ECG-based Models for Heartbeat Classification

Akshat Saini
Group 15
2020019
IIIT Delhi
akshat20019@iiitd.ac.in

Nakul Thureja
Group 15
2020528
IIIT Delhi
nakul20528@iiitd.ac.in

Nishaant Rastogi
Group 15
2020091
IIIT Delhi
nishaant20091@iiitd.ac.in

Abstract

Electrocardiography (ECG) is a widely utilized method for diagnosing cardiac disorders and monitoring cardiac health[1]. Nonetheless, manual interpretation of ECG signals can be subjective, time-consuming and error-prone. Therefore, there is a growing interest in developing automated methods for ECG analysis[2]. In this study, we aim to classify ECG data into five superclasses: Normal, Myocardial Infarction, Conduction Disturbance, Hypertrophy, and ST/T-Change, using traditional machine learning and deep learning techniques. We extracted various features from the raw ECG signals using hrv-analysis and neurokit2 for traditional Machine Learning models. We have also utilised the raw data and applied Convolutional Neural Network (CNN) based Deep Learning models. Our findings are based on the PTB-XL dataset, which comprises a large and diverse collection of 21837 clinical 12-lead ECGs obtained from 18885 patients of 10 seconds length. Additionally, we provide insights into the strengths and limitations of each model and discuss the challenges and opportunities for future research on ECG classification.

Keywords: ECG signals, Time-Frequency Domain Analysis, P wave, QRS complex, T wave, Machine Learning (ML), Deep Learning (DL), Convolutional Neural Networks (CNN), Support Vector Machine (SVM), Transfer Learning, PTB-XL, Feature Extraction, Classification Models, Confusion Matrix, F1 Score.

1 Introduction

1.1 Background

The significance of maintaining a healthy heart cannot be overstated as it is vital for sustaining a healthy body and mitigating the risk of cardiovascular disease, stroke, and heart attack. The heart is interconnected with the overall well-being of an individual, and in recent times, there has been a surge in cardiovascular fatalities, according to the World Health Organization. Therefore, there is a need for dependable and efficient heartbeat classifications. The ECG is the most optimal and extensively used biosignal for identifying heart diseases, and it is a non-invasive approach. This signal is measured on the surface of the human body and comprises 12 leads that represent the different directions of cardiac activation in three-dimensional space.

The ECG signal represents a graph of the heart's electrical activity and can be leveraged to categorise heartbeats and identify irregularities. The P, QRS complex, and T waves together constitute the ECG signal and can be used as features for diagnosing a range of cardiovascular diseases like cardiac arrhythmia. And the P-QRS-T wave amplitude and duration parameters offer clinically significant insights into the underlying cardiac disease[3].

1.2 Literature Survey

Numerous studies by various researchers have proposed different machine learning algorithms for ECG heartbeat classification, with two important avenues of research in this domain. The first research point involves feature extraction, and the second involves the actual classification.

There have been some researchers that have used the ECG data without any feature extraction[4]. In contrast, other researchers, such as Karpagachelvi et al., have relied heavily upon feature extraction methods such as discrete wavelet transform (DWT)[5], and some researchers, like Aziz et al., have also researched fractional Fourier Transforms (frFT) algorithms for feature extraction[6]. Ramanujan Fourier Transform is another method used by researchers like Uthaman[7]. Another approach involves converting ECG waves into images for feature extraction. Jun et al.[8] put forward a deep two-dimensional convolution method to classify ECG data and converted every ECG beat into a two-dimensional grey-scale image as the input data of the classifier.

After this, we move towards literature, focusing more towards the classification task, The usage of CNN can be found in many research papers regarding the classification of ECG waves [8][9]. Some researchers like Rabee and Barhumi have also implemented SVMs[10] In addition, as a commonly used classification algorithm, KNN is also applied to classify ECG data. Saini et al.[11] used it as a classifier to detect QRS waves of ECG signals. DNN is also a direction that some researchers, such as Śmigiel et al.[4] have taken on this particular task.

1.3 Objectives

The objective of this study is to classify electrocardiogram (ECG) signals into distinct diagnostic classes, such as Myocardial Infarction, Conduction Disturbance, Hypertrophy and ST/T-Change, with subclasses to aid in accurate diagnostic labelling.

Additionally, we aim to evaluate the distribution of these diagnostic labels across different metadata features, such as age and sex, to provide insights into the prevalence of these classes among specific demographic groups.

We also endeavour to conduct a comparative analysis of the classifications derived after feature extraction from the time-domain, frequency-domain, and time-frequency domain of ECG signals.

1.4 Scope

Automated evaluation and diagnosis of cardiovascular diseases, particularly cardiac arrhythmias, through machine learning techniques, hold significant potential. Large volumes of ECG data can be utilised to train machine learning models to classify heartbeats accurately as healthy or unhealthy based on learned parameters. Such applications offer the potential to enable early detection or prediction of abnormal cardiac activity, allowing patients to take preventive measures before conditions worsen. Additionally, the use of such models can accelerate ECG evaluation by assisting clinicians in diagnosing and treating cardiovascular diseases.

