SCHOOL OF SCIENCE & TECHNOLOGY

A Comparative Study on Machine Learning Models for Multi-class classification of Cardiac Arrhythmia

By

Gokila Sundaram

(N1082709)

Supervisor: Dr.Temitope Alade

Word Count: 11530 (Excluding References, Abstract, Indexes)

Project Report submitted in partial fulfilment of the requirements of Nottingham Trent University for the degree of MSc Data Science.

Student Signature: Gokila Sundaram Date: *01-Sep-2023*

Abstract:

Arrhythmia, a common cardiovascular disorder characterized by irregular heart rhythms, presents a significant health challenge globally. Accurate and timely classification of arrhythmia patterns is essential for effective patient management and treatment. In this research paper, we conduct a comprehensive comparative study focused on multi-class classification of arrhythmia utilizing a range of feature selection techniques in conjunction with various classification algorithms.

A diverse set of feature selection techniques are employed, including SelectKbest, Principal Component Analysis (PCA), Lasso Regression & Recursive Feature Elimination (RFE) Method. Subsequently, we evaluate the performance of popular classification algorithms such as Support Vector Machines (SVM), Random Forest, K-Nearest Neighbors (KNN), and LightGBM(Gradient Boosting machine), using the selected feature subsets. The experimental results are presented in a comparative framework, showcasing the effectiveness of different feature selection techniques in enhancing classification accuracy, precision, recall, and F1-score for each arrhythmia class.

Our findings indicate that RFE (Wrapper Technique) feature selection method is particularly well-suited for High dimensional dataset, while the Light classification model shows excellent performance in handling imbalanced datasets. This combination yielded the highest accuracy of 89% when employing a 90-10 data split.

Acknowledgement

I would like to express my sincere gratitude to my primary advisor, Dr.Temitope Alade, for his invaluable guidance, insightful feedback, and unwavering support throughout the entire research process. The study's orientation has been significantly influenced by his profound knowledge and unwavering commitment. I extend my appreciation to the "UC Irvine Machine Learning Repository" for generously providing the dataset essential for conducting this study.

Table of Contents

List of Figures:

List of Tables:

List of Equation:

CHAPTER 1:

INTRODUCTION

Electrocardiogram (ECG) signals are electrical signals that represent the activity of the heart over time (Merdjanovska, Rashkovska 2022). It's typically measured using electrodes placed on the skin of the chest, arms, and legs (Banaee et al. 2013). The signal is then amplified, filtered, and recorded for further analysis. The ECG signal consists of a series of waves and intervals that correspond to the different phases of the cardiac cycle (Kusumoto 2020).

There is a need for automated methods to analyse ECG signals. It can be analysed using various techniques, such as signal processing, pattern recognition, and machine learning, to extract information about the function of the heart and to diagnose various cardiac conditions (He et al. 2023).

ECG analysis can help diagnose various cardiac conditions, such as arrhythmias, Infraction, and heart block (Nesaragi et al. 2022). It's also used to monitor changes in cardiac function over time, such as during exercise or in response to medication.

Use of Machine Learning (ML) models in ECG analysis is a critical topic in the research area. Numerous research studies have experimented several ML models to detect anomalies in ECG signals (Ardeti et al. 2023). Accuracy of ML algorithm is crucial element of study, as these algorithms are often used for diagnostic purposes and can have a direct impact on patient care.

1.1 Background (or) Research Context:

In clinical settings, ECG signals are typically collected using electrocardiogram machines, which are specialized devices that use multiple electrodes to record the electrical activity of the heart (Kusumoto 2020). These machines can produce high-quality ECG signals and are commonly used for diagnostic purposes (Abubaker 2023). ECG signals can also be collected using wearable devices such as smartwatches and fitness trackers (Appelboom et al. 2014).

These devices typically use one or two electrodes to record the ECG signal and are designed for continuous monitoring of heart activity outside of clinical settings (Banaee et al. 2013). While the quality of the ECG signals collected by these devices may not be as high as those collected by electrocardiogram machines, they can still provide valuable information for monitoring heart health and detecting abnormal heart rhythms (Lown et al. 2020).

Anomaly detection is a technique used in data analysis to identify patterns in data that deviate significantly from the expected or "normal" behaviour (Sunny et al. 2022). These patterns are often referred to as outliers or anomalies. The process of anomaly detection involves first establishing a baseline of what is considered "normal" behaviour or data, and then monitoring new data to identify any deviations from this baseline (Liu et al. 2020). By employing ML models, signals can be classified as either normal or abnormal (Masud Shah Jahan et al. 2022).

Common Issues & Challenges of ECG Data Analysis are listed below:

- Analysing ECG signals manually are time-consuming, especially when dealing with large amounts of data (He et al. 2023).
- Human error is inherent in manual analysis, and it can lead to misinterpretation or overlooking subtle abnormalities in the ECG data (Ohn et al. 2019).
- Even among experienced healthcare professionals, there can be variability in the interpretation of ECG data. Different observers may have different opinions on what constitutes an unusual pattern, leading to inconsistency in predictions (Serhani et al. 2020).
- Manual analysis of ECG data does not lend itself well to real-time monitoring or immediate identification of unusual patterns (Lyon et al. 2018).
- Understanding the interrelation between different ECG leads can be challenging (Ohn et al. 2019).
- Interpreting 12-lead ECGs can be complex, leading to negative perceptions and attitudes towards ECG learning (Ohn et al. 2019).
- Since ECG signals can be collected using different devices, variations in the signal quality, recording parameters, Data types, Storage patterns and noise levels that can impact the accuracy of Machine learning models (Banaee et al. 2013).
- ECG signals are inherently high-dimensional due to their continuous nature and typically consist of multiple channels or leads (Jha, Kolekar 2020).
- Dealing with missing healthcare data poses an additional challenge as it can occur when collecting data using devices (Dissanayake, Md Johar 2021).
- Classifying different Patterns of ECG Signals, presents a highly challenging multi-class classification problem (Hu et al. 2022).

Identifying the difference between normal sinus rhythm and irregular heart rhythm is important for accurate ECG signal classification. Normal sinus rhythm is the regular rhythm of the heart, characterized by a P wave, a QRS complex, and a T wave, with a consistent interval between each beat (Kusumoto 2020). Irregular heart rhythms, on the other hand, can be caused by a variety of factors, such as arrhythmias, atrial fibrillation, or heart block, and can result in abnormal ECG patterns (Kusumoto 2020).

Figure 1 Represents the components of ECG Signal.

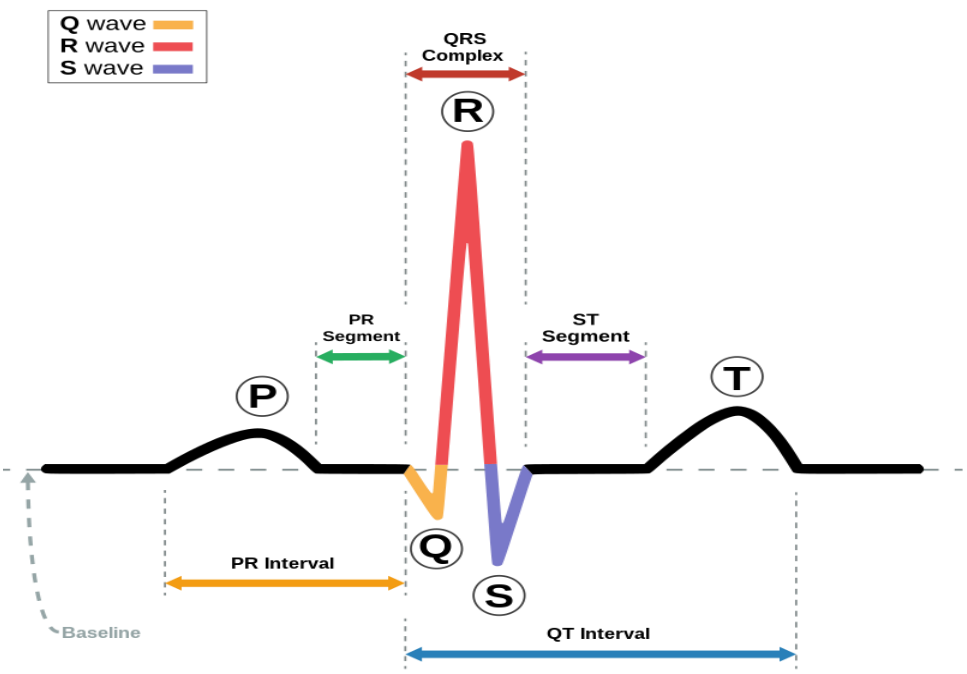


Figure 1:Components of ECG signal (Wikipedia Contributors 2019)

QRS detection is a process in which the QRS complex, which is a characteristic waveform of the electrocardiogram (ECG), is detected and delineated in order to extract important clinical information about the heart's electrical activity (Ardeti et al. 2023). The QRS complex is composed of three individual waveforms: the Q wave, the R wave, and the S wave represents the depolarization of the ventricles of the heart (Geweid, Chen 2022). The absence of the P wave is indicative of an arrhythmia like Atrial Fibrillation (Kusumoto 2020). QRS detection is important in many clinical applications, including the diagnosis of arrhythmias, myocardial Infraction, and heart failure (Geweid, Chen 2022).

To understand this research work, it is important to familiarize with the various classifications of heart diseases and gain an understanding of how ECG data can be interpreted. Normal Sinus Rhythm is characterized by consistent intervals between heartbeats, while Sinus Arrhythmia exhibits irregular intervals between heartbeats (Kleyko et al. 2020). Sinus Arrhythmia, which is considered a normal variation, occurs due to slight variations in heartbeat intervals while breathing in or out. Sinus arrhythmia (irregular rhythm as shown in Figure 2) is indicative of a well-functioning heart and is considered a positive sign of cardiac health (Cleveland Clinic medical professional).

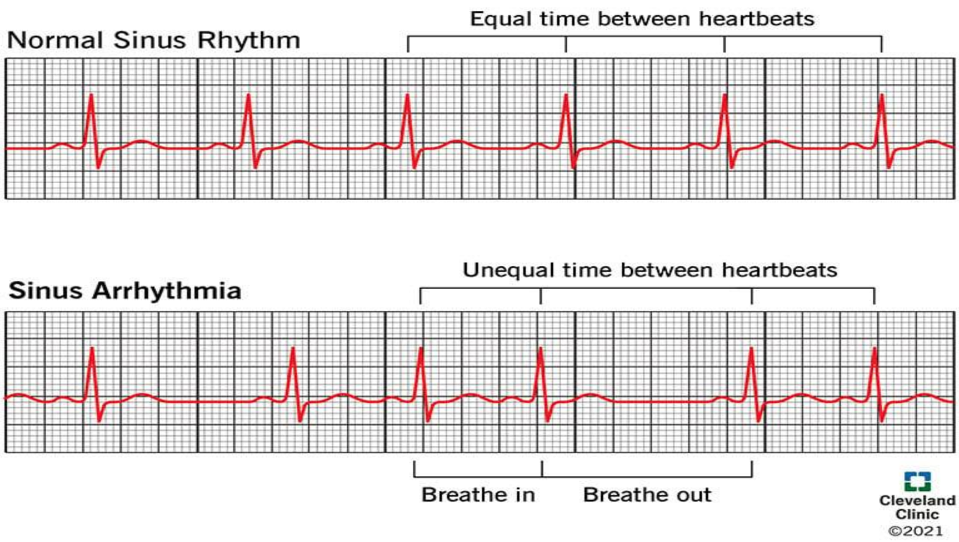


Figure 2: Normal Sinus Rhythm & Arrhythmia (Cleveland Clinic medical professional n.d.)

However, Certain types of arrhythmias (as shown in Figure 3) can have critical implications for human health when the heartbeat becomes excessively slow or fast. These abnormal rhythms can have significant consequences and may require medical intervention to restore a healthy heart rate and rhythm.

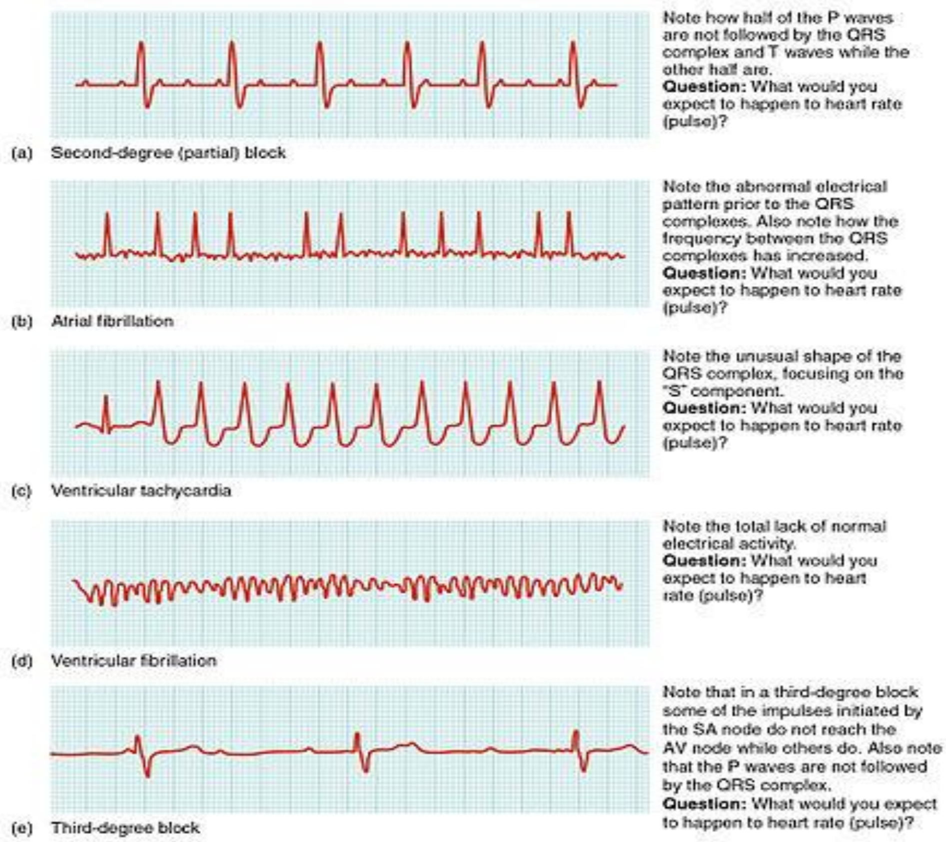


Figure 3: Different forms of Arrhythmia (Wikipedia 2023)

The number of classes used to detect arrhythmia vary depending on the specific study design and the goals of the research. In Binary classification, the ECG signals are classified into two classes, typically normal and abnormal (Badawi et al. 2021). The goal of binary classification is to distinguish between normal and abnormal ECG signals and to detect the presence or absence of arrhythmias.

