DS 504 – Final Project Report

Advancing Automated ECG Interpretation

Team 11: *Aayush Sangani, Alexander Mitchell, Antonela Tamagnini, Arsalan Saif, Tanish Kandivlikar*

1. Introduction

Cardiovascular diseases remain a leading cause of mortality worldwide, emphasizing the critical need for accurate and efficient diagnostic tools. Electrocardiograms (ECGs) provide a non-invasive and cost-effective means of assessing cardiac function, making them invaluable in the early detection and management of heart-related conditions. However, the interpretation of ECG recordings can be challenging, requiring extensive training and expertise, which can lead to potential errors or inconsistencies.

Recent advancements in machine learning and artificial intelligence have opened new methods for automated ECG analysis, promising improved accuracy, efficiency, and accessibility. In this context, our project aims to develop a robust classification model capable of identifying 18 distinct cardiac abnormalities from a diverse dataset comprising a subset of 43,100 12-lead ECG recordings.

By leveraging the power of machine learning algorithms and the wealth of available ECG data, we strive to create a reliable model that can assist healthcare professionals in making informed clinical decisions. Building upon the achievements of previous research efforts, our team is dedicated to further improving the performance and robustness of ECG classification models, ultimately contributing to enhanced patient care and better health outcomes.

2. Data collection

To obtain a diverse and comprehensive dataset, we gathered data from six distinct institutions located across various regions of the globe.

- Southeast University, China:

  https://www.kaggle.com/datasets/bjoernjostein/china-12lead-ecg-challenge-database

  https://www.kaggle.com/datasets/bjoernjostein/china-physiological-signal-challenge-in-2018

- St. Petersburg Institute of Cardiological Technics:

  https://www.kaggle.com/datasets/bjoernjostein/st-petersburg-incart-12lead- a
  rrhythmia-database

- The Physikalisch Technische Bundesanstalt (PTB):

  https://www.kaggle.com/datasets/bjoernjostein/ptb-diagnostic-ecg-database https://www.kaggle.com/datasets/bjoernjostein/ptbxl-electrocardiography- d atabase

- Georgia University:

  https://www.kaggle.com/datasets/bjoernjostein/georgia-12lead-ecg-challenge- database

3. Data Description

For each patient record, there are two accompanying files. The first file is a header type, containing details such as the patient's age, gender, and the diagnosis code assigned. The second file is a raw Matlab data file that captures the signal information from the ECG study. This file includes 12 simultaneously measured signals, which represent different viewpoints of the heart's electrical activity, capturing the movement of electrical impulses through the heart muscle from various angles:

- Lead I: Measures the electrical potential between the right arm and left arm. This lead looks at the lateral wall of the left ventricle.
- Lead II: Measures the electrical potential from the right arm to the left leg. This lead provides a view of the heart from the right shoulder to the left foot, which is useful for viewing the inferior wall of the heart.
- Lead III: Measures the electrical potential from the left arm to the left leg. This lead views the inferior wall between the left and right leg.
- Lead aVR: Focuses on the right atrium by measuring the electrical activity from the body's center to the right arm.
- Lead aVL: Measures the electrical activity from the center of the body to the left arm, focusing on the lateral wall of the left ventricle.
- Lead aVF: Measures the electrical activity from the center of the body to the feet, looking at the inferior wall of the heart.
- Leads V1 to V6 (Chest leads): These leads are placed in specific locations on the chest, around the heart:

V1 and V2 are positioned on the right and left side of the sternum, respectively, and mainly view the heart's septal wall. V3 and V4 transition between the septal and the anterior walls, with V4 directly over the anterior wall. V5 and V6 look at the lateral wall of the left ventricle, with V6 being more lateral than V5.

These 12 leads collectively provide a comprehensive 360-degree view of the heart's electrical activity, enabling clinicians to detect and localize disturbances in cardiac function. Each lead's unique perspective helps pinpoint where abnormal electrical activity starts and travels, which is vital for accurate diagnosis and treatment planning.

As the diagnosis of the patient is encoded in the header file, we used the international SNOMED CT (Systematized Nomenclature of Medicine -- Clinical Terms) unique codes from SNOMED International (https://www.snomed.org/), to decode the heart conditions.

To facilitate data analysis and model development, we converted and consolidated the datasets from various sources into a single CSV data frame (Table 1). This data frame includes information from the header files and Matlab files, and the decoded SNOMED CT values corresponding to the diagnosis codes. Our analysis focused on a random subset of 1,000 studies with valid diagnoses according to the International SNOMED CT record.

Table 1 presents the consolidated data frame. The first column represents the unique identifier for each report. Columns 2 through 13 contain the signal data extracted from the raw data files. The column containing the diagnosis code is derived from the patient's header file. Lastly, the columns displaying the diagnosis name and abbreviation are populated using the information from the SNOMED CT record.