1.5 Impact

This study presents the potential to have a significant impact on the field of cardiology as our study intends to develop a model that can help patients and doctors by providing real-time results and feedback. This feedback can potentially improve diagnostic accuracy and treatment efficiency[12], allowing for more timely interventions in the case of heart-related abnormalities. Our study can also be impactful in the future, as it may be able to potentially take advantage of the advances in wearable technology as our model could be integrated with the devices to enable non-stop monitoring of cardiac activity. This could be used to alert the patients and doctors of any concerning trends or changes. This type of continuous monitoring could be particularly beneficial for individuals with a history of heart disease or those with a higher risk of developing cardiac issues.

2 Materials and Methods

2.1 Dataset

The PTB-XL dataset, which is available on websites such as PhysioNet, is a large collection of 21801 clinical 12-lead electrocardiograms (ECGs) from 18869 patients[13] that we intend to use for this study. Each ECG record is 10 seconds long and has up to two cardiologists' annotations, who assigned potentially multiple ECG statements to each record. The annotations include 71 different ECG statements that follow the SCP-ECG standard. The dataset features ECG recordings with a sampling frequency of 500 Hz and downsampled recordings with a frequency of 100 Hz.

We chose the PTB-XL dataset over the MIT-BIH arrhythmia database due to its diverse range of signal quality, covering a wide age range with a balanced gender distribution and containing demographic information such as age, sex, height, weight, and race/ethnicity.

The data was collected between 1989 and 1996 at the Department of Cardiology of University Hospital Heidelberg and was released for public use in 2019 following a significant restructuring aimed at improving its usability and accessibility for the machine learning community. Overall, the PTB-XL dataset is a valuable and diverse resource that is suitable for a variety of research and development purposes in the field of ECG signal processing and related areas.

2.2 Exploratory Data Analysis

This report presents the key findings and observations from the exploratory data analysis (EDA) conducted on the dataset. The purpose of the EDA was to identify any issues with the data, understand its characteristics, and provide insights for the modelling stage. The analysis includes an examination of data distributions, identification of outliers, and exploration of variable relationships.

The first step in EDA involved checking the number of rows and columns, the names and data types of the columns, and the data distribution of each column.

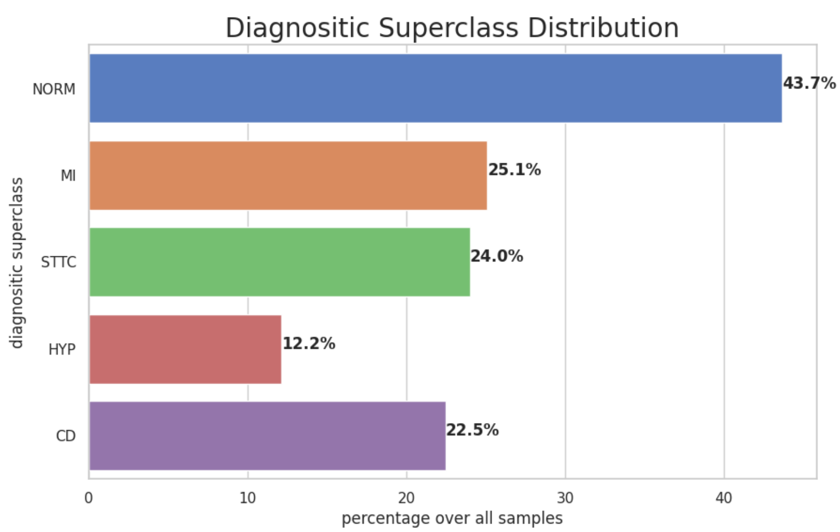


Figure 1: SuperClass label-wise distribution of PTB-XL dataset

We learned that the dataset contained 21801 entries for 12 leads spread across five fundamental superclass labels. These labels include:

- **NORM**: Normal ECG indicates a healthy heart with no signs of cardiac pathology.
- **MI**: Myocardial Infarction (heart attack) occurs when blood flow to a part of the heart is reduced or stopped, causing damage to the heart muscle.
- **STTC**: ST/T wave changes can indicate cardiac pathology or be a normal variant, and their interpretation depends on the clinical context and prior ECG findings.

- **CD**: Conduction Disturbance can cause irregular heartbeats (arrhythmias) and can be identified by abnormalities in the electrical impulses travelling through the heart.
- **HYP**: Hypertrophy is a condition in which the heart muscle becomes abnormally thick, making it harder for the heart to pump blood. Hypertrophic Cardiomyopathy (HCM) is a type of hypertrophy.