In Multi-class classification approach, the ECG signals are classified into multiple classes, each representing a different type of arrhythmia (Hu et al. 2022). The number of classes can vary depending on the specific arrhythmias of interest, but typically ranges from three to five. AAMI EC57 (Association for the Advancement of Medical Instrumentation) has developed a standard for the classification of ECG beats. These classes include normal beat (N), Supraventricular ectopic beat (S), Fusion of a normal and ventricular ectopic beat (F), Ventricular ectopic beat (V), and Unknown beat (U) (Mousavi, Afghah 2019).

Use of machine learning algorithms in Arrhythmia classification, enables healthcare professionals to make faster and more accurate diagnoses and treatment decisions. This can help reduce hospitalizations and emergency room visits, saving time and resources.

The primary contributions of this paper can be summarized as follows:

1. A refined wrapper feature selection technique, centred on the Gradient Boosting Classifier, is introduced to identify the most pertinent features.
2. The LightGBM classification algorithm is applied, employing the chosen parameters, to achieve elevated multiclass classification accuracy on the UCI-arrhythmia dataset.

   This research will be published in a Journal Article after having final review with Project Supervisor (Sundaram et al.2023).

Understanding the different stages involved in ECG data analysis is crucial for evaluating the accuracy of machine learning models. The ECG data analysis pipeline typically involves several stages, including data acquisition, pre-processing, feature extraction, Feature reduction and classification.

Conducting a literature review can offer comprehensive insights into the current body of research encompassing Arrhythmia detection, including detailed information on previous studies, methodologies employed, and their respective findings. Furthermore, it can help identify gaps, limitations, and potential avenues for future research within the literature.

The rest of the report is organised in the following manner: Chapter 2 offers an overview of the Literature Review across various phases of ML implementation.

Chapter 3 provides detailed information on the Project Methodology and Project Planning, Chapter 4 offers comprehensive insights into the conducted experiments, Data pre-processing, Model Training and Evaluation, Chapter 5 presents the Results and Discussion, while Chapter 6 elucidates the Conclusion and outlines Future Work.

CHAPTER 2:

LITERATURE REVIEW

Multiple research studies compared the performance of machine learning algorithms and deep learning methods using diverse datasets for training & validations, to classify Arrhythmia or cardiovascular diseases (Ardeti et al. 2023).

Machine Learning (ML) algorithms typically require domain expert knowledge for feature extraction, domain experts often extract relevant features from the raw data and provide them as input to the machine learning model (Hidde Bleijendaal et al. 2021). On the other hand, deep learning algorithms use a layered architecture to automatically learn hierarchical representations of the input data (Kashou et al. 2023). This implies that the deep learning models have the capability to extract valuable features directly from the raw data, eliminating the requirement for explicit feature engineering.

In Supervised ML the outcome is inferred from training data that has been labelled or annotated. Unsupervised learning, on the other hand, seeks to find hidden structures in datasets without any prior knowledge of reference outcomes or labels (Kleyko et al. 2020). It enables the algorithm to discover new groups and clusters of data that a person might not be aware of (Merdjanovska, Rashkovska 2022). Deep learning models are robust to noise and artifacts & can achieve high accuracy for ECG pattern prediction tasks, especially when trained on large datasets (Quer et al. 2021). However, lack of interpretability is a key challenge of using deep learning models in clinical settings. It can be difficult to understand how they arrived at their decision.

In Most of the research studies ECG signal analysis used for the diagnosis of arrhythmias, including atrial fibrillation, ventricular tachycardia, and atrioventricular block, myocardial infarction, and congestive heart failure (Zhang et al. 2020). ECG analysis also used for the diagnosis of neurological conditions, as well as for the detection of sleep apnea and other disorders (Pant et al. 2022).

Different ECG signal generators are employed in clinical settings including Electrocardiogram, Echocardiogram, Holter Monitors, Event Recorders, Wrist-worn wearables, and Biosensors (Ardeti et al. 2023). Each device records Sinus rhythms over a period of time. The electrical signal of heart is measured by these devices in different angles. When 10 electrodes are used in 12 different angles, it generates 12 lead ECG signals, whereas One electrode is used to record Single-lead ECG signals (Badawi et al. 2021).

Researchers have employed two approaches to predict Arrhythmia. The first approach involves selecting RAW ECG signals from Physionet Database and performing data analysis, starting from denoising, QRS complex detection, followed by feature extraction, feature engineering, and classification (Ardeti et al. 2023). In the second approach, research papers that have explored the classification of Arrhythmia, using publicly available databases like Kaggle and UCI Repository. These databases already contain denoised, clinical features extracted from ECG signals, eliminating the requirement to process RAW ECG signals from the scratch (Mustaqeem et al. 2018).

Research experiments have explored the multiclass classification of different types of Cardiac Arrhythmia (Ayar et al. 2023). It includes classification of 16 types of arrhythmias, where one class indicates the absence of disease, and the remaining fifteen classes represent electrocardiogram records of various subtypes of arrhythmias.

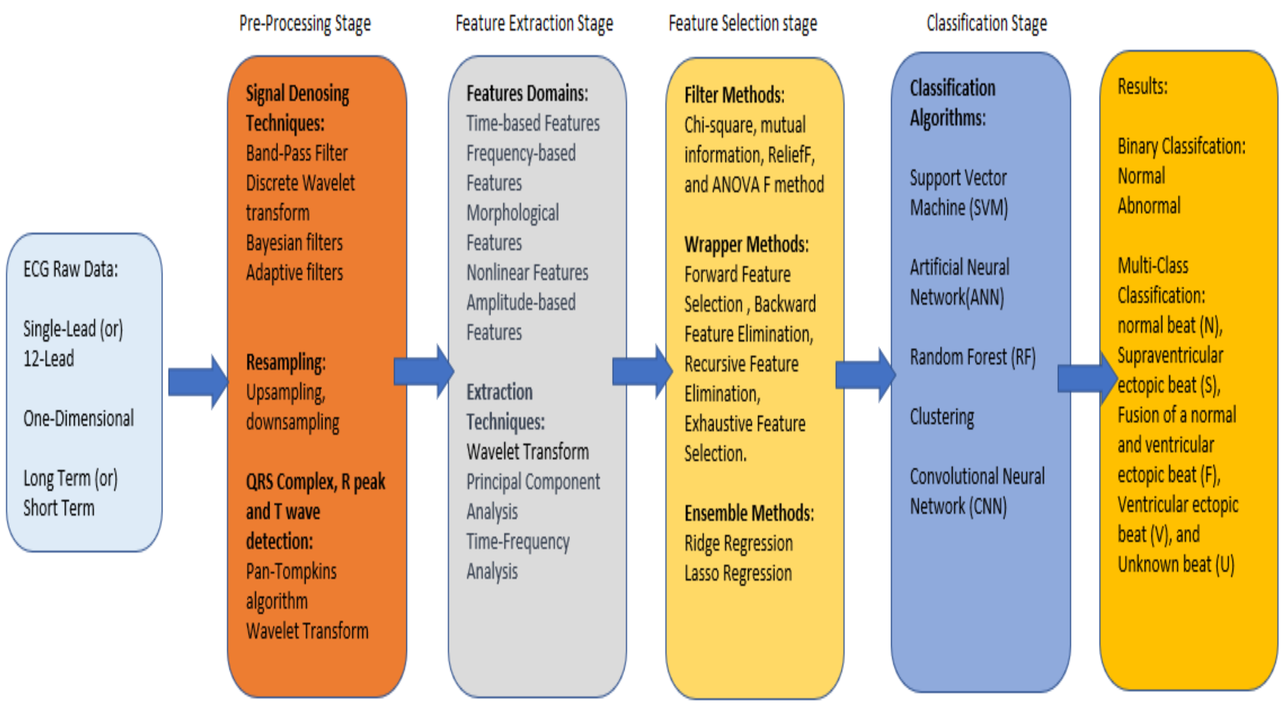


Figure 4: System Flow Diagram & Techniques

Figure 4 depicted above showcases the system flow of a machine learning model, highlighting the commonly employed techniques at each stage.

Further research focused on more comprehensive examination of existing literature, specifically targeting each phase of the Machine Learning System workflow, with the aim of delving deeper into the subject matter.

2.1 Data Pre-Processing:

Various noises associated with the ECG signals are eliminated during the pre-processing stage. Pre-processing and sampling the data at a certain frequency is performed before training ML models (Jha, Kolekar 2020).

The number of techniques and methods employed at each level of the ECG analysis are explained by a recent survey on computational ECG analysis. Filtering methods are used to reduce noise in ECG readings, such as *Standard Finite Impulse Response, Infinite impulse response, Band-Pass filters, Discrete Wavelet transform, Bayesian filters* & *Adaptive filters* (Merdjanovska, Rashkovska 2022).

Pan-Tompkin's algorithm is a commonly used technique for R peak detection in ECG signals (Jha, Kolekar 2020). It involves filtering the ECG signal using a bandpass filter and then detecting QRS complexes based on a combination of amplitude and slope criteria.

Accuracy of six traditional classifier models is evaluated by replacing missing values with mean or Median values of the features (Dissanayake, Md Johar 2021). Handling missing values is an important step to ensure accurate interpretation and diagnosis. One common approach to handle missing values is to use imputation techniques. Imputation involves filling in the missing values with estimated values based on the available data. Some commonly used imputation techniques in ECG analysis include mean imputation, regression imputation, and k-nearest neighbour imputation (Feng et al. 2021).

There are several research papers that discuss using different lengths of ECG signals. Very Short ECG Episodes (10 to 20 heartbeats) are used to detect Atrial fibrillation, R-peaks were located using the annotation files and the PhysioNet WFDB Toolbox (Masud Shah Jahan et al. 2022). The length and frequency of ECG signals impact the effectiveness of ML models. Based on an experimental study, it is evident that accuracy of ML models is improved by upsampling data from 250Hz to 360 Hz & signals with length of 3 seconds performed better than those with a length of 1 second or 5 seconds when using wavelength analysis for anomaly detection (Hu et al. 2022).

Time-Domain (One dimensional) features of ECG signals such as RR Interval, QRS Duration, QT interval, PR interval, ST segment deviation, T wave amplitude are used to segment heart rate and to detect the abnormal signals (Li et al. 2015). Statistical features of RR intervals used in the research studies to identify the heart rate variability (Masuda Shah Jahan et al. 2022).

A scalogram generated from the wavelet transform of an ECG signal provides a time-frequency (two-dimensional) representation of the signal, tensor decomposition (Mathematical Technique) is used to extract meaningful features from ECG (Nesaragi et al., 2022). This approach with Bagged Trees classifier resulted in 96.53% accuracy.

2.2 Feature Extraction:

Multiple feature extraction techniques are used to improve the efficiency of Machine Learning models (Jha, Kolekar 2020).

Discrete Wavelet Transform (DWT) is commonly used to extract features from ECG Signals (Jha, Kolekar 2020). The DWT breaks down a signal into a series of wavelets, each with a different frequency range and time duration, allowing for a more detailed analysis of the signal.

Recent studies witnessed that Pre-trained Deep learning techniques such as SqueezeNet, AlexNet & CNN model for feature extraction improved the accuracy of machine learning algorithms (Abubaker 2023). The accuracy of machine learning algorithms will be significantly impacted by feature extraction approaches, according to literature study. Instead of training a new model from scratch, the pre-trained model's weights are transferred to the new model, which is then fine-tuned to perform a new task (Khan et al. 2021). This approach (Transfer Learning) saved a significant amount of time and resources that would otherwise be required to train a model from scratch on a new dataset. In transfer learning, the pre-trained model is often a deep neural network that has been trained on a large dataset, such as ImageNet, for a specific task like image classification (Khan et al. 2021).

Gabor wavelets are used to decompose ECG signals into various frequency bands (Liu et al. 2020). The wavelet scattering transform is a relatively new technique in the field of ECG signal processing, and its application to ECG beat classification is an interesting contribution to the field. The authors have provided a detailed description of the wavelet scattering transform and its implementation in ECG beat classification.

2.3 Feature Selection:

A critical review on feature selection or dimensionality reduction techniques are discussed in Journals. Experimental research is performed on three major supervised feature selection techniques (Filter, Wrapper, and Embedded), and then performance of different classification algorithms is evaluated (Dissanayake, Md Johar 2021). Authors have used four filter-based feature selection algorithms, such as Chi-square, mutual information, ReliefF, and ANOVA F method, and four wrapper-based algorithms, namely Forward Feature Selection (FFS), Backward Feature Elimination (BFE), Recursive Feature Elimination (RFE) & Exhaustive Feature Selection (EFS). The embedded techniques Ridge Regression and Lasso Regression are also used. With a classification accuracy of 88.52% and a precision of 91.30%, the feature subset chosen using the backward feature selection technique has resulted in the highest classification performance. However, there are some limitations to the study. Firstly, the study uses a relatively small dataset with only 303 instances, which may not be representative of the entire population. Secondly, the study uses only 13 features, which may not be sufficient to capture all the relevant information for heart disease prediction. Appropriate feature selection process will increase the accuracy of the classification model and decreases the processing time.

To enhance the accuracy of Incremental backpropagation neural network (IBPLN) and Levenberg-Marquardt (LM) algorithm, researchers have utilized Correlation based Feature selection (CFS) in combination with linear forward selection (Mitra, Samanta 2013). This approach selected the most suitable features from a high-dimensional dataset containing 182 features, reducing it to 18 features. The application of CFS with linear forward selection has shown a substantial improvement in the accuracy of both IBPLN and LM algorithms.