ID	I	II	III	aVR	aVL	aVF	V1	V2	V3	V4	V5	V6	Diagnosis Code	Diagnosis Name	Abbreviation
A0001	-4.0	-23.0	-18.0	14.0	8.0	-20.0	56.0	99.0	41.0	1.0	-9.0	-20.0	426783006	sinus rhythm	SNR
A0001	2.0	-18.0	-20.0	8.0	10.0	-18.0	39.0	81.0	29.0	2.0	-17.0	-24.0	426783006	sinus rhythm	SNR
A0001	11.0	-19.0	-29.0	4.0	20.0	-24.0	35.0	79.0	21.0	3.0	-22.0	-30.0	426783006	sinus rhythm	SNR
A0001	8.0	-8.0	-16.0	0.0	12.0	-13.0	33.0	75.0	17.0	4.0	-17.0	-27.0	426783006	sinus rhythm	SNR
A0001	9.0	4.0	-5.0	-7.0	7.0	-1.0	32.0	78.0	18.0	5.0	-17.0	-27.0	426783006	sinus rhythm	SNR

Table 1, Data disposition

4. Methodology

First, we performed an in-depth Exploratory Data Analysis (EDA) to comprehend the dataset's characteristics, including the distribution of cardiac abnormalities, potential biases, and the quality of the ECG recordings. This step is crucial for gaining insights into the data and identifying any potential challenges or areas of concern.

Next, we preprocessed the ECG data to prepare it for the models. This process involved resampling the data to a consistent sampling rate, normalizing the signals to a common scale, and segmenting the ECG recordings into appropriate windows or segments. Preprocessing is essential for ensuring the data is in a suitable format for analysis by the deep learning algorithms.

An electrocardiogram (ECG) is a graphical representation of the electrical activity of the heart over time. The most common visualization of an ECG is a waveform plotted on a graph, as illustrated in Figure 1. Each distinct wave or deflection in the ECG waveform corresponds to a specific electrical event occurring during the cardiac cycle. By analyzing the characteristics of these waves, valuable insights into the heart's electrical functioning can be obtained.

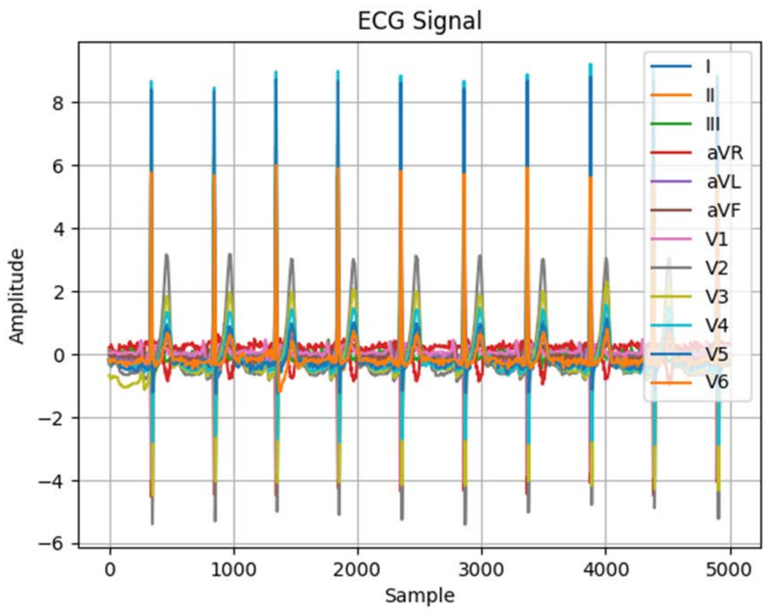

Figure 1, Example ECG signals

By meticulously analyzing the shape, duration, and timing of the waves in an ECG, medical professionals or, in our case, a deep learning model can assess various aspects of cardiac function. This analysis enables the evaluation of heart rate, rhythm, the presence of any anomalies, and ultimately, the diagnosis of potential heart conditions. These conditions may include arrhythmias, heart attacks, and heart failure. Thus, a careful understanding of ECG signals can provide powerful insights into the overall functionality and health of the heart.

To gain a comprehensive understanding of our dataset, we conducted an Exploratory Data Analysis (EDA). This analysis enabled us to examine the characteristics of the dataset, including the distribution of cardiac abnormalities, potential biases, and the quality of the ECG recordings. One of the key steps in this process was combining all the provided datasets into a single, universal dataset. Currently, our dataset comprises approximately 43,100 data points. Figure 2 showcases the individual ECG signals for a particular patient, providing a visual representation of the data we work with.

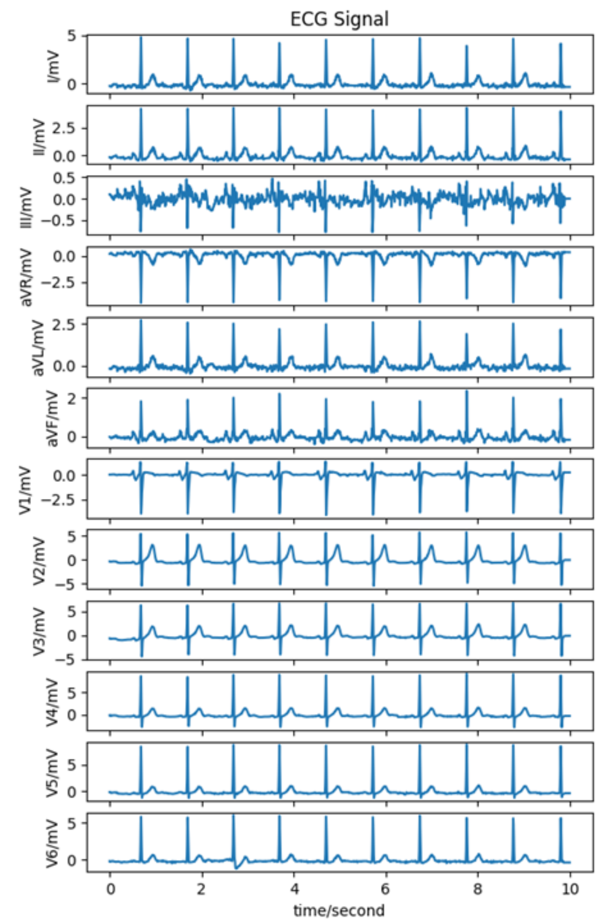

Figure 2, Individual ECG signals.