As can be seen from Figure 1, the data was found to be imbalanced with respect to the distribution of samples across different classes. Specifically, the NORM class had a higher number of entries compared to the other classes. To address this class imbalance issue, the Synthetic Minority Oversampling Technique (SMOTE) was employed. SMOTE is a widely used technique to generate synthetic samples for the minority class in imbalanced datasets. The method involves identifying the k-nearest neighbours of a minority sample and randomly selecting one of them to create a new synthetic sample on the line connecting them. In the current study, the SMOTE technique was implemented using the corresponding Python library.

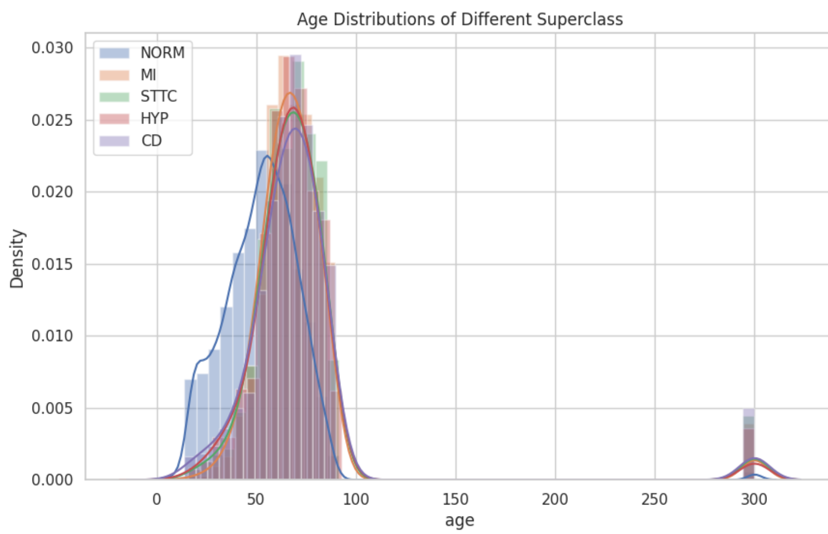
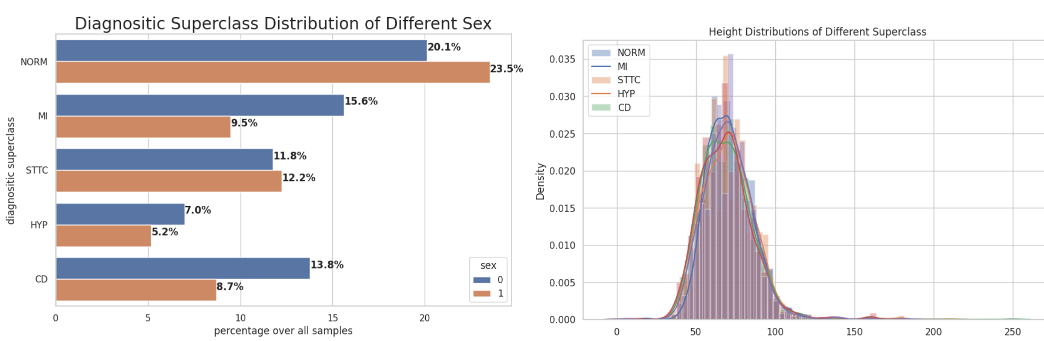


Figure 2: Age-wise distribution of PTB-XL dataset

From Figure 2, we can analyse the distribution of patients' age. The data can be seen to be distributed normally w.r.t. to each superclass, but we can see the NORM class has a lesser mean than other superclasses. This could potentially hint towards a relation that age has with heart diseases. Further, we can see that the dataset contained some outliers with respect to the age entries, and some signals were attributed to people aged over 300. These entries were discarded from the dataset.



a Sex-wise distribution of PTB-XL dataset

b Height-wise distribution of PTB-XL dataset

Figure 3: Age-wise distribution of PTB-XL dataset

Subsequent exploratory data analysis, as shown in Figures 3a and 3b, did not yield any further noteworthy findings. The distribution of the other variables in the dataset did not exhibit any remarkable patterns or trends, and there were no significant differences in the distributions of these variables across the different classes.

Also note that for our purposes, there were no missing values in the dataset.

Overall, the EDA was able to provide valuable insights into the data, and these findings were used to inform the next steps of the research.

2.3 Methodology

2.3.1 Feature Extraction

Feature extraction is a critical step in machine learning that involves selecting relevant features from raw data and transforming them into a suitable format. It helps to reduce the dimensionality of data, remove unwanted features, and identify new features that can improve the accuracy and efficiency of models.

Extracting R Peaks

We used the neurokit2 library to extract the R-peaks of the ECG signal. Neurokit2 is a Python toolbox for neurophysiological signal processing that provides easy access to advanced biosignal processing routines. To find R-peaks, neurokit2 uses the ECG-process method, which detects QRS complexes based on the steepness of the absolute gradient of the ECG signal. Subsequently, R-peaks are detected as local maxima in the QRS complexes[14]. The difference between these R-peaks was taken to obtain RR intervals.