Arrhythmias are commonly associated with irregular rhythm patterns, but they can also be identified by abnormal waveform shapes or missing components (Geweid, Chen 2022). A comprehensive analysis is performed considering a total of 55 features. Combination of various features utilized to effectively represent the rhythm. To narrow down the feature selection, an initial pre-selection process was performed using ANOVA in combination with stepwise discriminant analysis using the RAOV (Reduced Average Overlap) criterion (Geweid, Chen 2022). This led to the identification of 10 subsets of data, each with an increasing number of features. However, this study did not consider the limitations of ANOVA, as it sensitive to noise and outliers in the data and may not always produce the best subset of features for a given task.

Kleyko et al. (2020) conducted an experiment to automatically detect Atrial Fibrillation (AF) from short-term ECG signals. They trained machine learning (ML) models using

Long-term ECG Signals and utilized MATLAB source code to extract four different feature sets, each consisting of varying numbers of features (188 features, 171 features, 150 features, and 509 features). These feature sets were derived based on existing solutions proposed in literature reviews for ECG long-term signal analysis. To optimize the feature sets, the authors employed different feature selection methods, including Recursive Feature Elimination (RFE), Neighborhood Component Analysis (NCA), and a statistical approach. Among these methods, the RFE method yielded the most promising results compared to the other two methods, showcasing its effectiveness in feature selection for AF detection.

Principal Component Analysis (PCA), Linear Discriminant Analysis (LDA), and Independent Component Analysis (ICA) methods are used in literature to reduce the dimensionality of linear and non-linear features extracted using wavelet transform technique, it captures the most important information for subsequent analysis or classification tasks (Jha, Kolekar 2020).

A novel unsupervised feature reduction technique is presented for imbalanced dataset, using similarity-based clustering (Sun et al. 2022). Firstly, a new similarity measure is devised, which is used in a hierarchical clustering model that considers the samples. This enables the reasonable generation of new samples between the cluster center of each sample cluster and its nearest neighbors. Consequently, a hybrid sampling method is introduced to construct a balanced decision system.

2.4 Classification Algorithms:

Both Supervised and Unsupervised learning models are used in the multi-class classification of Arrhythmia. Deep learning approaches requires massive amounts of data to be trained, as many parameters that need to be learned from data (Kachuee et al. 2018).

A comprehensive overview of the computational techniques used for ECG analysis and interpretation, and their contribution to medical advances are discussed in this study (Lyon et al. 2018).

Widely used Supervised learning models are Support Vector Machine, K-Nearest Neighbors, Decision Tree & Random Forest (Merdjanovska, Rashkovska 2022).

ECG clustering algorithms & Deep learning models are employed as unsupervised learning methods to overcome challenges occurred in supervised methods. (Nezamabadi et al. 2023).

Research studies have investigated diverse frameworks to address the challenges encountered in automated classification using Machine learning models. Performance of multiclass classification is enhanced through the introduction of class hierarchies (Silva-Palacios et al. 2017). The authors propose a framework that leverages the hierarchical relationships among classes to improve the accuracy of the classification models. They conduct experiments on various datasets and demonstrate that the incorporation of class hierarchies leads to significant performance improvements compared to traditional flat multiclass classification approaches. The study highlights the potential of utilizing class hierarchies to enhance the accuracy and effectiveness of multiclass classification algorithms in various domains.

One-versus-all (OVA) is a multi-class classification strategy where multiple binary classifiers are trained to distinguish each class from all other classes individually. A novel method called Differential Partition Sampling Ensemble (DPSE) is introduced within the One-versus-all (OVA) framework to address serious class imbalance in multiclass classification tasks (Gao et al. 2021). DPSE utilizes sampling techniques to balance majority and minority samples in each binary training dataset. Safe, borderline, rare, and outlier samples are identified based on neighborhood information, and Random Undersampling (s-Random Undersampling) and SMOTE (Synthetic Minority Over-sampling Technique) are applied accordingly. Multiple sub-classifiers are trained on balanced datasets, and experiments on 27 KEEL public multiclass datasets demonstrate DPSE's superior performance compared to typical OVA, One-versus-One, and direct classification methods.

2.4.1 Support Vector Machine (SVM):

SVM with three different strategies are experimented in the research studies, such as One-Against-One (OAO), One Against-All (OAA), and error correction code (ECC) to classify different types of Arrythmias'. The classifiers' performance was assessed using accuracy, kappa statistics, and mean square root error, resulting in an evaluated accuracy of 92.07% (Mustaqeem et al. 2018). The approach involves transforming the multi-class classification problem into multiple binary classification problems. It then combines the results of the binary classification problems to obtain the final result.

Twin support vector machine based on multi-order moment matching (TSVM-M^3) is proposed to improve the computational performance of large-scale classification (Qiang et al. 2022).

A new Hybrid Approach of Dual Support Vector Machine (HA-DSVM) is introduced to classify ECG from short single-lead ECG recordings (Geweid, Chen 2022). The recordings are separated into four Categories: Noisy recordings, Normal Sinus Rhythms (NSR), various rhythms, and AF. In this research, performance of top five scoring models are compared with Hybrid SVM approach, based on the accuracy of each model in identifying R peaks, QRS Waves, P-R Interval, Q-T interval, S-T interval & P, QRS,T waves. Dual SVM model increased the accuracy than using original SVM method. The study combines the features extracted from time-domain, frequency-domain, and wavelet transform domains to classify AF. The challenges of AF classification from short ECG recordings can be addressed with the hybrid approach of dual SVM, and the classification performance is enhanced by the combining different features from multiple domains. The study only recorded an ECG with a single-lead, which would not have been sufficient to identify AF in all patients.

2.4.2 Neural Networks:

Deep learning models, such as convolutional neural networks (CNNs), can learn complex patterns and relationships directly from the raw ECG data, which can improve accuracy and reduce the need for manual feature engineering (Merdjanovska, Rashkovska 2022). It can be used for both binary and multi-class classification tasks, as well as for detecting multiple types of arrhythmias simultaneously (Khan et al. 2021). Deep learning models can be trained on imbalanced datasets to address the issue of class imbalance, as well as on large datasets to improve generalization and robustness (Hu et al. 2022).

A deep learning (DL) technique using long short-term memory (LSTM) is introduced to classify arrhythmia patients based on a high-dimensional cardiac arrhythmia dataset from the University of California, Irvine (UCI) repository (Ashfaq Khan, Kim 2021). To avoid the slow convergence of the LSTM, ADADELTA (adaptive learning rate optimization algorithm) learning rate technique is used. Model's performance is compared with research studies that use traditional ML models. The results show that the LSTM DL technique achieved the highest accuracy of 93.5% on the high-dimensional dataset. However, the paper does not provide specific details about the loss metrics used for evaluating the proposed deep learning model.

Artificial Neural Network (ANN) is proposed for Real-time monitoring of ECG signals and anomaly detection (Badawi et al. 2021). Datasets are resampled to standard frequency of 125 Hz, then converted to a binary dataset of two categories such as Normal & abnormal signals (9 different abnormal heartbeat categories). Real-time signals are acquired from the event monitor via Raspberry Pi. By employing an Artificial Neural Network (ANN) architecture with two hidden layers alongside the input and output layers, notable improvements in accuracy (99.3%) and processing time were observed compared to other machine learning models. However, its usability is limited to predicting only abnormal signals, as all abnormal ECG signals are grouped into a single category. Thus, it cannot be utilized for multi-class classification, where distinguishing between the different abnormal heartbeat categories would be necessary.

A Generalized methodology is developed to process nonuniform formats of ECG images collected from different equipment's, Single Shoot Detection (SSD) MobileNet v2-based Deep Neural Network architecture is used to detect cardiovascular diseases from 12-lead-based ECG images (Khan et al. 2021). ECG images are resized to 300KB to reduce the time for training the model. LabelImg tool is used to label the dataset. Proposed algorithm is used to detect four classes, Myocardial infarction (MI), Abnormal heartbeat, Previous history of MI & Normal. However, the study did not include a comparison of the proposed model's performance on alternative formats of ECG in order to suggest a scalable system.

A novel deep convolutional neural network is presented to classify imbalanced ECG heartbeats & achieved high classification performance (Sellami, Hwang 2019). However, it is important to emphasize that this approach has exclusively been evaluated on datasets that have been annotated, highlighting the crucial role of clinical experts in ensuring accurate annotation.

Increasing the number of layers in a neural network has an influence on the classification performance of high-dimensional time series data like ECG datasets (Sellami, Hwang 2019). To expedite the training process, batch normalization is incorporated after each convolutional layer. Furthermore, to mitigate overfitting, used 75% dropout before layers that are prone to overfitting. The classification process relies solely on raw ECG signals without any pre-processing or feature extraction. By comparing three inputs, each consist of a distinct number of heartbeats, the most optimal input is determined. This approach may introduce bias and prolong the training time.

(Kachuee et al. 2018) Authors employed a transferable Deep Neural network model which can used to classify both arrhythmia & myocardial infarction. The network is trained on a large dataset of annotated ECG recordings to learn representative features that can be transferred across different datasets. By employing transfer learning, the authors achieve promising results in ECG heartbeat classification, even with limited amounts of labelled data. To address the vanishing gradient problem (where gradients diminish as they propagate through many layers) and facilitate the flow of gradients through many layers, residual skip connections with five blocks are employed. These skip connections enable the gradients to bypass multiple layers and flow directly through the network. This enables the effective training of deep networks. The proposed approach is evaluated on multiple benchmark datasets, and the results demonstrate the effectiveness and transferability of the learned representations. The study compares the performance of their model with other state-of-the-art methods, showcasing its competitive classification accuracy.

2.4.3 K-Nearest Neighbors (KNN):

K-Nearest Neighbors (KNN) is a simple non-parametric algorithm, yet powerful machine learning algorithm used for ECG classification in literature (Sun et al. 2022).

KNN algorithm with an ensemble framework, utilizing random subspace sampling is used to classify 5 types of cardiac arrhythmias (Ramasamy et al. 2022). The optimization process is performed using the Jaya algorithm. Proposed method aims to address the challenges of detecting arrhythmias from noisy and complex ECG signals, which can be difficult due to the variability in the data and the presence of artifacts. The method uses Fourier–Bessel series expansion (FBSE) to transform the raw ECG cardiac arrhythmia beats into more meaningful sequences, which are then used as inputs to an (random subspace K-nearest neighbor) ERS-KNN classifier. Using PCA, the features are acquired from FBSE sequences. Cross-validation technique is performed on the whole dataset to evaluate and select the best learning parameter values. This novel FBSE approach influenced the performance of KNN model and resulted in 98% accuracy. The use of low-dimensional FBSE features in the classification improved the processing time of KNN.

Adaptive weighted K-nearest neighbors (AWKNN) algorithm is proposed for feature reduction in imbalanced datasets using similarity-based feature clustering (Sun et al. 2022). To overcome challenges of over-sampling and under-sampling techniques in machine learning, authors proposed clustering based feature selection (Hybrid sampling) model to group similar features together, reducing the dimensionality of the dataset and to improve classification performance. This method is evaluated on 29 imbalanced datasets from UCI Machine learning repository. The results show that the proposed method achieves higher classification accuracy compared to other methods such as traditional K-nearest neighbors and decision trees. The method is also shown to be effective in reducing the dimensionality of the dataset while maintaining high classification accuracy.

Best performance of KNN is achieved where ECG signals are decomposed using Wavelet Scattering Transform method (Liu et al. 2020). Recent studies in the field of ECG signal analysis often employ KNN models for classification tasks and deep neural network models for efficient feature extraction from ECG signals. The combination of these techniques enables accurate classification based on learned features from the ECG data (Li, Zhang 2023).

2.4.4 Random Forest (RF):

RF Model's classification accuracy increased as the number of decision trees increased (Ganeshkumar, YSKumaraswamy 2012). RF method is proposed to classify ECG signals into different types of arrhythmias. The method involves feature extraction from the ECG signals using statistical and morphological features and training the RF classifier using a large dataset of annotated ECG signals. The method is evaluated on the MIT-BIH arrhythmia database, which contains a large collection of ECG signals with different types of arrhythmias.

ECG rhythms are pre-trained by Convolutional Neural Network & fed into RF model for heartbeat classification (Zou et al. 2022). In this study, two distinct sets of features are extracted from ECG dataset, one contains heartbeat (Part 1) segmentation and other one includes Rhythm segmentation (Part 2). In contrast to training a heartbeat classifier, training an ECG recording classifier requires a different kind of annotated dataset (Zou et al. 2022). ECG Rhythms are pre-trained & labelled by CNN. Further, conventional features extracted from Part 1 and CNN labelled features from Part 2 are concatenated and employed to train RF model. A novel transfer learning approach is presented in this study, where pattern discovered from the ECG classifier is transferred to a heartbeat classifier via segment label.

Highest accuracy of RF model is resulted by using novel amplitude and slope-based features of P-waves (Evangelia Myrovali et al. 2022). P-waves are detected, and Statistical features of P-wave are used to predict Paroxysmal Atrial Fibrillation (PAF). In this study, the performance of machine learning models and previously reported feature extraction techniques were compared to predict PAF. The results showed that the RF model achieved the highest accuracy, surpassing the performance of the feature extraction methods described in the literature. Notably, the RF model utilized innovative features extracted from P-waves, contributing to its superior predictive capabilities.

2.4.5 Clustering

To mitigate the impact of subject-specific variations and enable the generalization of Deep Learning algorithms across different subjects, a novel multi-level unsupervised domain adaptation framework is proposed (He et al. 2023) to detect arrhythmias.

Recent research has focused on investigating the effectiveness of adaptive filtering and feature extraction algorithms for noise removal in bio-signals (Sun et al. 2022). This is particularly crucial because traditional algorithms designed for extracting ECG signal parameters encounter limitations when employed in real-time signal processing scenarios. By utilizing adaptive filtering and feature extraction techniques, these studies aim to improve the accuracy and reliability of bio-signal analysis in real-time applications.

A novel feature extraction technique based on unsupervised Multiresolution Teager Energy Operator (MTEO) algorithm & local extremum search, is proposed to extract ECG parameters from movement contaminated signals and from stationary ECG recordings (Butkevičiūtė et al. 2022). Proposed method extracted 9 ECG parameters Q, R, S, T amplitudes, Q-T, S-T, R-R, QRS intervals and T wave intervals (Butkevičiūtė et al. 2022). The study involves a comparison of multiple QRS detection algorithms using the MIT-BIH arrhythmia database to assess their accuracy. Additionally, the research includes an evaluation of various open-access QRS complex detection algorithms alongside a newly proposed algorithm for ECG pattern recognition in the presence of movement. However, the study exclusively focuses on individuals without any heart-related abnormalities or disorders. Consequently, the findings of this research are not applicable for diagnostic purposes concerning cardiac diseases.