The dataset contains 85 different possible labels, representing a wide range of cardiac ailments. However, for our research, we focused on 18 specific labels that were the most representative and frequently occurring within the dataset. To prepare the ECG data for training our models, we performed several preprocessing steps, including filtering and normalization. The total size of the dataset is approximately 20GB when stored in CSV format.

Figure 3 illustrates the distribution of the most common diagnoses present in the dataset, providing insights into the prevalence of various cardiac conditions. Additionally, Figure 4 depicts the age distribution of the patients represented in the dataset. It is important to note that the dataset exhibits a significant skew toward an older demographic. This skew could potentially limit the model's applicability in a general population setting; however, the dataset remains highly valuable for studying cardiac conditions in individuals aged 40 and above, particularly those over 70 years old.

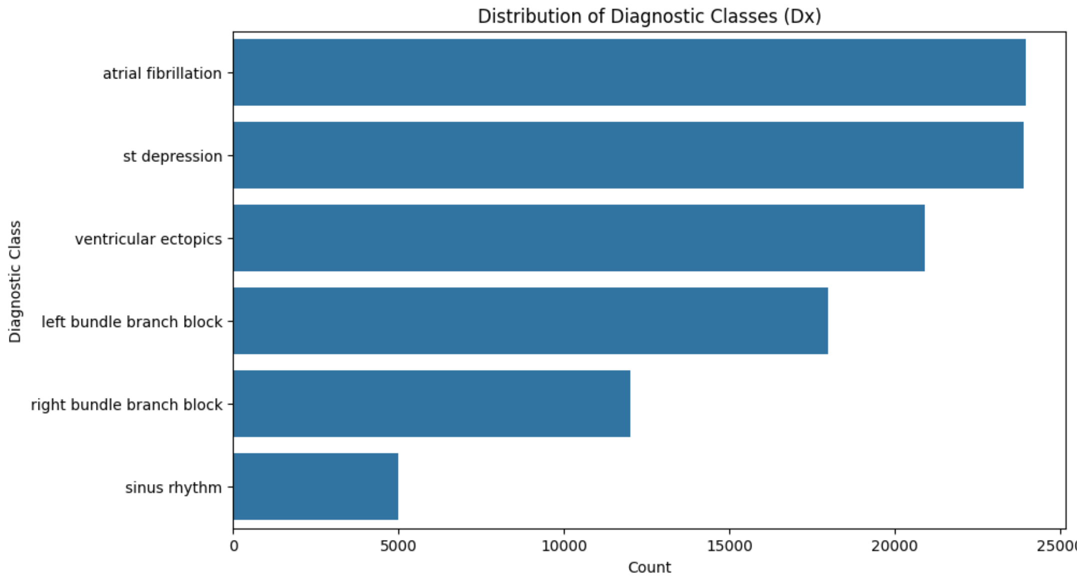

Figure 3, Diagnostic classes

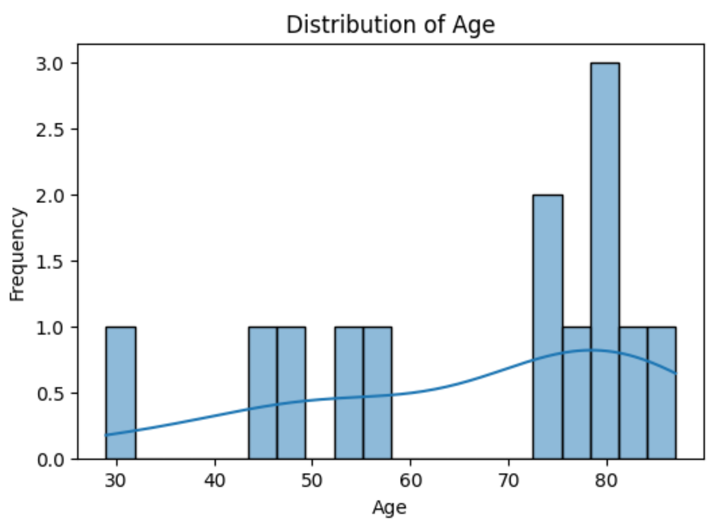

Figure 4, Distribution of Age

Figure 5 shows the cross correlation between the different ECG signals. This is very important for understanding how the different electrical leads interact with each other. The largest correlation is between leads V2 and V3, while the largest negative correlation is between leads aVR and II.