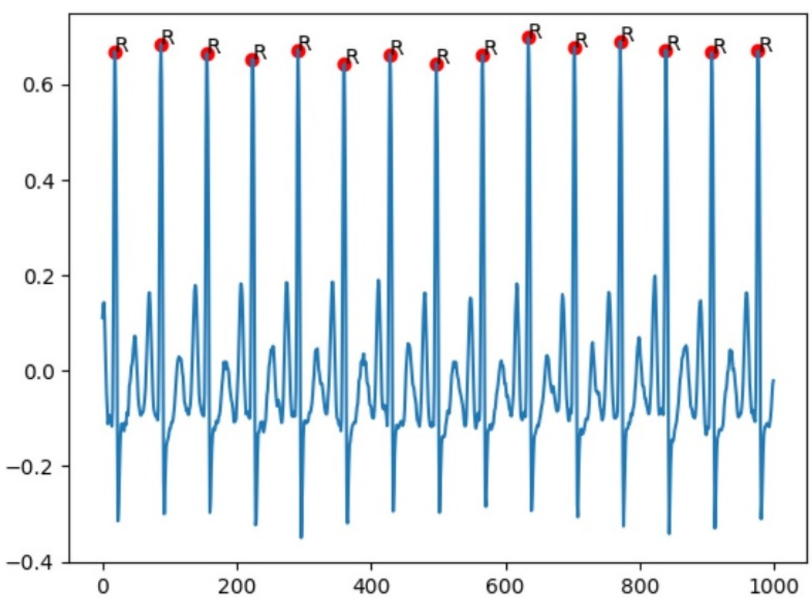


Figure 4: Extracted R Peaks from a raw ECG Signal

Heart Rate Variability Analysis

We used Python library functions to extract time-domain and frequency-domain features, respectively, from the RR intervals as calculated previously. The features extracted from HRV analysis can be used as inputs in ML models for predicting cardiovascular health outcomes, diagnosing diseases, and evaluating the effectiveness of interventions[15]. These features provide valuable insights into the autonomic nervous system activity and cardiac function, which can be used to improve the accuracy and performance of the models.

Finally, after feature extraction, we were able to bring down the feature tuple from the dimensions of (1000,12) to a mere (23).

2.3.2 Models

- **Classic ML Models**

  We have implemented and used several machine learning (ML) models, including **Random Forest, XGBoost, LGBM, and SVM**, to classify the ECG signals into different cardiac arrhythmias. Normally these models may not be ideal for datasets with a large number of features. However, after feature extraction from ECG signals, the dimensionality of the feature space is reduced, making these models somewhat suitable for ECG classification tasks.

  Random Forest is a popular ML algorithm that uses an ensemble of decision trees to classify data. It is robust to noise and outliers and can handle high-dimensional feature spaces, making it a suitable algorithm for our ECG classification task.

  XGBoost and **LGBM** are gradient-boosting algorithms that are widely used in ML for classification tasks. These algorithms use multiple weak classifiers to create a strong classifier that can accurately classify the data. XGBoost and LGBM are known for their fast and accurate performance and are suitable for large datasets with high-dimensional feature spaces, such as our ECG classification task.

  SVM is a classic ML algorithm that is commonly used for classification tasks. It works by finding the optimal hyperplane that separates the different classes in the feature space. SVM is suitable for high-dimensional feature spaces and is robust to noise and outliers.

  Although we used various ML algorithms for ECG classification, the overall results were not satisfactory (discussed in Section 3). This may be due to the complexity of the ECG signals and the limited discriminatory power of traditional ML algorithms. Therefore, we decided to employ DL models next.

- **DL Models**

  As the next step in our ECG classification project, we used various DL models, including Inception,**LSTM, GRU, AlexNet, and VGG16**, to classify the ECG signals into different cardiac arrhythmias. We also used an ensemble model that combined the predictions of VGG16, AlexNet, and LSTM models to improve the classification accuracy.

  Inception is a DL architecture that uses multiple layers of convolutional and pooling layers to extract features from the input data. This architecture is usually suitable for image classification tasks, however, we have modified the input layer to accommodate raw ECG signal data and kept the same architecture for the rest of the layers.

  LSTM and **GRU** are DL architectures that are suitable for time-series data analysis, such as ECG signals. These architectures can capture the temporal dependencies in the data and are robust to noise and variations in the signal. We used LSTM and GRU to classify the ECG signals by processing them as sequences of data.

  AlexNet and **VGG16** are DL architectures that were originally designed for image classification tasks. However, they can also be used for ECG classification by converting the input layer to accommodate raw ECG signal data and keeping the same architecture for the rest of the layers. These architectures are widely used in DL-based ECG classification studies and have shown promising results.

