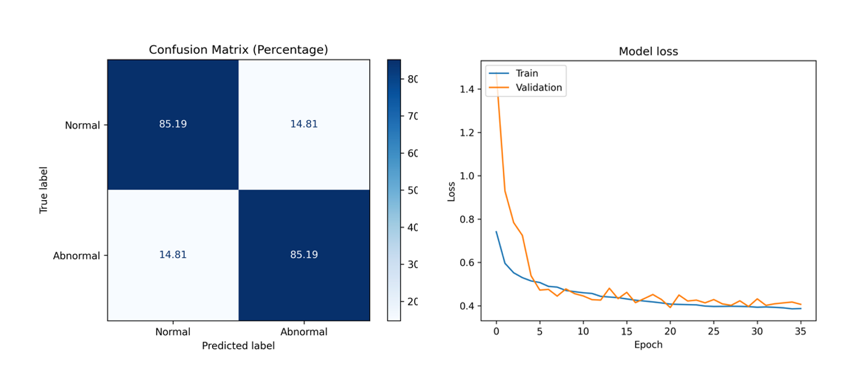


Figure 4. Training validation Loss and Confusion Matrix of the Binary Classification

1. *Multiclass Classification*

The multiclass CNN-LSTM model achieved a test accuracy of 51.76%. The confusion matrix shows that the model best classified NORM (57.14%), STTC (56.27%), and CD (56.54%), with lower accuracy in differentiating HYP and MI due to overlapping waveform patterns.

Training and validation accuracy curves indicate steady convergence over 30 epochs. Although the multiclass model’s performance is lower than the binary version, it demonstrates meaningful diagnostic capacity across several categories. These results suggest that with further optimization and advanced techniques (e.g. data augmentation), multiclass ECG diagnosis can be further improved.

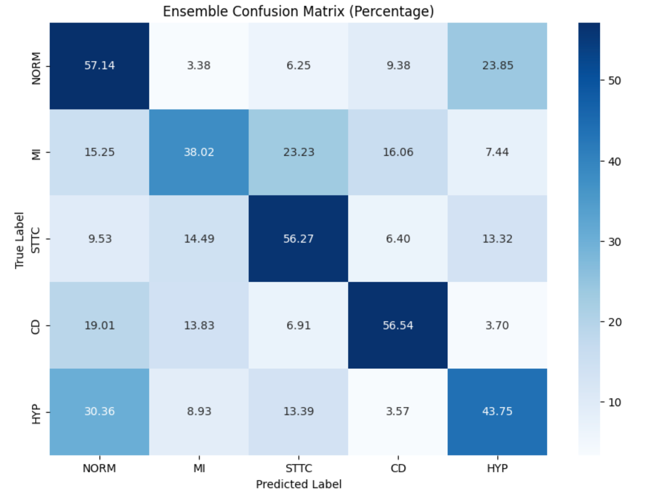


Figure 5. Confusion Matrix of the Multiclass Classification

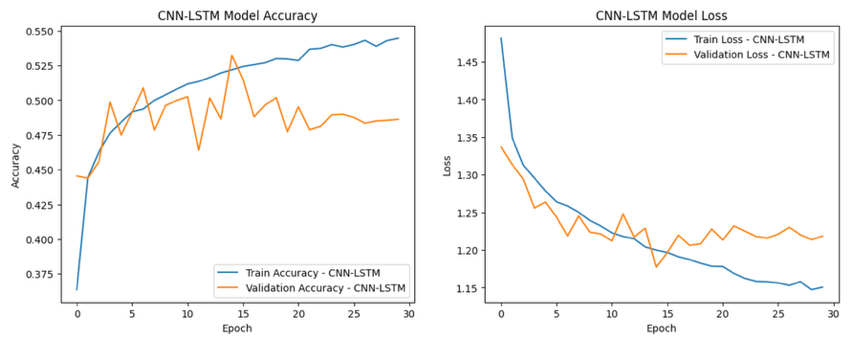


Figure 6. Training validation Loss and Confusion Matrix of the Binary Classification

# Conlcusion

This project presents a comprehensive deep learning pipeline for automated ECG-based heart disease diagnosis. Leveraging the PTB-XL dataset, we developed a signal processing workflow that includes denoising, normalization, and robust QRS complex detection across all 12 leads. These physiological features—QRS width, RR intervals, and amplitude—formed the foundation for accurate, data-driven cardiac classification.

We first implemented a hybrid CNN-LSTM architecture for binary classification, achieving a strong test accuracy of 85.19%, with balanced sensitivity and specificity. The explicit use of QRS morphology greatly enhanced the model’s ability to capture spatial and temporal ECG dynamics.

Building on this success, we extended our approach to multiclass classification, targeting five diagnostic categories: NORM, MI, STTC, CD, and HYP. The ensemble CNN-LSTM model achieved a test accuracy of 51.76%, with the highest performance observed in distinguishing Normal, ST/T Changes, and Conduction Disorders. These results demonstrate the feasibility of deep learning in supporting clinical ECG interpretation across multiple disease types.

Future work will focus on improving multiclass accuracy through architectural optimization, interpretability tools like SHAP and LIME, and testing on external datasets. Ultimately, we aim to integrate this pipeline into real-time diagnostic environments, contributing to scalable and explainable AI solutions in cardiology.

##### References

[1] L. Alzubaidi *et al.*, “Review of deep learning: concepts, CNN architectures, challenges, applications, future directions,” *Journal of Big Data*, vol. 8, no. 1, p. 53, Mar. 2021, doi: 10.1186/s40537-021-00444-8.

[2] “Long Short-Term Memory Networks (LSTM)- simply explained! | Data Basecamp.” Accessed: Apr. 07, 2024. [Online]. Available: https://databasecamp.de/en/ml/lstms

[3] “Heart | Structure, Function, Diagram, Anatomy, & Facts | Britannica.” Accessed: Sep. 18, 2024. [Online]. Available: https://www.britannica.com/science/heart

[4] U. F. O. Themes, “ECG Leads,” Thoracic Key. Accessed: Apr. 07, 2024. [Online]. Available: https://thoracickey.com/ecg-leads/

[5] “What does QRS complex represent in ECG?,” BYJUS. Accessed: Apr. 07, 2024. [Online]. Available: https://byjus.com/neet/what-does-qrs-complex-represent-in-ecg/

[6] P. Wagner *et al.*, “PTB-XL, a large publicly available electrocardiography dataset,” *Sci Data*, vol. 7, no. 1, p. 154, May 2020, doi: 10.1038/s41597-020-0495-6.

[7] Y. M. Ayano, F. Schwenker, B. D. Dufera, and T. G. Debelee, “Interpretable Machine Learning Techniques in ECG-Based Heart Disease Classification: A Systematic Review,” *Diagnostics*, vol. 13, no. 1, Art. no. 1, Jan. 2023, doi: 10.3390/diagnostics13010111.

[8] A. Kalimi, “List of Machine Learning Models,” Medium. Accessed: Mar. 23, 2024. [Online]. Available: https://medium.com/@codekalimi/list-of-machine-learning-models-61b51ad492f1