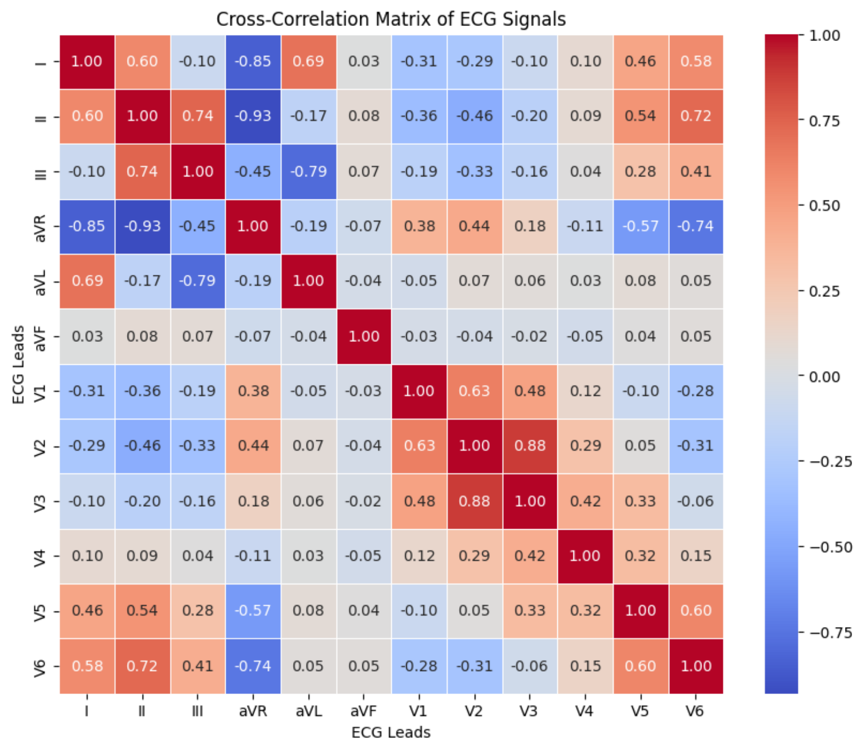

Figure 5, Cross-correlation matrix of ECG signals

Abnormal events can be hard to spot especially since they can appear frequently in a single signal. Figure 6 showcases different abnormal events appearing on just lead I. They primarily happen on the top or bottom of the amplitude of a heart movement. Figure 7 is another example showcase of a signal for a single patient.

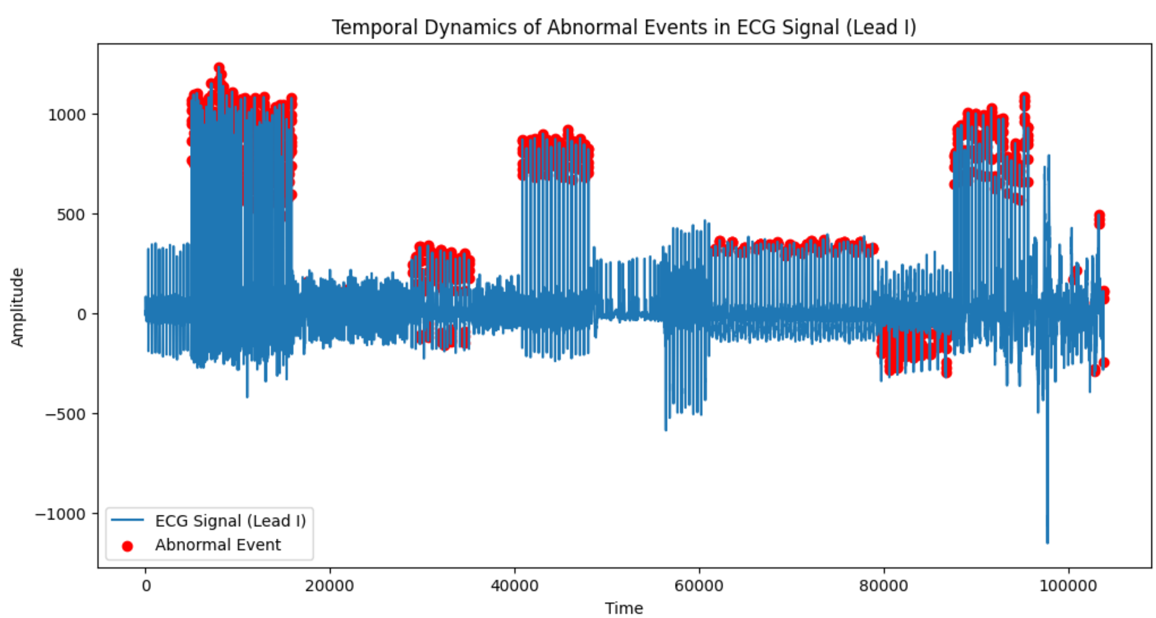

Figure 6, Abnormal events in ECG signal (lead 1)