  Finally, we used an **Ensemble Model** that combined the predictions of **VGG16, AlexNet, and LSTM** models to improve the classification accuracy. The ensemble model is a popular technique in DL that combines the strengths of multiple models to achieve better performance than any single model alone. By combining different DL architectures with different strengths and weaknesses, we aimed to improve the classification accuracy of our ECG classification task.

The DL models used in our ECG classification project achieved higher classification accuracy compared to the traditional ML models, indicating their potential usefulness in the field of ECG analysis and cardiac arrhythmia diagnosis. The results are discussed in length under Section 3.

- **Transfer Learning Models**

  There is a lack of signal-based deep learning models; however, there is an abundance of very successful image-based deep learning models that have been developed for image classification tasks, such as VGG16, Inception, and EfficientNet.

  Therefore in order to benefit from these trained models, we plotted the ECG signals using Einthoven's triangle on lead 1, lead 2, and lead 3 to create a single graph for each signal data. Einthoven's triangle is a representation of the three bipolar limb leads of an ECG that allows for the measurement of the electrical activity of the heart from three different perspectives.

  By plotting the ECG signals in this way, we were able to create a visual representation of the electrical activity of the heart and identify patterns and anomalies in the signals that may be indicative of specific cardiac arrhythmias. This allowed us to understand the underlying mechanisms of the heart better and provided additional insights into the classification of ECG signals.

  After converting the ECG signals into images, we were able to leverage the power of image-based DL models, specifically VGG16 and EfficientNet. Dense layers were further added to the respective architectures to get the desired outputs. By adding layers at the end of pre-trained models, we were able to leverage the learned features from the pre-trained layers while also customizing the model architecture to fit the characteristics of the ECG signals better.

  The transfer learning models were not able to perform as well as raw DL models, probably because of the loss of features during image conversion.

2.3.3 Hyper parameter Tuning

We performed hyperparameter tuning to optimize the performance of our DL models. We used various hyperparameters such as learning rate, number of epochs, batch size, number of layers, number of filters, kernel size, activation function, and dropout rate. We employed grid search and random search methods to try different combinations of values for these hyperparameters. Our implementation of hyperparameter tuning resulted in improved accuracy and better results for the same architectures.

2.4 Novelty

The novelty in our research project is the extensive comparison of ML models on extracted features, DL models on raw data, and transfer learning after converting ECG data into images, all in a single study. This comparison allowed us to determine the most effective approach for analyzing ECG data.

In addition, we utilized an ensemble technique that combined three DL models: VGG16, AlexNet, and LSTM, which has not been explored much before. Through this approach, we were able to achieve comparable results to the state-of-the-art methods.

We believe this new approach has significant implications for improving the accuracy of ECG analysis and showcases the potential of combining different techniques to enhance the performance of biomedical data analysis.

2.5 Application

This research project has significant applications in the field of biomedical data analysis, particularly in the diagnosis and treatment of various cardiovascular diseases. By determining the most effective approach for analyzing ECG data through the extensive comparison of ML models and DL models, this project could lead to improved patient outcomes.

The overall application of this study in the field of cardiology is significant, as the developed model can potentially improve diagnostic accuracy and treatment efficiency. Providing real-time results and feedback can allow for timely interventions in the case of heart-related abnormalities, which can be crucial in improving patient outcomes.

2.6 Evaluation metrics

We will be using these evaluation metrics used for the heart-beat classification problems:

Area Under the Curve (AUC): measures of the classifier's ability to distinguish between positive and negative classes.

Accuracy: measures the proportion of correct predictions out of total predictions, giving an overall assessment of the model's performance.

Precision, Recall, and F1-score: adapted for multi-class problems by computing them for each class and then averaging them using techniques like micro-averaging and macro-averaging.

Confusion Matrix: shows the number of true positives, true negatives, false positives, and false negatives for each class, providing insight into the types of errors the model is making.

These metrics provide a comprehensive understanding of the model's performance across all classes, highlighting areas where the model may struggle to predict certain classes accurately.

3 Results

3.1 Failure of Machine Learning Models

We performed feature extraction using various different libraries like neurokit2, hrv-analysis and ecg-rhthyms that focus on extracting crucial information from R-peaks and help in reducing the dimentionality of the dataset being given to the Machine Learning Model.

The results for the performance of Machine Learning Model after feature extraction using neurokit2 followed by performing hrv-analysis on the R-intervals for Lead 0 of the PTB-XL dataset. This feature extraction technique reduced the dimensions of each record from (1000, 12) to (23,). These 23 features include time domain features and frequency domain features. These extracted features were passed to the below Machine Learning Models.

When evaluating the performance of machine learning models, it is important to consider metrics such as F1 score and Recall, in addition to test accuracy. Test accuracy measures the overall accuracy of the model in predicting the correct class for a given input, while F1 score and Recall measure the accuracy of the model in predicting specific classes[19].