2.5 Evaluation & Performance Metrics:

To assess the effectiveness of ML models, standard performance indicators are computed in the literature (Nezamabadi et al. 2023). To guarantee generalizability, the performance is tested using hidden datasets (Zhao et al. 2022).

Literature review highlights the widespread utilization of the five-fold cross-validation strategy in the development of training and testing models (Nesaragi et al. 2022). This approach helps assess the performance and generalizability of machine learning algorithms effectively.

In a related thread, authors emphasize the use of the ten-fold cross-validation method, which provides a more rigorous evaluation of model performance by dividing the dataset into ten subsets for training and testing purposes (Ramasamy et al. 2022). This technique aids in obtaining robust and reliable results.

The performance is typically assessed by its accuracy, sensitivity, specificity, and positive predictive value (PPV) (Ramasamy et al. 2022). These metrics serve as important indicators to gauge the effectiveness and reliability of the algorithm's outcomes.

To evaluate the classification performance of the compared algorithms, various metrics are employed, including the receiver operating characteristic (ROC) curve, the area under the ROC curve (AUC), F-Score, G-mean, and mean F-measure (MFM) (Sun et al. 2022). These metrics offer valuable insights into the accuracy, sensitivity, and overall effectiveness of the classification models.

Confusion matrix is incorporated to assess the performance of classification models in the context of ECG anomaly detection (Khan et al. 2021). Various parameters are measured by Confusion Matrix such as true positive (TP), false positive (FP), true negative (TN), and false-negative (FN) (Ramasamy et al. 2022). It serves as a valuable tool in assessing the strengths and weaknesses of different classification algorithms, allowing researchers to fine-tune their models and optimize their performance.

Table 3: Summary of Literature Review

Author /Year	Title	Classifier	Number of classes	Performance	Performance Metrics	Feature Selection Techniques
Abubaker 2023	Detection of Cardiovascular Diseases in ECG Images Using Machine Learning and Deep Learning Methods	SVM ,KNN, DT,RF, NB	4	SVM Accuracy with different feature extraction techniques 99.47%, 97.87%, and 97.66%	Accuracy, Precision, Recall, F1 score	SqueezeNet, AlexNet , Transfer Learning
Ayar et al. 2023	NSICA: Multi-objective imperialist competitive algorithm for feature selection in arrhythmia diagnosis.	liner regression, SVM, Naive Bayes, & decision tree	16	Accuracy = 83.76 % Precision = 84.23% Recall = 83.76% F1Score = 83.84%	Accuracy, Precision, Recall, F1 score, AUC Curve	Non-dominated sorting imperialist competition algorithm
Lown et al. 2020	Machine learning detection of Atrial Fibrillation using wearable technology	SVM	2	Sensitivity:99.2%	Accuracy, Precision, Recall, F1 score, AUC Curve	Wavelet transformation used to reduce volume of data
R. Abirami, Vincent 2019	Cardiac Arrhythmia Detection Using Ensemble of Machine Learning Algorithms	Ensemble Learning	16	Accuracy: 85%	Accuracy, Precision, Recall, F1 score, AUC Curve, Kappa Statistics	PCA
Masud Shah Jahan et al. 2022	Short-term atrial fibrillation detection using electrocardiograms: A comparison of machine learning approaches	Support Vector Machine (SVM), Decision Tree (DT), Random Forest (RF), Stacking Classifier (SC), Extreme Gradient Boosting (XGBoost), and Adaptive Boosting (AdaBoost)	2	XGBoost Specificity: 96.1	False Positive Rate, Specificity, Sensitivity	
Zou et al. 2022	Heartbeat Classification by Random Forest With a Novel Context Feature: A Segment Label	Random Forest	5	Accuracy 96%	Accuracy, Precision, Recall, F1 score	Convolutional Neural Network
Ramasamy et al. 2022	Detection of cardiac arrhythmias from ECG signals using FBSE and Jaya optimized ensemble random subspace K-nearest neighbor algorithm	Ensemble random subspace KNN classifier	5	Accuracy : 98.0% for noisy signals and 98.1% for noise-free signals	Accuracy, Specificity, Sensitivity	Fourier Bessel series expansion (FBSE) feature extraction method
Ganeshkumar, YSKumaraswamy 20	Investigating Cardiac Arrhythmia in ECG using Random Forest Classification	Random Forest	6	Accuracy 92.16 with 30 decision trees		Discrete Cosine Transform (DCT)
Geweid, Chen 2022	Automatic classification of atrial fibrillation from short single-lead ECG recordings using a Hybrid Approach of Dual Support Vector Machine	Hybrid Support Vector Machine	4	Accuracy: AFr rhythm: 93.25% Noisy rhythm: 91.85% Other rhythm: 96.13% normal rhythm: 96.83%		ANOVA with a stepwise discriminant analysis

In this chapter, it has been observed that experiments encompassed the utilization of both machine learning and deep learning models for arrhythmia classification. The primary objective was to classify Arrhythmia into binary or multi-class categories, typically ranging from 4 to 6 groups.

However, it's noteworthy that there has been limited research dedicated to multi-class classification involving as many as 16 target classes. Consequently, there's a significant need to explore optimal feature selection techniques and classification algorithms to enhance the accuracy in this context.

CHAPTER 3: Project Methodology

3.1 Research Focus Area:

Distinguishing between multiple classes or categories in multi-class classification presents significant challenges in the field of machine learning, primarily due to the inherent complexity and intricacy involved.

Multi-class classification presents challenges due to factors such as an increased number of classes, class imbalance, overlapping classes, high-dimensional data, computational complexity, and the curse of dimensionality (Gao et al. 2021). As the number of classes grows, the problem becomes more complex, and classifying data into multiple categories becomes more difficult. Class imbalance can result in biased models that perform poorly on minority classes (Sáez et al. 2016). Overlapping classes make it harder to distinguish between similar instances. High-dimensional data introduces noise and complexity, while computational complexity escalates with more classes and features (Mustaqeem et al. 2018). The curse of dimensionality highlights the difficulty of accurately representing and generalizing from high-dimensional data, leading to overfitting (Gao et al. 2021).

Considering the insights gathered from the literature review, numerous factors have been identified as influencing the accuracy of classification models. Hence, the focus of this study will revolve around the multi-class classification problem in the context of Arrhythmia, as well as the various factors that have an impact on the accuracy of classification.

3.2 Aims and Objectives.

The aim of this research is to determine a classification algorithm with the highest accuracy by utilizing diverse feature selection techniques on a high dimensional dataset. The study will investigate different factors influencing accuracy of ML models and propose an accurate machine learning algorithm for the multi-class classification of Arrhythmia.

3.2.1 Objectives:

- Establish effective preprocessing approaches to clean, transform and handle missing values in the the data.
- Explore four optimal Feature selection techniques for identifying essential features from high-dimensional datasets.
- Examine six different machine learning models for the classification of Arrhythmia
- Perform experiments for multi-class arrhythmia classification using four distinct feature sets.
- Evaluate Model accuracy using appropriate metrics.
- Compare ML model accuracy and suggest optimal algorithms and feature selection techniques.

3.2.2 Tasks:

This research involves a variety of tasks, which are listed below and follows a standard approach to implement machine learning model in the Arrhythmia classification.

Initial Task of the research project will involve engaging in discussions with the project supervisor to receive valuable guidance regarding the chosen research area. Maintaining regular communication and seeking input from subject matter experts will be instrumental in effectively achieving the project goals while ensuring alignment and minimizing deviations.

Part 2 of the research project will focus on the project planning and implementation.

Data Collection: There are several ways to gather medical research data.

1.Experimental Approach (Primary Data): Randomized clinical trials of body wearable sensors or ECG monitors are used to collect the sensitive health data, however, these studies typically do not offer accurate annotations and relevant labels on physiological signals (Banaee et al. 2013).
2.Clinical or Online Databases (Secondary Data): Medical Information from Hospitals and research labs is available online, and this data been used to evaluate the models in many research papers.
3.Simulated data: Multiple methods are used to simulate the human vital signs. Using python, health data can be simulated.

As there are several legal issues involved collecting health data, online free medical research database UCI Machine Learning Repository is utilized for research in this paper. Upon selecting the data, it will be thoroughly discussed with the supervisor and obtain their approval for utilizing the dataset.

A detailed summary of tasks, dependencies and deliverables are presented in the following table.

Table 2: Tasks and Estimated Deadline

Phase	**Task**	**Dependencies**	**Deliverables**	**Estimated Deadline**
P1: Literature Review	1A. Explore Google Scholar, Science Direct Journals, Conference Papers on the Multiclass classification of Arrhythmia.	NA	List of References	10/05/2023
	1B. Create a draft version of Literature Review including: 1. Challenges in Arrhythmia classification 2. Feature Selection Techniques and ML models for Classification 3. Evaluation Metrics 4. Project Planning	1A	Draft version of Literature Review	22/05/2023

	1C. Update the Literature Review with feedback received from Supervisor.	1B	Finalized Literature Review Document	21/06/2023
	1D. Conclusion on the Research Topic with Supervisor	1C	NA	30/06/2023
P2: Data Collection	2A. Explore the options for Data Collection: 1. Experimental Approach 2. Clinical Trails 3. Simulated Data 4. Public Database	NA	List of finalized datasets	02/07/2023
	2B. Discussion with Supervisor on the finalized dataset	2A	Approval from Supervisor	05/07/2023
P3: Data Preprocessing	3A. Data Cleaning 1. Missing Value Imputation 2. Formatting of Data Types	2B	NA	07/07/2023
	3B. Exploratory Data Analysis on: 1. Data Distribution 2. Data Visualization 3. Understand the Relationship between variables	3A	NA	08/07/2023
P4: Feature Selection	4A. Feature selection using “Select Kbest” method	3B	Kbest Feature subset (1)	10/07/2023

	4B. Feature Selection using “PCA” method	NA	PCA Feature subset (2)	10/07/2023
	4C. Feature Selection using “Lasso Regression” method	NA	Lasso Feature Subset (3)	12/07/2023
	4D. Feature selection using “RFE”	NA	RFE feature Subset(4)	12/07/2023
P5: Train Test Split	5A. Split the Feature sets into Training & Testing Datasets	4A,4B,4C,4D	NA	14/07/2023
	5B. Standardize the Training & Testing Data	5A	NA	14/07/2023
P6: Model Training	6A. Train the Classification Models using Kbest Feature subset	5B	NA	17/07/2023
	6B. Train the Classification Models using PCA Feature subset	5B	NA	19/07/2023
	6C. Train the Classification Models using Lasso Regression Feature subset	5B	NA	20/07/2023
	6D. Train the Classification Models using Wrapped Based Feature subset	5B	NA	26/07/2023
	Discussion with Supervisor on the proposed framework & Receive approval on the approach	5B	Supervisor Record form	01/08/2023
P7: Model Evaluation	7A. Evaluate the performance of classification Models for all 4 feature sets	6A,6B,6C,6D	NA	08/08/2023
P8: Comparison	8A. Compare the performance of ML Models based on Evaluation metrics	7A	NA	11/08/2023

	Accuracy, Precision, Recall, AUC-ROC.			
P10: Draft Report	Draft Report Submission on the Research Topic	Tasks 1 to 8	Draft Report	18/08/2023
P11: Final Dissertation Report	Detailed Analysis on the proposed Framework and Report Writing	Tasks 1 to 8	Final Report	01/09/2023

3.3 Resources required.

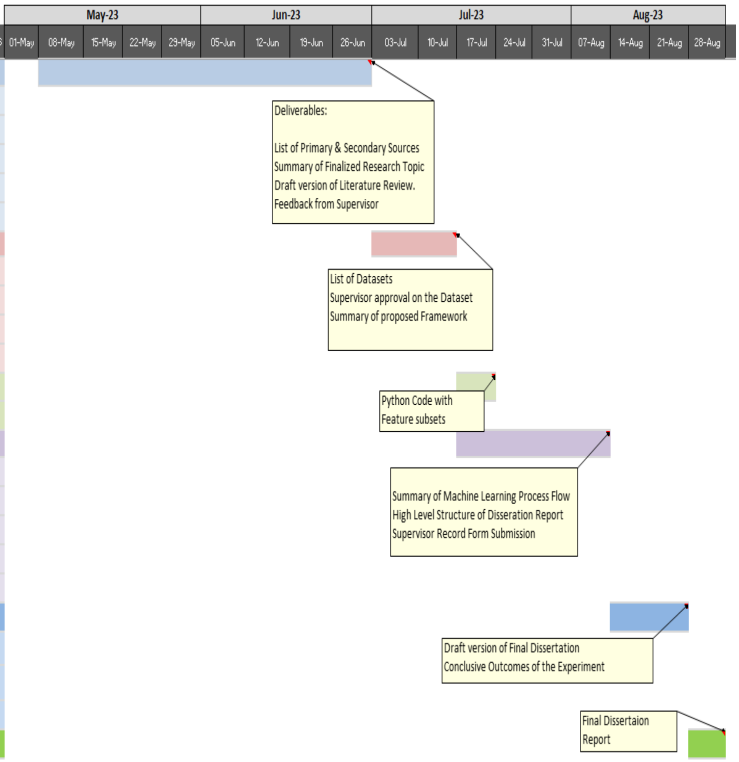
1. Primary Sources: Conference proceedings & research papers, Journal articles are required to gain insights on the novel methodologies and the recent research findings in ECG anomaly detection.
2. Secondary Data Sources such as Survey article are required, which is typically a comprehensive review or summary of existing research and literature on a particular topic. It aims to provide an overview and analysis of the current state of knowledge in a specific field or subject area. In this sense, survey articles often synthesize information from various primary and secondary sources, making them similar to secondary sources.
3. Tertiary Sources such as Encyclopaedia, Wikipedia, Clinical websites, provide healthcare professionals, students, and the general public with reliable and accessible information related to various medical topics. These sources are useful in general understanding for ECG components and cardiovascular diseases.
4. Medical Research Arrhythmia Dataset, to test and verify accuracy of different Machine learning algorithms. It provides access to a 12-Lead ECG signals, encompassing various age groups and Arrhythmias.
5. Machine learning libraries that provide tools and frameworks for implementing machine learning algorithms, such as *scikit-learn, Pandas, numpy* etc.
6. Python (version 3.10.12) programming language for coding the Machine learning project.
7. Integrated Development Environments (IDEs) such as Kaggle or Jupyter notebook, Google colab for Python.