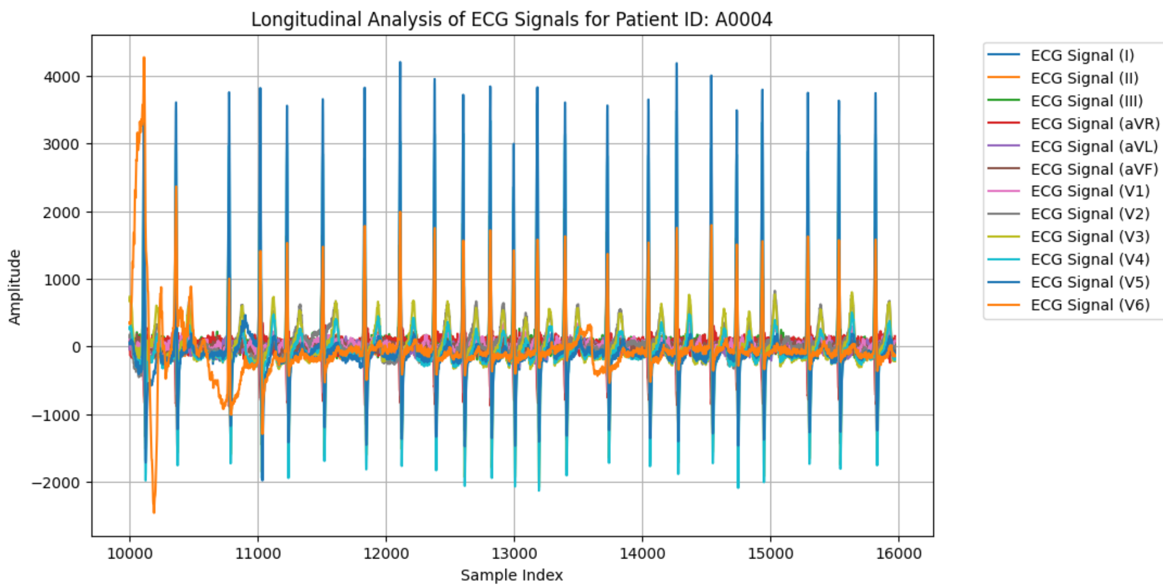

Figure 7, ECG for patient A0004

5. Model Training and Evaluation

Following are the Deep Learning models implemented for this task:

Artificial Neural Networks

It is a straightforward feedforward neural network that consists of a fully connected layer which takes the input data and transforms it using a set of weights and biases, followed by ReLU activation function that introduces non-linearity into the model, allowing it to learn more complex patterns in the data and finally a second fully connected layer maps the output of the hidden layer to the number of classes in the target variable.

ANNs are relatively straightforward to implement and train, making them a good starting point for binary or multi-class classification tasks. Despite their simplicity, ANNs can capture complex relationships in the data through layers of neurons, making them effective for a wide range of problems, including medical diagnosis from signals like ECG.

Test Accuracy (72.12%): This accuracy shows that the model correctly predicts the diagnostic class for approximately 72.12% of the test data. This is a decent starting point but highlights potential areas for improvement, either through more complex model architectures or more data.

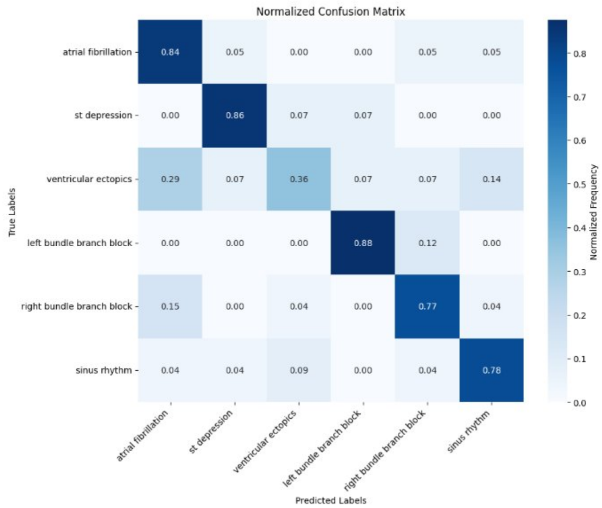

Figure 8, Confusion Matrix for ANN

Fully Connected Networks

It is a straightforward, yet slightly deeper architecture compared to the initial ANN, consisting of first fully connected layer that applies a linear transformation, followed by a ReLU activation function to introduce non-linearity to the model, allowing it to capture more complex patterns in the data, then a second fully connected layer for data processing and finally the output layer for classification.

FCNs are easily adaptable and scalable to different sizes of datasets and feature dimensions without needing significant architectural changes.

Test Accuracy (74.04%): This model correctly predicts the diagnostic category for 74.04% of the test set, which is a slight improvement over the initial ANN model. This suggests that adding more layers (depth) to the network helped capture more complex relationships in the data.

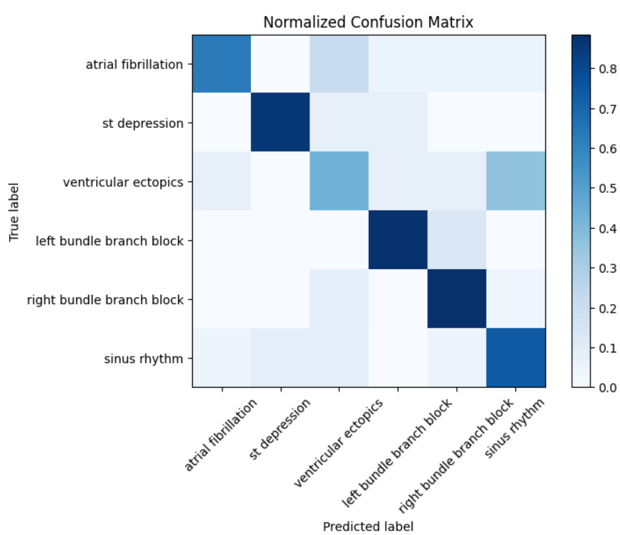

Figure 9, Confusion Matrix for FCN

Convolutional Neural Networks

This model is designed to exploit the spatial structure of ECG data, which is represented as time-series data. It contains 2 convolutional layers, the first one designed to capture short-term relationships in the data while the second one helps to further abstract and enhance features extracted by the first layer. Then the output of the last convolutional layer is flattened into a 1D vector to be fed into the 2 fully connected layers where the first one processes features extracted and flattened from the convolutional layers and the final output layer classifies the input into one of the diagnostic categories.