In cases where a model has a high test and train accuracy but a low F1 score and Recall, it often means that the model is performing well in predicting the majority class, but struggling to predict the minority class correctly. This is particularly common in imbalanced datasets, where one class has significantly more samples than the other. We had used the imblearn library to correct the imbalance and still, the poor performance suggests that it can be due to the feature extraction performance on large datasets.

The performance of classical ML models with feature extraction in ECG signal classification is often limited due to their reliance on handcrafted features, which may not capture all the relevant information from the signals.

Our experimental results have also indicated that these models may suffer from high dimensionality and redundancy of the features, which can affect their performance and generalization. Furthermore, these models are often based on domain knowledge and expert rules for feature design, which may not be able to adapt to the diversity of ECG signals.

Therefore, alternative approaches such as DL models that can learn more complex and abstract features directly from the raw ECG signals have been proposed and have achieved state-of-the-art performance on various ECG classification tasks.

Table 1: Machine Learning Model Performance after Feature Extraction

Model	Super Class	AUC	Precision	Recall	micro F1	Accuracy
Random Forest						
	NORM	0.648	0.541	0.239	0.331	0.596
	MI/C	0.534	0.140	0.035	0.055	0.825
	STTC	0.533	0.105	0.031	0.048	0.861
	CD	0.528	0.259	0.074	0.116	0.775
	HYP	0.514	0.191	0.056	0.086	0.856
XGBoost						
	NORM	0.66	0.584	0.216	0.315	0.607
	MI/C	0.50	0.114	0.013	0.023	0.839
	STTC	0.56	0.191	0.035	0.060	0.875
	CD	0.52	0.293	0.034	0.061	0.793
	HYP	0.53	0.182	0.027	0.047	0.867
LGBM						
	NORM	0.66	0.584	0.216	0.315	0.607
	MI/C	0.50	0.114	0.013	0.023	0.839
	STTC	0.56	0.191	0.035	0.060	0.875
	CD	0.52	0.293	0.034	0.061	0.793
	HYP	0.53	0.182	0.027	0.047	0.867
SVM						
	NORM	0.49	1.0	0.0	0.0	0.573
	MI/C	0.53	1.0	0.0	0.0	0.855
	STTC	0.58	1.0	0.0	0.0	0.893
	CD	0.54	1.0	0.0	0.0	0.796
	HYP	0.54	1.0	0.0	0.0	0.882

3.2 Success of Deep Learning Model

After realising that the feature extraction done using hrv-analysis is not giving good results on classical machine learning models, we decided to move to deep learning models like Long Short-Term Memory(LSTM) networks, Gated Recurrent Unit (GRU) and VGG16. We had to modify some of these architectures for the PTB-XL datasets and thus, we utilised modified implementations[16] for these architectures.

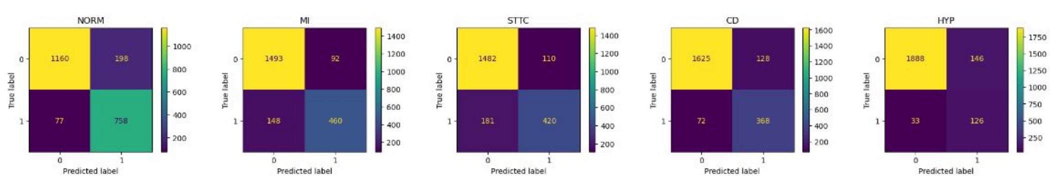


Figure 5: Confusion Matrix for AlexNet

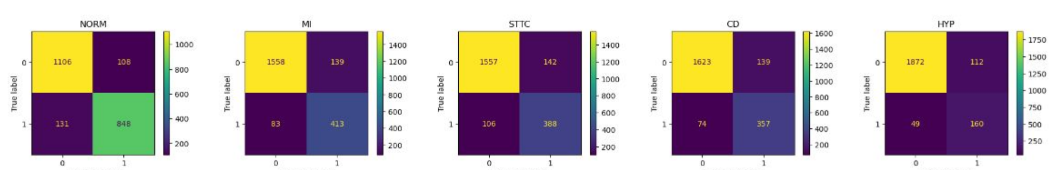


Figure 6: Confusion Matrix for LSTM

We then tried various combinations for ensembling and found the best results with the VGG16, AlexNet and LSTM combination with Area Under Curve(AUC) = 0.948.