3.4 Time plan

Following Gantt chart shows the deadline for each stage & tasks involved to complete the research successfully on time.

Table 1: Gannt Chart

TASK	START	END	DAYS
Literature Review			
Identify the Machine Learning Algorithms used for Arrhythmia detection	10-May-23	20-May-23	10
Understand the challenges in ECG Dataset patterns	22-May-23	02-Jun-23	11
Understand the limitations of Machine learning Models	05-Jun-23	17-Jun-23	12
Explore the Machine Learning model evaluation techniques	19-Jun-23	23-Jun-23	4
Identify the Performance metrics of ML models	26-Jun-23	30-Jun-23	4
Data Selection & Pre-Processing			
Identify relevant dataset for experiment	01-Jul-23	05-Jul-23	4
Review the Dataset with Supervisor	04-Jul-23	05-Jul-23	1
Data Cleaning & Exploratory Data Analysis	06-Jul-23	10-Jul-23	4
Missing Value Imputation and Create 4 Feature Subsets	11-Jul-23	16-Jul-23	5
Data Sampling			
Split your dataset into training, validation, and testing sets	16 Jul 2023	17 Jul 2023	1
Model Training & Report Writing			
Train the Classification Models using Kbest Feature Set	18-Jul-23	20-Jul-23	2
Train the Classification Models using PCA Feature Set	21-Jul-23	22-Jul-23	1
Train the Classification Models using Lasso Regression Feature Set	24-Jul-23	26-Jul-23	2
Train the Classification Models using Wrapped Based Feature Set	28-Jul-23	01-Aug-23	4
Supervision Record Form Submission	31-Jul-23	01-Aug-23	1
Model Evaluation & Report Writing			
Evaluate the performance of classification Models for all 4 feature sets	02-Aug-23	07-Aug-23	5
Compare the performance of ML Models based on Evaluation metrics Accuracy, Precision, Recall, AUC-ROC	08-Aug-23	11-Aug-23	3
Draft Report Submission	17-Aug-23	18-Aug-23	1
Report Writing & Project Submission	21 Aug 2023	1 Sep 2023	11



3.5 Proposed System:

Previous research studies have employed techniques such as Filter, Wrapper, and regression models for feature selection, accompanied by the utilization of classification models such as SVM, RF, NB, MLP, and KNN for multi-class classification purposes (M Devadas 2021). It's been observed that Features extracted using Recursive Feature Elimination (Wrapper Method) in conjunction with RF classifier, returns highest accuracy in small datasets (Dissanayake, Md Johar 2021).

This study will involve a comprehensive comparison between the current models employed for multi-class classification and the subsequent detailed description of the unique framework.

A unique framework for multi-class classification:

1.Conduct experiments involving alternate approaches such as SelectKbest and Recursive Feature Elimination (RFE) in conjunction with Gradient Boosting for feature selection.
2.Explore the application of the LightGBM model for classification through experimentation.

RFE based on the Feature importance score derived from a "Gradient Boost "(GB) model will be applied to datasets with high dimensionality. Because GB model aims to improve model accuracy by focusing on reducing errors in predictions, potentially leading to better performance compared to Random Forest.

LightGBM classification Model, a specialized high-performance machine learning framework designed for gradient boosting. LightGBM employs a leaf-wise tree growth strategy and histogram-based binning, which significantly enhances training speed and makes it highly efficient, especially for large datasets. It also has mechanisms to directly address imbalanced data, making it particularly suitable for such scenarios.

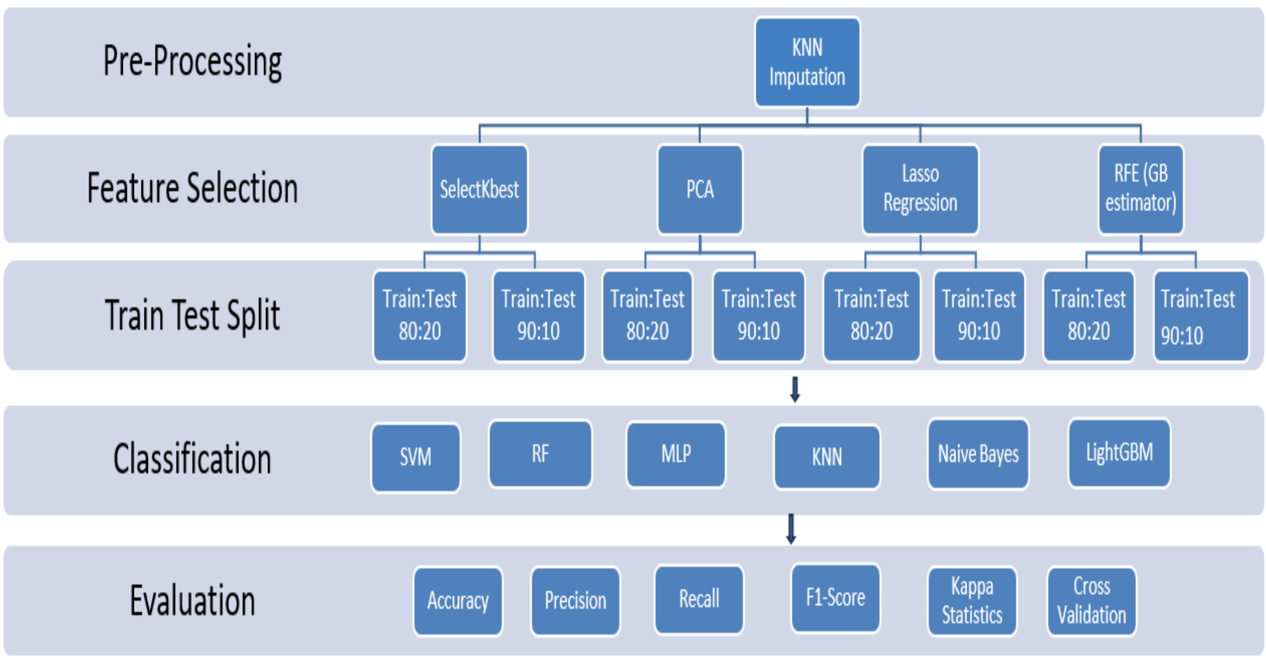


Figure 4: Flowchart of ML Process Flow

Figure 4 above illustrates the methodology employed in this research project.

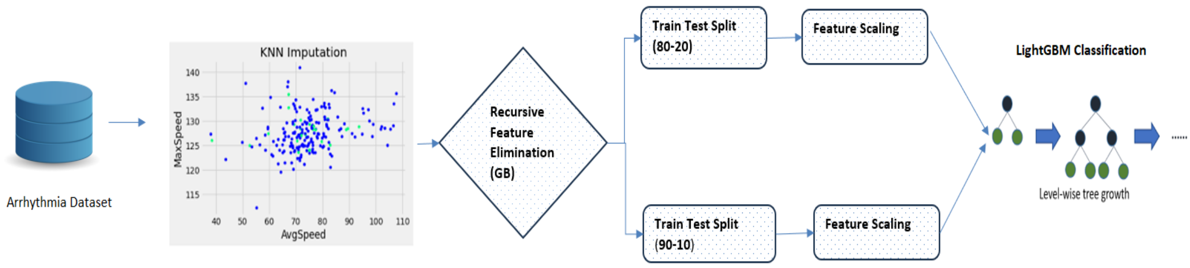


Figure 5: Proposed System with High Accuracy

Figure 5 above represents the proposed framework that yields the highest accuracy. Detailed approach on the proposed framework & methodology will be discussed in Chapter 4.

CHAPTER 4: Experimental Study:

4.1 Dataset details:

We will use the Arrhythmia dataset from public database UCI Machine Learning Repository (Guvenir et al. 1998). This data was contributed by Dr. H. Altay Guvenir (1998). The dataset consists of clinical features extracted from 12-lead ECG signals, obtained from the IBM-Mt. Sinai Hospital Program. Arrhythmia is classified into 16 groups.

The ECG signals in the dataset consist of recordings obtained from 12 different electrodes or channels, which are known as DI, DII, DIII, AVL, AVR, AVF, and V1 to V6. These channels capture the electrical activity of the heart from various perspectives, forming a comprehensive 12-lead ECG system.

4.1.1 Data Analysis:

The dataset consists of 452 samples, each containing 279 features. Among these features, 206 are represented by numerical values, while the remaining features are categorical. The classes within the dataset are labelled as follows: Class 01 represents a 'normal' ECG, classes 02 to 15 correspond to different types of Arrhythmia, and class 16 includes unclassified cases. Out of the 280 features, 17 have zero values, and 5 features have missing values.

The target class in the dataset exhibits imbalanced data, with the majority class being patients with a normal heartbeat, having the highest number of observations.

Conversely, the minority class consists of patients with various types of Arrhythmia and is characterized by a significantly lower number of observations. Notably, there are no observations for patients with First-degree Atrioventricular block, second-degree AV block, or third-degree AV block (Class 11, 12, 13), which is an important observation to highlight.

Variance analysis shows that there are some features have significantly high variance value such as 'J', 'T', 'QRS', 'PRinterval', 'chV3_QRSTA' ..etc. These columns could potentially contribute more to the classification process by providing more discriminatory information. On the other hand, columns with very low variances, such as chDI_SPwave', 'chAVL_SPwaveAmp', 'chV5_SPwaveAmp', 'chV6_SPwaveAmp', etc., have little variability, which might indicate that they hold limited discriminatory power for classification purposes.

The correlation matrix reveals notable instances of high correlation among features, with correlation coefficients reaching 0.71, also indicates moderate correlations between certain features with coefficients around 0.3.

Subsequent analysis indicates the presence of outliers in a majority of the features; however, the frequency of outliers is relatively moderate and not excessively high.

4.1.2 Pre-processing:

In the context of this research investigation, we have opted not to eliminate the correlated attributes or Outliers. Correlated features often contain redundant information, but they can also contribute unique insights to the dataset. Removing them might result in a loss of valuable information, potentially leading to a less accurate or representative model.

If outliers are not true anomalies but rather valid data points, their removal can introduce bias into the dataset and subsequently bias the analysis and model outcomes.

The dataset consists of both integer and floating-point data types. Notably, one of the features (designated as "J") exhibits the highest proportion of missing values, accounting for 83% of the data. Following an extensive investigation into various techniques for handling missing data, as documented in the study (Feng et al. 2021) opted to employ the KNN imputation method to address the gaps in the dataset. Refrained from choosing mean or median imputation due to the absence of a normal distribution in the dataset. Furthermore, it's worth noting that no duplicate entries are present in the dataset.

A marked imbalance exists between the major and minority classes within the dataset. Specifically, there are 245 instances of patients with normal ECG readings, while 207 instances depict irregular ECG signals. In this study, data analysis will be performed without any dataset sampling. This decision is grounded in the understanding that applying sampling techniques to a medical dataset may not produce real-time results that accurately reflect the clinical context.

4.1.3 Feature Selection:

In this study, four feature selection techniques are employed, such as SelectKbest, PCA (Principal Component Analysis), Lasso Regression and RFE with Gradient boost classifier. These techniques considered for the following reasons.

(1) SelectKbest: Employs various statistical tests (ANOVA, chi-square, Information Gain**)** to measure relationship between features and Target. The statistical test calculates a score or a metric that reflects the strength of the relationship between the feature and the target variable. The higher the score, the more relevant or significant the feature is deemed to be (Brownlee 2019).

Since our research study focuses on classification, ANOVA F score is used as scoring function in selectKbest. F score is calculated by comparing the mean difference between features & Target for each feature. Higher F-scores indicate that the means of the target variable differ significantly across the levels of the feature, suggesting that the feature has potential predictive power for the target variable.

$$F = \frac{(Between\ Sum\ of\ Squares/\ Degree\ of\ Freedom\ for\ between)}{(Within\ Sum\ of\ Squares/Degree\ of\ Freedom\ for\ Within)} \qquad \textit{Equation 1}$$

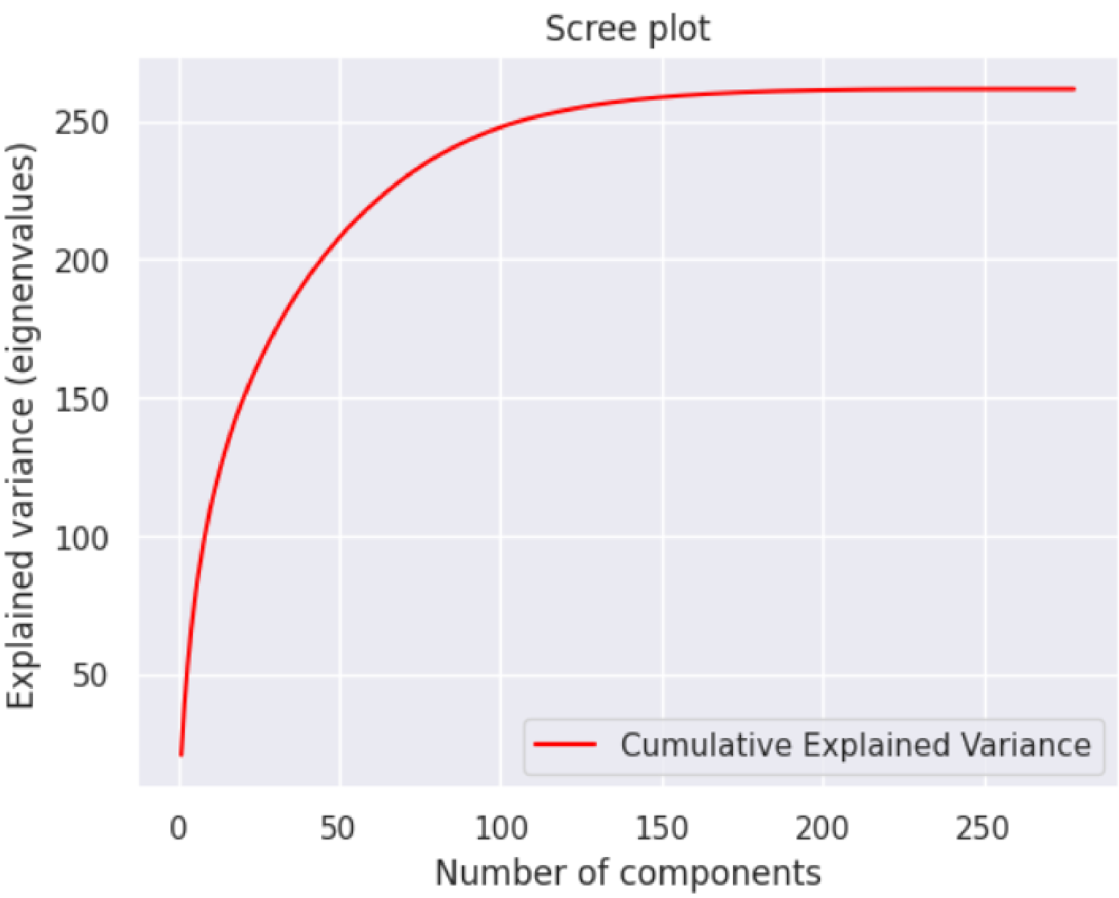
Subsequently, Kbest utilizes the F-scores to establish a ranking for each feature. The features with higher ranks, denoting enhanced potential for predicting the target variable, are subsequently selected. In our study, we will adopt this approach and choose the top 94 features from Kbest for further analysis.