CNNs are highly effective at automatic feature extraction from spatial or temporal data. In the context of ECG signals, CNNs can identify important patterns in the data without explicit manual feature engineering.

Test Accuracy (75%): The CNN model achieves an accuracy of 75%, indicating it correctly classifies 75% of the test data. This is an improvement over the previous ANN and FCN models, suggesting that the convolutional layers are effective at extracting useful features from ECG data.

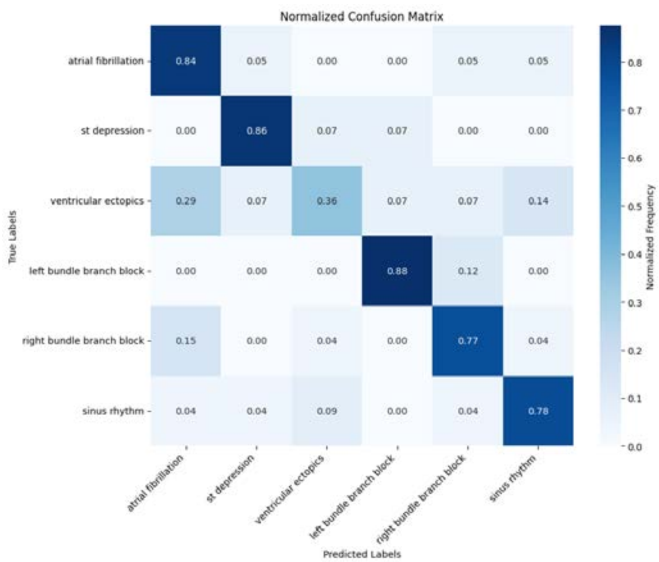

Figure 10, Confusion Matrix for CNN

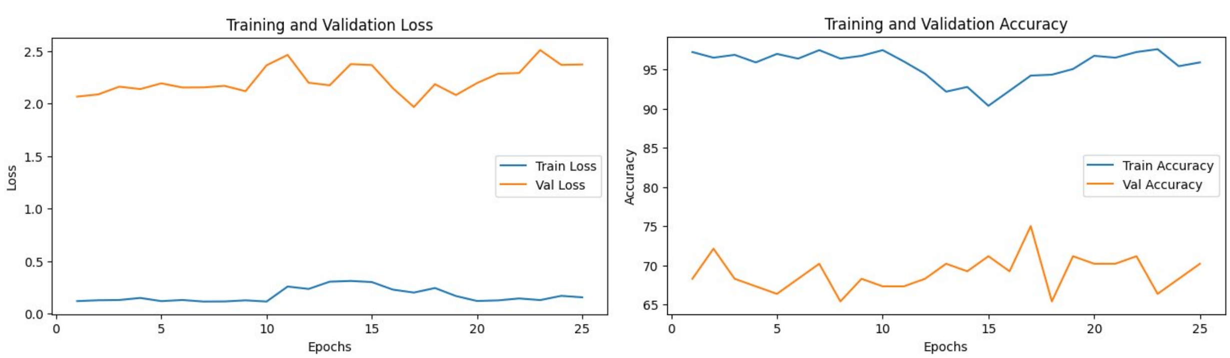

Figure 11, Training and Validation Accuracy and Loss Graphs for CNN

The training accuracy starts high and remains relatively stable throughout the training process, hovering around 95%. This high level of accuracy indicates that the model is effectively learning the training data and consistently applying what it has learned across each epoch.

The validation accuracy is significantly more variable and generally lower than the training accuracy, averaging around 70%. Such a difference typically points to overfitting, where the model learns the training data too well, including noise and specifics not applicable to unseen data.

The higher training accuracy compared to validation accuracy in the CNN model suggests overfitting, where the model learns the training data too well, including its noise and minor fluctuations, but fails to generalize effectively to unseen data due to the following reasons:

Insufficient regularization: Lack of methods like dropout, L2 regularization, or insufficient use of data augmentation which helps to prevent the model from becoming too fitted to the training data.

Complex model architecture: Too many layers or neurons could make the model overly complex for the amount of training data available, allowing it to memorize rather than generalize.

Recurrent Neural Networks

This model is particularly suited to handling sequence data, such as time-series ECG signals where the temporal relationship between data points is crucial, due to its ability to process sequences one element at a time while maintaining a memory (hidden state) of what has been processed.

The RNN Layer, which is the core of the model, processes each sequence in the input batch by batch, maintaining an internal state that captures information about the sequence processed so far. The hidden state is initialized to all zeros and this state is updated as each element of the input sequence is processed.

Test Accuracy (64.42%): The accuracy is lower than what was achieved with CNN and FCN models, which might suggest that the simple RNN is less effective at capturing the dependencies in the ECG data or might be suffering from issues like vanishing gradients, especially with longer sequences.