Deep learning models are performing better due to their ability to learn features directly from raw signals without requiring manual feature engineering. The architecture of these models comprises

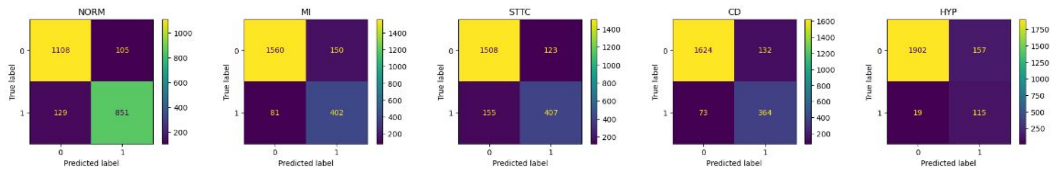


Figure 7: Confusion Matrix for VGG16

Table 2: Deep Learning Model Performance based on Evaluation Metrics

Model	**params**	**AUC**	**Precision**	**Recall**	**micro F1**	**Accuracy**
m-Inception	112,581	0.9403	0.63	0.68	0.7285	75
LSTM	204,421	0.9385	0.63	0.68	0.7353	75
GRU	208,645	0.9318	0.61	0.65	0.7229	73
m-AlexNet	1,172,837	0.9381	0.63	0.67	0.7225	75
m-VGG16	3,880,325	0.9368	0.61	0.62	0.7333	74
Ensemble(VGG16,AlexNet,LSTM)	-	0.94813	0.766	0.766	0.766	0.766

convolutional layers that capture local and global patterns from the ECG signals and fully connected layers that perform classification. [17]

The advantage of these models is that they can learn complex and abstract features from the signals, which can be more discriminative and robust than handcrafted features. Additionally, CNNs can take advantage of large amounts of data and achieve state-of-the-art performance on various ECG classification tasks.

3.3 Transfer Learning

With an average AUC of around 0.5 transfer learning's effectiveness was limited in our situation.

One potential reason for this is that the pre-trained model may not have learned the appropriate features necessary for the new task. In these instances, fine-tuning the pre-trained model with data specific to the new task may be necessary to improve performance.

Another factor that may impact the success of transfer learning is the resolution of the data. Higher-resolution data can provide more detailed information, leading to improved model accuracy. Therefore, it is crucial to carefully evaluate the relevance of the pre-trained model and the quality of the data when using transfer learning techniques.

Research suggests that fine-tuning pre-trained models with task-specific data can significantly improve performance, especially when compared to training a new model from scratch. However, the success of this approach depends on the quality and relevance of the pre-trained model and the task-specific data.

4 Conclusion

In conclusion, the Ensemble Model performs best in classifying the ECG signals into the 5 super-classes, as we have combined the VGG16, AlexNet and LSTM. The advantages of these models lie in their ability to learn and extract relevant features from input data, whether it is images or sequential data. These architectures have shown great success in their respective domains and have been widely used in various applications.

Ensembling these models together overcomes the shortcomings of each other but also promote the performance of classification task. As VGG16 can learn discriminative features from the raw ECG signals, without requiring manual feature engineering, which makes it suitable for automated ECG classification, while LSTM can capture the long-term dependencies and temporal patterns in the ECG signals, which may be crucial for accurate diagnosis of cardiac diseases[18].

The ML models' performance is often limited due to the reliance on handcrafted features. These handcrafted features are based on domain knowledge and expert rules, which may not capture all the relevant information from the signals. As a result, these models may miss subtle patterns or abnormalities in the ECG signals that can be critical for diagnosis. Furthermore, the process of designing handcrafted features can be time-consuming and requires significant domain knowledge. Therefore, the performance of these models can be limited by the low domain knowledge of the practitioners.

While performing the transfer learning, we had to first convert all the 12 lead ECG signals to images, and due to the large dataset, we had to compress these images drastically. This resulted in poor performance from VGG16 model that takes image data as input in contrast to our custom VGG16 architechture. We also performed transfer learning with the EffecientNetB0 model, which also gave similar results.

Thus, we can conclude that our Ensemble Model performs best with an AUC score of 0.94813.

5 Limitations and Future Research

To address the limitations of the results obtained in our study, there are several strategies that can be implemented in our current work. First, we can explore more advanced deep-learning models that are capable of more sophisticated feature extraction. This could involve experimenting with different architectures, such as convolutional neural networks (CNNs) or recurrent neural networks (RNNs), to better capture the nuances of complex data.

Secondly, some domain knowledge was lacking for feature extraction which can be easily overcome with the help of a domain expert to create better features for the classical ML models.

Lastly, we can explore ways to improve transfer learning techniques, such as incorporating higher-resolution images, to improve their ability to generalize to new domains. By implementing these strategies in our research projects, we can contribute to advancing the state-of-the-art in machine learning and overcome the current limitations in the field.

6 Distribution of Work

The distribution of work for this research project on ECG classification was organized as follows. The exploratory data analysis (EDA) stage was performed by Akshat Saini, Nakul Thureja, and Nishaant Rastogi. Specifically, they examined the dataset, conducted data preprocessing, and gained insights into the characteristics and patterns present in the ECG signals.

Feature extraction using NeuroKit and HRV-Analysis was primarily carried out by Nakul Thureja and Nishaant Rastogi. They utilized these tools to extract relevant features from the ECG signals, focusing on parameters related to heart rate variability. Akshat Saini took the responsibility of feature extraction using R-peaks.