PCA: It seeks to convert data with high dimensions into a reduced-dimensional representation, while preserving a significant portion of the original data's variability.PCA achieves this by finding a set of orthogonal (uncorrelated) axes, called principal components, along which the data exhibits the most variance.

The original dataset is standardized using a standard scaler before applying the PCA model to retain its variability, while ensuring that all features contribute equally to the PCA analysis.

In PCA, the parameter **n_components** refer to the number of principal components that to be retained after the dimensionality reduction (Brownlee 2019). N_components in PCA is initially set to Total number of features (279). PCA is then applied on the dataset to plot the cumulative explained variance against the number of principal components. The plot explains that the cumulative explained variance doesn't significantly increase after reaching 150 or the average cumulative explained variance was around 150 components. Hence, a substantial portion of the original data's variance can be retained with 150 components.

Figure 7: Variance Plot



Lasso Regression: It's known as Least Absolute Shrinkage and Selection Operator. In standard linear regression, the objective is to minimize the sum of squared residuals between the predicted values and the actual target values (Brownlee 2019). In lasso regression, an additional term is added to the objective function, which is proportional to the absolute values of the model coefficients (L1 norm). The objective function for lasso regression is:

*Objective = Sum of squared residuals + λ * Sum of absolute values of coefficients*
Equation 2

The regularization strength is modulated by the regularization parameter λ (lambda) (Brownlee 2019). A higher λ leads to more aggressive regularization. As λ increases, more coefficients are pushed towards zero, resulting in a simpler model with fewer non-zero coefficients. This effectively removes those features from the model, leading to automatic feature selection. Features with non-zero coefficients are considered relevant and retained in the model.

Lasso Regression model is applied on the dataset under experiment, 183 features are automatically selected by the model. These selected features are deemed to be the most relevant for making accurate predictions based on the Lasso's regularization process.

RFE (Gradient Boost (GB)): Determines relevant features through an iterative process that leverages the predictive power of the Gradient Boosting model. Gradient Boosting is an ensemble learning algorithm that combines multiple weak learners (typically decision trees) to create a strong predictive model (Brownlee 2019). This model is trained on the dataset and provides a measure of feature importance. Based on the feature importance scores, a selection criterion is applied to choose the most important features. GB Model learns complex relationships in the data allows it to identify the most relevant features for the classification task. In this research, we will select the top 94 features by considering their importance scores assigned by the Gradient Boosting model.

Table 4: Summary of Feature Subsets:

Description	**Feature selection Method**	**Number of Features**
Feature Set 1	SelectKbest	94
Feature Set 2	PCA	150
Feature Set 3	Lasso Regression	183
Feature Set 4	RFE (Gradient Boost)	94

4.1.4 Model Training:

In this study, we will utilize five different classification models, namely Random Forest (RF), Multi-Layer Perceptron (MLP), Naive Bayes (NB), Support Vector Machine (SVM), and k-Nearest Neighbors (KNN), to train and test the feature sets.

The feature subsets will be divided into training and test datasets using a standard ratio of 20% for testing and 80% for training with “Random state” as 2. To address the class imbalance between major and minority classes, passed an argument “Stratify” to maintain the proportion of different classes remains consistent in both the training and testing sets. The main purpose of using **stratify** is to prevent a situation where one of the classes is underrepresented or absent in the testing dataset, which could lead to biased or unreliable evaluation results.

The training dataset will be employed by the classification models to learn patterns, while the test dataset will be utilized to assess the models' performance.

The training and testing datasets will undergo normalization to mitigate the influence of higher numeric values.

Each feature subsets are individually trained & tested by classification models. And the best parameters for each classification models are determined by "GridSearchCV", a technique used to find the parameter values that results in highest accuracy.

Model performances are evaluated using the test dataset and also involves "Cross-validation" method with "K-fold" as 5. In this approach, the feature subsets undergo five random folds, each evaluating performance independently.

4.1.5 Evaluation:

Following metrics are used to evaluate the performance of ML models on the testing Data set:

Confusion Matrix, *Accuracy, Precision, Recall, F1 Score, & Kappa Statistics, Classification Report.*

Confusion matrix is widely used in multi-class classification. It generates a matrix, where each row of the confusion matrix corresponds to a true class, and each column corresponds to a predicted class. All other metrics are computed from confusion Matrix (Grandini et al. 2020).

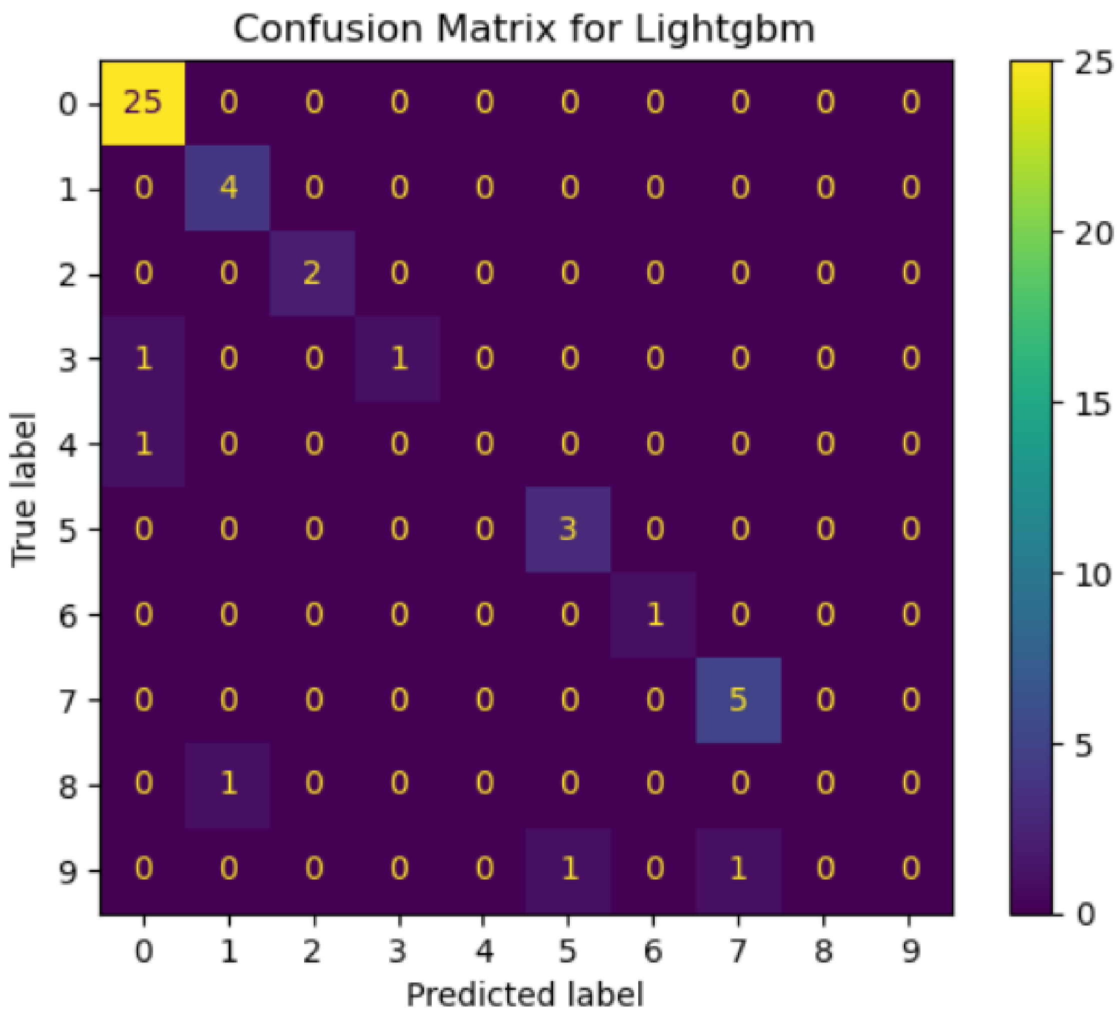


Figure 8: Confusion Matrix for LightGBM Model

Confusion Matrix for Multi-class classification is shown in Figure 8. Class 0 has true label & predicted label as 25, which shows the model has predicted all true positive cases, however Class 9 has 2 samples, but the model predicted class 9 as class 5 and 6. These are indeed False positive.

Precision, Recall, and F1 score are computed for each class separately, and the mean score of Precision and Recall is additionally determined, referred to as the Macro Average of Precision and Recall. Macro averaging ensures that each class has equal contribution to the final evaluation in an imbalanced dataset.

Macro average score tend to decrease if there are very a smaller number of samples in minority class. Hence the weighted average is employed to assess performance because it computes the average considering the relative proportions of each class within the dataset.

N= Total number of Classes.

Accuracy is calculated using the formula below:

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \qquad \text{Equation 3}$$

"True Positive" indicates the test correctly identifies someone has the disease.

"True Negative" signifies the test correctly shows someone doesn't have the disease.

"False Positive" result represents an erroneous prediction when a healthy patient is incorrectly labelled as diseased.

"False Negative" represents the most unfavourable scenario, where a patient with a disease is inaccurately classified as healthy.

Similarly, Specificity of individual classes are calculated from confusion Matrix as follows:

$$Specificity\ (or)\ Precision_{\ (Class\ x)} = \frac{TP_{Class\ x}}{TP_{Class\ x} + FP_{Class\ x}} \qquad \text{Equation 4}$$

Recall score for each individual classes are derived from the confusion Matrix as follows:

$$Sensitivity\ (or)\ Recall_{\ (Class\ x)} = \frac{TP_{class\ x}}{TP_{class\ x} + FN_{class\ x}} \qquad \text{Equation 5}$$

As it's a multi-class classification the weighted average can be calculated as follows:

Weighted Average =

$$\frac{(precision_{class\ x} * support_{class\ x}) + (precision_{class\ y} * support_{class\ y}) + \cdots ....(precision_{class\ N} * support_{class\ N})}{Total\ Number\ of\ Support\ or\ Samples}$$

Equation 6

The classification report in Python, typically generated using libraries like Scikit-Learn, presents these metrics in a clear and organized format for each class in the dataset. It provides not only the numerical values but also allows for easy visualization.

Cohen's Kappa Statistics (K) is used to measure the level of agreement between the predictions made by a model and the actual labels of the data. It quantifies how well the model's predictions match the true labels, while also considering the possibility of agreement occurring by chance (Grandini et al. 2020). And it's less sensitive to imbalances in the dataset.

$$K = \frac{c \times s - \sum_{k}^{K} p_{k \times t_k}}{s^2 - \sum_{k}^{K} p_{k \times t_k}} \qquad \text{Equation 7}$$

4.2 Project Risks

Table 5 Project Risks & Mitigation Plan

Risk Category	Risk description	Likelihood	Severity	Mitigation Plan
Project Risk	Poor quality Dataset will impact the accuracy of the results.	Low	High	Use approved medical research dataset. Use data visualization techniques to identify potential issues and outliers in the dataset. Use of Robust Machine learning algorithms to handle noisy or incomplete Data.
Schedule Risk	Task completion takes longer than estimated.	Low	High	Conducting regular progress reviews with Supervisor and status updates to identify any potential issues or delays early on.
Legal Risk	Data Privacy & Security of health data	High	High	Use of De-identified dataset
Technical Risk	Overfitting of Machine learning Model	High	High	To avoid overfitting, use techniques such as cross-validation, regularization, and early stopping.
	Underfitting of Machine Learning Model	High	High	Choose an appropriate algorithm, engineer additional features, or increase the complexity of the model.

	Data Imbalance	High	High	Use techniques such as oversampling, under sampling, or weighted loss functions to address this issue.
	Prediction Error	High	High	Utilize ensemble modelling, combine multiple individual models, known as base models or learners, to make predictions or classifications.
	Data Drifting (ECG data collected from different sources or at different time points may have variations).	Low	High	Robust techniques for domain adaptation or transfer learning may be required to address these challenges and ensure the models generalize well to new data.
	Biased Results	Low	High	Inaccurate or inconsistent annotations can introduce errors and biases into the analysis, leading to unreliable results. Use approved medical research data.

4.3 Professional issues

The proposed machine learning model for arrhythmia classification should follow the PSEL framework to ensure that it is designed, implemented, and maintained to a high standard of quality and reliability.

Professional Issues: The model should follow best practices, including coding standards & testing. It is also important to ensure that the software used, is secure, reliable, and scalable. Thorough testing should be performed on the ECG anomaly detection model to ensure that it meets the requirements and performs accurately. Conducting fraudulent research, falsifying data, or plagiarizing the work of others can compromise the validity and reliability of research findings.

As Intellectual property rights (IPR) are important to consider on comparative study of existing research papers, proper citation is critical component to ensure accuracy and integrity of research.