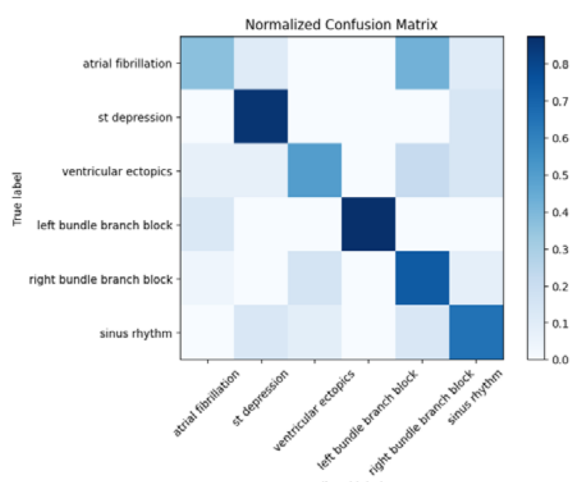

Figure 12, Confusion Matrix for RNN

Long Short-Term Memory

This model is an advanced type of RNN designed to avoid the long-term dependency problem typical of standard RNNs which makes it particularly effective for tasks involving sequences with important relationships at various time lags. It contains LSTM layer, the core of the model, that can maintain a long-term memory using a series of gates that regulate the flow of information and help the LSTM decide what to keep in or leave out of the cell state, allowing it to capture longer dependencies in data sequences. At the start of each sequence processing, the hidden and cell states are initialized to zeros, providing a clean slate for sequence processing after which the final hidden state is used to make a prediction.

Unlike traditional RNNs, LSTMs can maintain a stable gradient over many time steps, which helps during the training process, especially on longer sequences.

Test Accuracy (70.19%): This accuracy is a notable improvement over the simple RNN model and demonstrates the LSTM's ability to correctly classify a higher percentage of the test data. This suggests better handling of temporal dependencies in ECG signals.

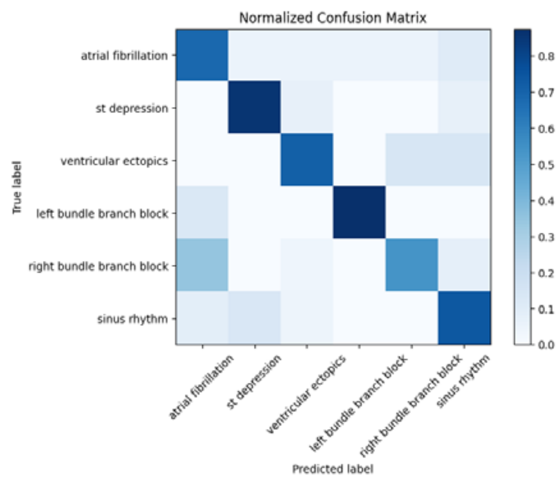

Figure 13, Confusion Matrix for LSTM

Transformers

This model focuses on using self-attention mechanisms to process sequences, offering several advantages over RNNs and LSTMs. It contains an embedding layer which is the initial layer for linear transformation for processing in the transformer encoder layer which processes the entire input sequence simultaneously, leveraging the self-attention mechanism to weigh the importance of different parts of the input data after which the output for the last token is passed through a fully connected layer to produce the final class predictions.

Transformers can capture dependencies between any two points in the sequence, regardless of their distance, thanks to the self-attention mechanism.

Test Accuracy (56.73%): This is lower than the results from LSTM and even some simpler models, which might suggest that the Transformer architecture as configured may not be fully suited or optimized for this particular type of sequence data without further tuning.