The subsequent stages involved running classical machine learning models such as SVM, Random Forests,etc which was handled by Nakul Thureja. Nishaant Rastogi was responsible for running deep learning models, employing techniques such as convolutional neural networks (CNNs) to classify the ECG signals effectively.

Furthermore, Akshat Saini took charge of running transfer learning models, leveraging pre-trained models to enhance classification performance. Lastly, the collective effort of Akshat Saini, Nakul Thureja, and Nishaant Rastogi was focused on analyzing the results, making observations, and drawing meaningful inferences from the experiments conducted.

7 Code Availability

The code for loading the dataset, building the models and training the models can be found at 'https://github.com/Nakul-Thureja/Machine-Learning-Project'

The datasets, pickle files and h5 files have been uploaded on Kaggle for easy access and can be found at 'https://www.kaggle.com/datasets/nakulthureja/ml-project-models'

8 References

[1] National Health Service. "Electrocardiogram (ECG)." NHS, 2022, https://www.nhs.uk/conditions/electrocardiogram/

[2] Bertsimas, D., Mingardi, L., Stellato, B. (2021). Machine Learning for Real-Time Heart Disease Prediction. IEEE journal of biomedical and health informatics, 25(9): 3627–3637.

[3] National Center for Biotechnology Information. "Electrocardiogram (ECG)." StatPearls [Internet]., U.S. National Library of Medicine, 9 Nov. 2021.

[4] Śmigiel S, Pałczyński K, Ledziński D. ECG Signal Classification Using Deep Learning Techniques Based on the PTB-XL Dataset. Entropy (Basel). 2021 Aug 28;23(9):1121

[5] Karpagachelvi, S. et al. "ECG Feature Extraction Techniques - A Survey Approach." ArXiv abs/1005.0957 (2010): n. Page.

[6] Aziz, S., Ahmed, S. Alouini, MS. ECG-based machine-learning algorithms for heartbeat classification. Sci Rep 11, 18738 (2021).

[7] Uthaman, Neena Prasanth, C. (2022). Time Frequency Analysis of ECG Signal Using Ramanujan Fourier Transform Mrs Neena Uthaman. International Journal of Scientific Research.

[8] Zhang, Dengqing Chen, Yuxuan Chen, Yunyi Ye, Shengyi Cai, Wenyu Chen, Ming. (2021). An ECG Heartbeat Classification Method Based on Deep Convolutional Neural Network. Journal of Healthcare Engineering. 2021.

[9] Deep Learning Based Cardiac Arrhythmia Classification using Multi-Modal ECG Signals. IEEE Access, vol. 9, 2021, pp. 60133-60143.

[10] A. Rabee and I. Barhumi, "ECG signal classification using support vector machine based on wavelet multiresolution analysis," 2012 11th International Conference on Information Science, Signal Processing and their Applications (ISSPA), Montreal, QC, Canada, 2012, pp. 1319-1323.

[11] Saini I, Singh D, Khosla A. QRS detection using K-Nearest Neighbor algorithm (KNN) and evaluation on standard ECG databases. J Adv Res. 2013 Jul;4(4):331-44.

[12] Azman, Mohamed Prakash, Sai Kompalli, Venkata Sai Kiran Nandu, Eslavath. (2020). Real-Time Detection and Prediction of Heart Diseases from ECG Data Using Neural Networks.

[13] Wagner, P., Strodthoff, N., Bousseljot, R., Samek, W., Schaeffter, T. (2022). PTB-XL, a large publicly available electrocardiography dataset (version 1.0.3). PhysioNet.

[14] M. Ciocîrlan and A. Udrea, "Classification and Feature Extraction of Biological Signals Using Machine Learning Techniques," 2022 8th International Conference on Control, Decision and Information Technologies (CoDIT), Istanbul, Turkey, 2022, pp. 780-784

[15] Ishaque, S., Khan, N. M., Krishnan, S. (2021). Trends in Heart-Rate Variability Signal Analysis. Frontiers in Digital Health, 3. https://doi.org/10.3389/fdgth.2021.639444

[16] Razinw. (2021). Deep learning. Kaggle. https://www.kaggle.com/code/razinw/deep-learning

[17] Zhang, J., Gajjala, S., Agrawal, P., Tison, G. H., Hallock, L. A., Beussink-Nelson, L., ... Narayan, S. M. (2018). Fully automated echocardiogram interpretation in clinical practice: feasibility and diagnostic accuracy. Circulation, 138(16), 1623-1635

[18] Saadatnejad, S., Oveisi, M., Hashemi, M. (2020). LSTM-Based ECG Classification for Continuous Monitoring on Personal Wearable Devices. IEEE journal of biomedical and health informatics, 24(2), 515–523. https://doi.org/10.1109/JBHI.2019.2911367

[19] Davis, J., Goadrich, M. (2006). The relationship between Precision-Recall and ROC curves. Proceedings of the 23rd international conference on Machine learning, 233-240.