Social issues: ECG anomaly detection is primarily used for emergency medical attention. The suggested method shouldn't cause erroneous warnings. It is essential to minimize the occurrence of false positives to ensure accurate identification of genuine anomalies that require urgent medical attention.

Ethical Issues: Research is conducted in a manner that is just, fair, and respectful of individuals and communities. It is crucial to speculate whether completing the project and achieving the project aim are ethically justifiable. Online free medical data may contain sensitive personal information that must be protected to maintain the privacy and confidentiality of individuals whose data is being used. The dataset should be used in ethical manner.

Legal issues: Ensuring the protection of healthcare data is a critical priority, and this research must adhere to applicable data protection legislation (such as the UK Data Protection Act, 2018). This compliance is essential for upholding the privacy and confidentiality of individuals whose data is being utilized.

CHAPTER 5: Results & Discussion

Six classification models were trained using four different feature subsets. The parameters for these models were fine-tuned using the 'GridSearchCV' technique. The feature subset selected through Kbest, combined with the lightgbm classification model, achieved the highest accuracy at 81.3% using an 80:20 train-test split (80% Training Data, 20% testing Data). The experiment was conducted for both an 80:20 data split and a 90:10 data split.

RFE with Gradient Boost as classifier returned highest accuracy of 89% with 90:10 Data split.

Table 6: Summary of metrics with 80:20 Split:

Feature Selection	No. of Features	Accuracy (%)	Precision (%)	Recall (%)	Kappa Statistics
Kbest	94	81.3	77	81	0.7
PCA	150	73	76	74	0.6
Lasso Regression	183	80	83	80	0.67
RFE (GB)	94	79	77	78	0.67

Kbest feature subset has returned highest accuracy, precision, Recall, & Kappa Statistics. Research shows that removing correlated features from Kbest feature subset decreases accuracy of classification models. However, performance is not affected by applying PCA.

The models' accuracy exhibited improvement when the dataset was partitioned in a 90:10 ratio.

Table 7: Accuracy of models with 90:10 Train Test Split

Feature Selection	No. of Features	SVM	RF	MLP	KNN	NB	lightgbm
Kbest	94	78	87	80	72	8	83
PCA	150	78	65	67	72	50	73
Lasso Regression	183	70	78	70	63	10	80.43
RFE(GB)	94	83	85	83	69	10	89

LightGBM for RFE feature set has returned the highest accuracy. Accuracy of Random Forest classifier with Kbest Features returned second highest accuracy of 87%.

In terms of precision score, both the RFE and Lasso Regression yielded the top precision value of 82%. This score signifies the models' ability to accurately classify and predict positive instances while minimizing false positives, demonstrating their strong performance in this aspect of the evaluation. This balanced and high precision score is indicative of the models' efficacy in making precise and reliable positive predictions, contributing to their robust overall performance.

Despite employing a stratified train-test split, an imbalance persists in the class distribution between the training and testing datasets. This discrepancy arises due to the limited number of samples for specific classes, often as few as one or two instances. This disparity significantly influences the precision score of those specific minority classes, and also the average precision score of the model. To mitigate this, we have used Weighted-average precision score & cross validation metrics with 30-fold to ensure each fold maintains a similar class distribution and evaluated the model score. Cross-validation score helps to assess how well the model generalizes to different subsets of data.

Table 8: Weighted Average Precision Score of Classification Models with 90:10 Train Test Split

Feature Selection Method	No. of Features	SVM (%)	RF (%)	MLP (%)	KNN (%)	NB (%)	lightgbm (%)
Kbest	94	66	80	70	64	0.01	76
PCA	150	83	77	74	81	63	81
Lasso Regression	183	72	82	72	73	75	83
RFE (GB)	94	82	76	79	62	10	82

The recall metrics further validates that the RFE-based feature selection yields the highest performance when applied to the LightGBM model.

Table 9: Recall score of Classification Models with 90:10 Train Test Split

Feature Selection Method	No. of Features	SVM (%)	RF (%)	MLP (%)	KNN (%)	NB (%)	lightgbm (%)
Kbest	94	78	87	80	72	0.09	83
PCA	150	78	65	67	72	50	74
Lasso Regression	183	70	78	70	63	11	80
RFE(GB)	94	83	85	83	70	11	89

Likewise, when examining the Kappa Statistics, we observe a commendable score of 0.83 for the Wrapper Method. This score, which approaches the upper limit of 1, indicates a substantial level of agreement beyond what would be expected by chance alone. In other words, the Wrapper Method's Kappa Statistics underscores a strong level of concordance between its predictions and the actual outcomes, signifying its effectiveness in capturing meaningful patterns and making accurate classifications. This result further reinforces the Wrapper Method's efficacy and reliability in enhancing model performance.

Table 10: Kappa Statistics for Classification Models with 90:10 Train Test Split

Feature Selection	No. of Features	SVM	RF	MLP	KNN	NB	lightgbm
Kbest	94	0.65	0.8	0.7	0.53	-0.01	0.72
PCA	150	0.61	0.26	0.5	0.42	0.25	0.54
Lasso Regression	183	0.52	0.62	0.46	0.3	0.05	0.68
RFE (GB)	94	0.74	0.75	0.74	0.45	0.01	0.83

Cross-validation scores imply that LightGBM model is generalizing well. It's capturing patterns in the data that allow it to make accurate predictions on unseen data, which is crucial for real-world applications. Achieving an accuracy score of 87% across 30 stratified K-fold cross-validation runs is a very positive sign for the model's performance.

Table 11: Cross Validation Score of LightGBM Model on different Featureset

Feature Selection	No. of Features	Stratified Kfold (%)
Kbest	94	87
PCA	150	73
Lasso Regression	183	87
RFE (GB)	94	87

Factors Influencing the accuracy of the Model:

- The accuracy of the model tends to decrease when correlated features are eliminated from the dataset.
- Discarding categorical features that solely contain zero values also leads to a reduction in model accuracy. Thus, based on this experiment, it is advisable not to remove variables with constant values without domain expert knowledge and detailed analysis on the interdependencies between the features in high-dimensional dataset.
- Increasing number of decision trees in LightGBM classification model reduces the accuracy of model as it is overfitting on the testing data.

- Accuracy of Random Forest classifier is improved by hyper tuning the criterion to ‘Gini’, as it aims to minimize impurity during node splitting, leading to better separation of classes.
- Using a single hidden layer in a Multi-Layer Perceptron (MLP) yields higher accuracy compared to multiple layers, as excessive complexity causes overfitting by memorizing data rather than learning patterns.
- The highest accuracy of 48% on PCA-selected features suggests that the Naive Bayes model might struggle with the complexity of the multi-class classification task.

The table below presents a comparison of the proposed system's accuracy with that of State-of-the-Art methods, utilizing the same UCI dataset for experimentation.

Table 12: Comparison of Accuracy with State-of-the-Art Methods:

Author	Method	Train\Test Split	No. of selected features	Accuracy (%)
	Proposed (RFE (estimator: Gradient Boost) + LightGBM)	90 - 10	94	89
(Ayar et al. 2023)	NSICA Algorithm	10-Fold CV	27	83.76
(Ayar et al. 2021)	CDC+C4.5 algorithm	80-20	78	88
(R. Abirami, Vincent 2019)	Ensemble ML	75-25		85
(Sean Shensheng Xu et al. 2017)	FDR with DNN	10-Fold CV	236	82.96
(Khan et al. 2015)	KNN	20-Fold CV	60	88.2
(Fazel et al. 2014)	Linear SVM	10-Fold CV	15	73
(Mitra, Samanta 2013)	IBPLN +LM	68-32	18	87.71
(Namsrai et al. 2013)	NB	70-30	205	70.5
(M. Jadhav et al. 2011)	Modular NN	90-10	198	78.89
(Kohli et al. 2010)	SVM based approach	50-50	166	73.8

CHAPTER 6: Conclusion & Future Work

This study suggests that utilizing Recursive Feature Elimination (RFE) in conjunction with a Gradient Boosting (GB) classifier is the optimal choice for feature selection in high-dimensional datasets. Furthermore, when dealing with multi-class classification in imbalanced datasets, the LightGBM classification method proves to be the most appropriate solution.

Specifically, this method achieves an impressive accuracy rate of 89%, signifying the proportion of correctly predicted instances among the total. Furthermore, it demonstrates an elevated precision level of 82%, indicating the capacity of the model to accurately identify positive instances among those predicted as positive. Additionally, the method showcases a peak recall score of 89%, illustrating its ability to capture a high percentage of actual positive instances within the dataset.

These achievements are particularly notable given the complex nature of the data, which is characterized by having a high number of features (high-dimensional) and an unequal distribution of classes (imbalanced) with 16 types of Arrhythmias. Successfully navigating such challenges highlights the effectiveness of the proposed approach in handling intricate real-world scenarios.

However, as part of ongoing research and exploration, it is recommended to direct future efforts toward enhancing the precision of classification models, especially when dealing with multi-class classification tasks for the target variable. In multi-class scenarios, the objective is to ensure that the model not only accurately identifies positive instances within each class but also minimizes misclassifications and false positives.

This emphasis on precision is crucial in scenarios where the consequences of misclassification can be significant, such as in medical diagnoses or fraud detection.

While Recursive Feature Elimination (RFE) stands out as an effective method for feature selection, it's important to acknowledge that its processing time can be comparatively higher when compared to other techniques like KBest, Lasso regression, and PCA. This presents an avenue for future research to delve into, as optimizing the efficiency of RFE could potentially yield substantial benefits in terms of both time and resource utilization.

I. Appendices:

Definitions for all terms are referenced from Cleveland Clinic Website (Cleveland Clinic medical professional n.d.)

Term	**Definition**
Amplitude	Magnitude of the electrical potentials generated during different phases of the cardiac cycle.
Arrhythmia	Irregularities in the timing and pattern of P waves, QRS complexes, and T waves.
Atrial fibrillation	Atrial fibrillation is an irregular heart rhythm that begins in your heart’s upper chambers (atria). (Cleveland Clinic medical professional n.d.)
ECG rhythm classifier	Aims to classify the overall rhythm or pattern of the heartbeats
Heart Rate	Time intervals between successive R peaks (R-R intervals)
Heartbeat Classifier	Focuses on detecting individual heartbeats within the ECG signal
Myocardial Infraction	A heart attack (medically known as a myocardial infarction) is a deadly medical emergency where your heart muscle begins to die because it isn’t getting enough blood flow (Cleveland Clinic medical professional n.d.).
P wave	The P wave represents atrial depolarization
QRS Complex	Represents ventricular depolarization
ST Segment and T Wave Abnormalities	The ST segment is the period between ventricular depolarization (QRS complex) and repolarization (T wave). Deviations in the ST segment or T wave morphology can indicate ischemia, injury, or myocardial infarction (Cleveland Clinic medical professional n.d.).
T wave	Represents ventricular repolarization.

II. References:

Abubaker, M., 2023. Detection of Cardiovascular Diseases in ECG Images Using Machine Learning and Deep Learning Methods. *IEEE Transactions on Artificial Intelligence*, 4(2), pp.1–1. https://doi.org/10.1109/tai.2022.3159505.

Guvenir,H. et al., 1998. Arrhythmia. UCI Machine Learning Repository. [online]. Available at: https://doi.org/10.24432/C5BS32 [Accessed 7 July 2023].

Appelboom, G. et al., 2014. Smart wearable body sensors for patient self-assessment and monitoring [online]. *Archives of Public Health*, 72(1). Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4166023/.

Ardeti, V.A. et al., 2023. An overview on state-of-the-art electrocardiogram signal processing methods: Traditional to AI-based approaches. *Expert Systems with Applications*, 217, p.119561. https://doi.org/10.1016/j.eswa.2023.119561.

Ashfaq Khan, M., Kim, Y., 2021. Cardiac Arrhythmia Disease Classification Using LSTM Deep Learning Approach. *Computers, Materials & Continua*, 67(1), pp.427–443. https://doi.org/10.32604/cmc.2021.014682.

Ayar, M. et al., 2021. Chaotic-based divide-and-conquer feature selection method and its application in cardiac arrhythmia classification. *The Journal of Supercomputing*, 78(4), pp.5856–5882. https://doi.org/10.1007/s11227-021-04108-5.

Ayar, M. et al., 2023. NSICA: Multi-objective imperialist competitive algorithm for feature selection in arrhythmia diagnosis. *Computers in Biology and Medicine*, 161, pp.107025–107025. https://doi.org/10.1016/j.compbiomed.2023.107025

Badawi, A.A., Ahmed Noah Badr, Khalid Elgazzar, 2021. ECG Real-time Monitoring and Heart Anomaly Detection Reimagined. *The Internet of Things*. https://doi.org/10.1109/wf-iot51360.2021.9595639.

Banaee, H., Ahmed, M., Loutfi, A., 2013. Data Mining for Wearable Sensors in Health Monitoring Systems: A Review of Recent Trends and Challenges. *Sensors*, 13(12), pp.17472–17500. https://doi.org/10.3390/s131217472.

Brownlee, J., 2019. How to Choose a Feature Selection Method for Machine Learning [online]. *Machine Learning Mastery*. Available at: https://machinelearningmastery.com/feature-selection-with-real-and-categorical-data/.

Butkevičiūtė, E., Bikulčienė, L., Blažauskas, T., 2022. The unsupervised pattern recognition for the ECG signal features detection. *Biomedical Signal Processing and Control*, 78, p.103947. https://doi.org/10.1016/j.bspc.2022.103947.

Cleveland Clinic medical professional, Sinus Arrhythmia: Causes, Symptoms and Treatment [online]. *Cleveland Clinic*. Available at: https://my.clevelandclinic.org/health/diseases/21666-sinus-arrhythmia.

Dissanayake, K., Md Johar, M.G., 2021. Comparative Study on Heart Disease Prediction Using Feature Selection Techniques on Classification Algorithms Liao, T. W., ed. *Applied Computational Intelligence and Soft Computing*, 2021, pp.1–17. https://doi.org/10.1155/2021/5581806.

Evangelia Myrovali et al., 2022. Identifying patients with paroxysmal atrial fibrillation from sinus rhythm ECG using random forests. *Identifying patients with paroxysmal atrial fibrillation from sinus rhythm ECG using random forests*, 213, pp.118948–118948. https://doi.org/10.1016/j.eswa.2022.118948.