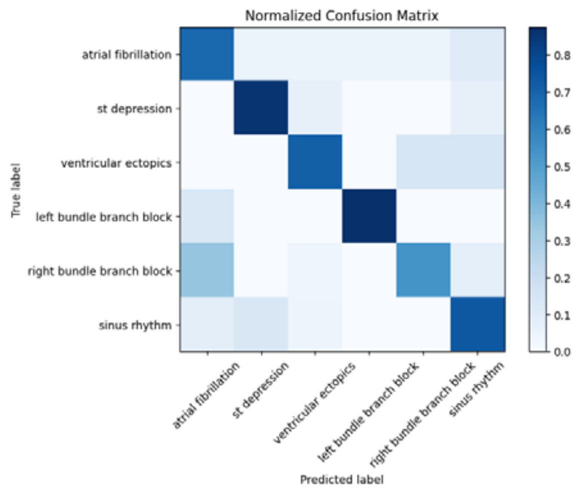

Figure 14, Confusion Matrix for TF

6. Comparison and Analysis

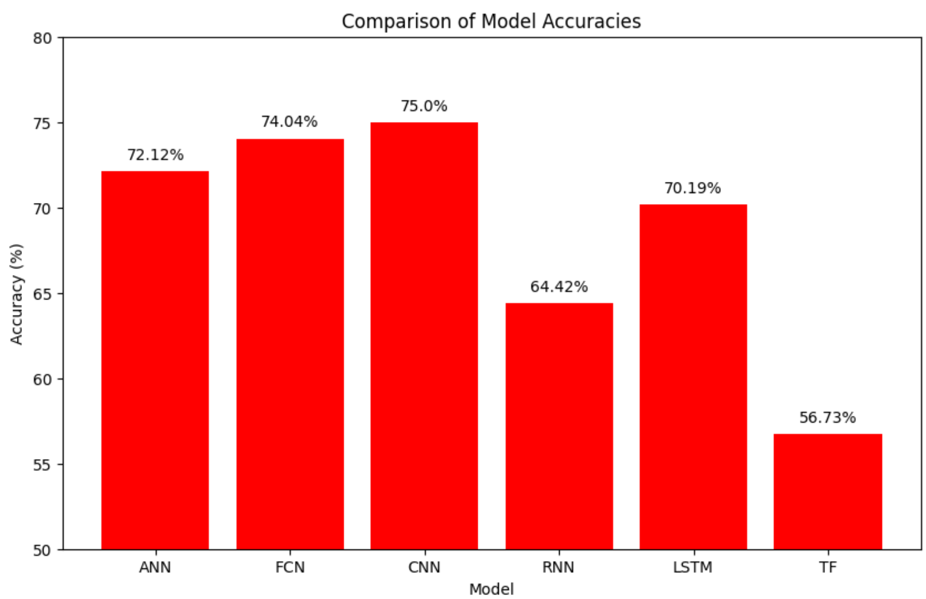

Figure 15, Comparison of Model Accuracies

Figure 14 shows the results of the models trained. The CNN model, which often performs exceptionally well on tasks involving spatial and temporal data due to several intrinsic characteristics of its architecture, yielded the highest accuracy due to the following reasons:

Effective Feature Extraction: CNNs are designed to automatically detect and utilize hierarchical patterns in data. For ECG signals, this means that CNNs can efficiently capture underlying features such as peaks, intervals, and amplitude variations without requiring manual feature engineering. These features are crucial for accurate ECG interpretation and subsequent classification.

Spatial and Temporal Relationships: While primarily known for their success in spatial data processing (like images), CNNs can also process temporal patterns by treating time as a spatial dimension. In the context of ECG data, which has inherent temporal properties, CNNs can apply their convolutional filters to learn from the temporal sequence of heartbeats, effectively capturing patterns that span across time steps.

Robustness to Shifts and Distortion: CNNs are inherently robust to local variations in the input data. In the case of ECG signals, this means that CNNs can recognize the same cardiac features (like QRS complexes) even if they are slightly shifted in time or distorted, which is common in real-world data due to differences in heart rate or minor noise interferences.

Layered Architecture: The layered structure of CNNs allows them to learn different levels of representation from the data. Lower layers might learn to detect simple edges (basic signal transitions in ECG), while deeper layers can abstract these features to detect more complex patterns such as the morphology of specific waveforms characteristic of certain heart conditions.

7. Conclusion

In this project, we developed and compared multiple deep learning models to classify ECG signals into various diagnostic categories. We employed a range of models, including ANNs, FCNs, CNNs, RNNs, LSTMs, and Transformers, to understand their performance in handling the complexities of ECG data.

Key Findings:

- Model Performance: The models generally achieved accuracies ranging between 70% to 75%, with the CNN model showing the best performance, highlighting its capability in capturing spatial and temporal patterns effectively.
- Overfitting Issues: The CNN model, despite its high accuracy, showed signs of overfitting as indicated by a discrepancy between training and validation accuracy. This suggests a need for better regularization techniques or model adjustments.
- Challenges with RNN and LSTM: While the RNN and LSTM models effectively captured temporal dependencies, they were slightly less accurate than the CNN, possibly due to the challenges in tuning and the tendency of RNNs to forget longer sequences.
- Transformers' Adaptability: The Transformer model, though innovative, did not outperform traditional models in this specific task, indicating that further tuning and exploration are necessary to leverage its full potential for time-series data.

Implications for Future Work:

- Data Augmentation and Regularization: To combat overfitting and improve model robustness, future efforts could focus on enhancing data preprocessing and implementing advanced regularization methods.

- Hybrid Models: Combining the strengths of CNNs and RNNs/LSTMs in a hybrid architecture could potentially offer better accuracy and robustness by capturing both spatial and temporal relationships more effectively.
- Expanding Data Diversity: Increasing the diversity and volume of the training data can help improve the models' generalization capabilities, crucial for medical diagnostic applications.
- Therefore, this project underscores the potential of deep learning techniques in medical diagnostics, particularly in automating and improving the accuracy of ECG signal classification. While challenges remain, the insights gained lay a solid foundation for further research and development in this critical area of healthcare technology.

8. References

[1] Goldberger AL, Amaral LAN, Glass L, Hausdorff JM, Ivanov PCh, Mark RG, Mietus JE, Moody GB, Peng CK, Stanley HE. PhysioBank, PhysioToolkit, and PhysioNet: Components of a New Research Resource for Complex Physiologic Signals. *Circulation* 101(23):e215-e220 [Circulation Electronic Pages; **http://circ.ahajournals.org/content/101/23/e215.full**]; 2000 (June 13). PMID: 10851218; doi: 10.1161/01.CIR.101.23.e215

[2] Perez Alday EA, Gu A, J Shah A, Robichaux C, Ian Wong AK, Liu C, Liu F, Bahrami Rad A, Elola A, Seyedi S, Li Q, Sharma A, Clifford GD* Reyna MA*. Classification of 12-lead ECGs: The PhysioNet/Computing in Cardiology Challenge 2020. Physiol. Meas. 2021 Jan 1;41(12):124003. doi: 10.1088/1361-6579/abc960