Fazel, A., Haider, B., Algharbi, F., 2014. CS229: Machine Learning [online]. *cs229.stanford.edu*. Available at: https://cs229.stanford.edu/proj2014/AlGharbi%20Fatema,%20Fazel%20Azar,%20Haider%20Batool,%20Cardiac%20Arrhythmias%20Patients.pdf.

Feng, S., Hategeka, C., Grépin, K.A., 2021. Addressing missing values in routine health information system data: an evaluation of imputation methods using data from the Democratic Republic of the Congo during the COVID-19 pandemic. *Population Health Metrics*, 19(1). https://doi.org/10.1186/s12963-021-00274-z.

gajawada, sampath kumar, 2019. ANOVA for Feature Selection in Machine Learning [online]. *Medium*. Available at: https://towardsdatascience.com/anova-for-feature-selection-in-machine-learning-d9305e228476.

Ganeshkumar, R., YSKumaraswamy, Dr., 2012. Investigating Cardiac Arrhythmia in ECG using Random Forest Classification. *International Journal of Computer Applications*, 37(4), pp.31–34. https://doi.org/10.5120/4599-6557.

Gao, X. et al., 2021. A multiclass classification using one-versus-all approach with the differential partition sampling ensemble. *Engineering Applications of Artificial Intelligence*, 97, p.104034. https://doi.org/10.1016/j.engappai.2020.104034.

Geweid, G.G.N., Chen, J.D.Z., 2022. Automatic Classification of Atrial Fibrillation from Short Single-Lead ECG Recordings using a Hybrid Approach of Dual Support Vector Machine. *Expert Systems with Applications*, p.116848. https://doi.org/10.1016/j.eswa.2022.116848.

Grandini, M., Bagli, E., Visani, G., 2020. Metrics for Multi-Class Classification: An Overview [online]. *arXiv:2008.05756 [cs, stat]*. Available at: https://arxiv.org/abs/2008.05756.

He, Z. et al., 2023. A novel unsupervised domain adaptation framework based on graph convolutional network and multi-level feature alignment for inter-subject ECG classification. *A novel unsupervised domain adaptation framework based on graph convolutional network and multi-level feature alignment for inter-subject ECG classification*, 221, pp.119711–119711. https://doi.org/10.1016/j.eswa.2023.119711.

Hidde Bleijendaal et al., 2021. Computer versus cardiologist: Is a machine learning algorithm able to outperform an expert in diagnosing a phospholamban p. Arg14del mutation on the electrocardiogram? *Heart Rhythm*, 18(1), pp.79–87. https://doi.org/10.1016/j.hrthm.2020.08.021.

Hu, R., Chen, J., Zhou, L., 2022. A transformer-based deep neural network for arrhythmia detection using continuous ECG signals. *Computers in Biology and Medicine*, 144, p.105325. https://doi.org/10.1016/j.compbiomed.2022.105325.

Jha, C.K., Kolekar, M.H., 2020. Cardiac arrhythmia classification using tunable Q-wavelet transform based features and support vector machine classifier. *Biomedical Signal Processing and Control*, 59, p.101875. https://doi.org/10.1016/j.bspc.2020.101875.

Kachuee, M., Fazeli, S., Sarrafzadeh, M., 2018. ECG Heartbeat Classification: A Deep Transferable Representation [online]. *2018 IEEE International Conference on Healthcare Informatics (ICHI)*, pp.443–444. Available at: https://arxiv.org/abs/1805.00794 [Accessed 26 May 2020].

Kashou, A.H., May, A.M., Noseworthy, P.A., 2023. Comparison of two artificial intelligence-augmented ECG approaches: Machine learning and deep learning. *Journal of Electrocardiology*. https://doi.org/10.1016/j.jelectrocard.2023.03.009.

Khan, A. et al., 2015. Identifying best feature subset for cardiac arrhythmia classification. https://doi.org/10.1109/sai.2015.7237188.

Khan, A.H., Hussain, M., Malik, M.K., 2021. Cardiac Disorder Classification by Electrocardiogram Sensing Using Deep Neural Network Khan, A., ed. *Complexity*, 2021, pp.1–8. https://doi.org/10.1155/2021/5512243.

Kleyko, D., Osipov, E., Wiklund, U., 2020. A Comprehensive Study of Complexity and Performance of Automatic Detection of Atrial Fibrillation: Classification of Long ECG Recordings Based on the PhysioNet Computing in Cardiology Challenge 2017. *Biomedical Physics & Engineering Express*, 6(2), p.025010. https://doi.org/10.1088/2057-1976/ab6e1e.

Kohli, N., Verma, N.K., Roy, A., 2010. SVM based methods for arrhythmia classification in ECG. https://doi.org/10.1109/iccct.2010.5640480.

Kusumoto, F., 2020. *ECG interpretation: from pathophysiology to clinical application* [eBook]. Cham: Springer. Available at: https://doi.org/10.1007/978-3-030-40341-6.

Li, H. et al., 2015. Novel ECG Signal Classification Based on KICA Nonlinear Feature Extraction. *Circuits, Systems, and Signal Processing*, 35(4), pp.1187–1197. https://doi.org/10.1007/s00034-015-0108-3.

Li, Z., Zhang, H., 2023. Fusing deep metric learning with KNN for 12-lead multi-labelled ECG classification. , 85, pp.104849–104849. https://doi.org/10.1016/j.bspc.2023.104849.

Liu, Z. et al., 2020. Wavelet Scattering Transform for ECG Beat Classification. *Computational and Mathematical Methods in Medicine*, 2020, pp.1–11. https://doi.org/10.1155/2020/3215681.

Lown, M. et al., 2020. Machine learning detection of Atrial Fibrillation using wearable technology Tolkacheva, E. G., ed. *PLOS ONE*, 15(1), p.e0227401. https://doi.org/10.1371/journal.pone.0227401.

Lyon, A. et al., 2018. Computational techniques for ECG analysis and interpretation in light of their contribution to medical advances [online]. *Journal of The Royal Society Interface*, 15(138), p.20170821. Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5805987/.

M Devadas, R., 2021. *CARDIAC ARRHYTHMIA CLASSIFICATION USING SVM, KNN AND NAIVE BAYES ALGORITHMS* [eBook]. International Research Journal of Engineering and Technology (IRJET). Available at: https://www.irjet.net/archives/V8/i5/IRJET-V8I5721.pdf.

M.Jadhav, S., L.Nalbalwar, S., A.Ghatol, A., 2011. MODULAR NEURAL NETWORK BASED ARRHYTHMIA CLASSIFICATION SYSTEM USING ECG SIGNAL DATA [online]. *Research Gate*. Available at: https://www.researchgate.net/profile/Ashok-Ghatol/publication/265352173_Modular_neural_network_based_arrhythmia_classification_system_using_ECG_signal_data/links/56e5939008aedb4cc8ae72d8/Modular-neural-network-based-arrhythmia-classification-system-using-ECG-signal-data.pdf.

Masud Shah Jahan et al., 2022. Short-term atrial fibrillation detection using electrocardiograms: A comparison of machine learning approaches. *International Journal of Medical Informatics*, 163, pp.104790–104790. https://doi.org/10.1016/j.ijmedinf.2022.104790.

Merdjanovska, E., Rashkovska, A., 2022. Comprehensive survey of computational ECG analysis: Databases, methods and applications [online]. *Expert Systems with Applications*, 203, p.117206. Available at: https://www.sciencedirect.com/science/article/pii/S0957417422005917 [Accessed 7 September 2022].

Mincholé, A. et al., 2019. Machine learning in the electrocardiogram. *Journal of Electrocardiology*, 57, pp.S61–S64. https://doi.org/10.1016/j.jelectrocard.2019.08.008.

Mitra, M., Samanta, R.K., 2013a. Cardiac Arrhythmia Classification Using Neural Networks with Selected Features. *Procedia Technology*, 10, pp.76–84. https://doi.org/10.1016/j.protcy.2013.12.339.

Mitra, M., Samanta, R.K., 2013b. Cardiac Arrhythmia Classification Using Neural Networks with Selected Features. *Procedia Technology*, 10, pp.76–84. https://doi.org/10.1016/j.protcy.2013.12.339.

Mousavi, S., Afghah, F., 2019. Inter- and Intra- Patient ECG Heartbeat Classification for Arrhythmia Detection: A Sequence to Sequence Deep Learning Approach [online]. *IEEE Xplore*, pp.1308–1312. Available at: https://ieeexplore.ieee.org/abstract/document/8683140 [Accessed 9 December 2020].

Mustaqeem, A., Anwar, S.M., Majid, M., 2018. Multiclass Classification of Cardiac Arrhythmia Using Improved Feature Selection and SVM Invariants. *Computational and Mathematical Methods in Medicine*, 2018, pp.1–10. https://doi.org/10.1155/2018/7310496.

Namsrai, E. et al., 2013. A Feature Selection-based Ensemble Method for Arrhythmia Classification. *Journal of Information Processing Systems*, 9(1), pp.31–40. https://doi.org/10.3745/jips.2013.9.1.031.

Nesaragi, N. et al., 2022. Automated diagnosis of coronary artery disease using scalogram-based tensor decomposition with heart rate signals. *Medical Engineering & Physics*, 110, p.103811. https://doi.org/10.1016/j.medengphy.2022.103811.

Nezamabadi, K. et al., 2023. Unsupervised ECG Analysis: A Review [online]. *IEEE Reviews in Biomedical Engineering*, 16, pp.208–224. Available at: https://authors.library.caltech.edu/114361/1/Unsupervised_ECG_Analysis_A_Review.pdf [Accessed 10 March 2023].

Ohn, M., Souza, U., Ohn, K., 2019. A qualitative study on negative attitude toward electrocardiogram learning among undergraduate medical students. *Tzu Chi Medical Journal*, 0(0), p.0. https://doi.org/10.4103/tcmj.tcmj_91_19.

Pant, H., Dhanda, H.K., Taran, S., 2022. Sleep apnea detection using electrocardiogram signal input to FAWT and optimize ensemble classifier. *Measurement*, 189, p.110485. https://doi.org/10.1016/j.measurement.2021.110485.

Qiang, W. et al., 2022. TSVM-M3: Twin support vector machine based on multi-order moment matching for large-scale multi-class classification. , 128, pp.109506–109506. https://doi.org/10.1016/j.asoc.2022.109506.

Quer, G. et al., 2021. Machine Learning and the Future of Cardiovascular Care. *Journal of the American College of Cardiology*, 77(3), pp.300–313. https://doi.org/10.1016/j.jacc.2020.11.030.

R. Abirami, Vincent, P., 2019. Cardiac Arrhythmia Detection Using Ensemble of Machine Learning Algorithms. *Advances in intelligent systems and computing*, pp.475–487. https://doi.org/10.1007/978-981-15-0184-5_41.

Ramasamy, K., Balakrishnan, K., Velusamy, D., 2022. Detection of cardiac arrhythmias from ECG signals using FBSE and Jaya optimized ensemble random subspace K-nearest neighbor algorithm. *Biomedical Signal Processing and Control*, 76, p.103654. https://doi.org/10.1016/j.bspc.2022.103654.

Sáez, J.A., Krawczyk, B., Woźniak, M., 2016. Analyzing the oversampling of different classes and types of examples in multi-class imbalanced datasets. *Pattern Recognition*, 57, pp.164–178. https://doi.org/10.1016/j.patcog.2016.03.012.

Sean Shensheng Xu, Mak, M.-W., Chun Shun Cheung, 2017. Deep neural networks versus support vector machines for ECG arrhythmia classification. https://doi.org/10.1109/icmew.2017.8026250.

Sellami, A., Hwang, H., 2019. A robust deep convolutional neural network with batch-weighted loss for heartbeat classification. *Expert Systems with Applications*, 122, pp.75–84. https://doi.org/10.1016/j.eswa.2018.12.037.

Serhani, M.A. et al., 2020. ECG Monitoring Systems: Review, Architecture, Processes, and Key Challenges [online]. *Sensors*, 20(6), p.1796. Available at: https://www.mdpi.com/1424-8220/20/6/1796/htm.

Silva-Palacios, D., Ferri, C., Ramírez-Quintana, M.J., 2017. Improving Performance of Multiclass Classification by Inducing Class Hierarchies. *Procedia Computer Science*, 108, pp.1692–1701. https://doi.org/10.1016/j.procs.2017.05.218.

Sun, L. et al., 2022. Feature reduction for imbalanced data classification using similarity-based feature clustering with adaptive weighted K-nearest neighbors. *Information Sciences*, 593, pp.591–613. https://doi.org/10.1016/j.ins.2022.02.004.

Sunny, J.S. et al., 2022. Anomaly Detection Framework for Wearables Data: A Perspective Review on Data Concepts, Data Analysis Algorithms and Prospects. *Sensors*, 22(3), p.756. https://doi.org/10.3390/s22030756.

Sundaram, Gokila. et al., 2023. Comparative study on Machine Learning Models for Multi-class classification of Arrhythmia.<Journal Name: To be concluded>. <Date: To be submitted in Sep 2023>

Wikipedia Contributors, 2020. Arrhythmia [online]. *Wikipedia*. Available at: https://simple.wikipedia.org/wiki/Arrhythmia [Accessed 3 May 2022].

Wikipedia Contributors, 2019. Electrocardiography [online]. *Wikipedia*. Available at: https://en.wikipedia.org/wiki/Electrocardiography [Accessed 19 April 2019].

Zhang, X. et al., 2020. Automated detection of cardiovascular disease by electrocardiogram signal analysis: a deep learning system. *Cardiovascular Diagnosis and Therapy*, 10(2), pp.227–235. https://doi.org/10.21037/cdt.2019.12.10.

Zhao, W. et al., 2022. Machine learning for distinguishing right from left premature ventricular contraction origin using surface electrocardiogram features. *Heart Rhythm*, 19(11), pp.1781–1789. https://doi.org/10.1016/j.hrthm.2022.07.010.

Zou, C. et al., 2022. Heartbeat Classification by Random Forest With a Novel Context Feature: A Segment Label. *IEEE Journal of Translational Engineering in Health and Medicine*, 10, pp.1–8. https://doi.org/10.1109/jtehm.2022.3202749.