

A  
MAJOR PROJECT REPORT  
ON

Identifying Cardiovascular Disorder Using  
ECG Image Analysis

*Submitted in partial fulfilment of the requirements for the award of the degree of*

BACHELOR OF TECHNOLOGY

*By*

Devansh Rautela (2018314)  
Anmool Kumar (2018185)  
Mohit Bajeli (2018491)

*To*

Mrs. Himadri Vaidya

Associate Professor

Department Of Computer Science & Engineering



DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING  
GRAPHIC ERA HILL UNIVERSITY, DEHRADUN  
DEHRADUN – 248002 (INDIA)

# CERTIFICATE

We, **Devansh Rautela, Anmool Kumar** and **Mohit Bajeli**, students of **B.TechCSE VIII Semester**, Department of Computer Science and Engineering, Graphic Era Hill University, Dehradun, declare that the technical project work entitled “Identifying cardiovascular disorder using ECG image analysis” has been carried out by us and submitting in partial fulfilment of the course requirements for the award of degree in Bachelor Of Technology of Graphic Era Hill University, Dehradun during the academic year **2023-2024**. This synopsis has not been submitted to any other university for the award of any other degree or diploma.

Place: Dehradun

Date: 30/05/2024

# ACKNOWLEDGEMENT

Here by we are submitting the project report on “Identifying cardiovascular disorder using ECG image analysis”, as per the scheme of Graphic Era Hill University, Dehradun. We express the deepest sense of gratitude to Mrs. Himadri Vaidya, Asst. Professor, Department of Computer Science and Engineering for the invaluable guidance extended at every stage and in every possible way. I am very much thankful to all the faculty members of the Department of Computer Science and Engineering, friends and my parents for their constant encouragement, support and help throughout the period of project conduction.

DEVANSH RAUTELA

Roll No : 2018314

Sec : C

ANMOOL KUMAR

Roll No : 2018185

Sec : C

MOHIT BAJELI

Roll No : 2018491

Sec : E

# ABSTRACT

This research focuses on digitizing paper-based electrocardiogram (ECG) charts to train machine learning models for detecting cardiovascular diseases (CVD), particularly abnormal heartbeat, myocardial infarction, history of myocardial infarction, and normal cases, due to dataset limitations. The digitization involves converting ECG charts into digital formats, allowing advanced data processing and machine learning techniques. Preprocessing steps include Gaussian filtering for noise reduction, thresholding to highlight relevant signals, grayscale conversion for simplicity, and resizing ECG images into a standardized format of 13 leads, subsequently converted into one-dimensional signals for analysis. The study utilizes Support Vector Machine (SVM), K-nearest neighbors (KNN), and Logistic Regression, and proposes an ensemble technique using a Voting-Based Classifier to combine individual predictions, achieving an accuracy of 92.5%, recall of 91%, precision of 92%, and an F1-score of 92%. This approach outperforms existing methods, automating cardiac abnormality detection, reducing manual examination, and leading to faster and more accurate diagnoses, ultimately improving patient outcomes. Despite promising results, limitations include the need for more comprehensive datasets and refining preprocessing and machine learning techniques. In conclusion, digitizing ECG records and applying machine learning can revolutionize CVD detection, providing a valuable tool for early diagnosis and enhanced healthcare.

Index Terms—Electrocardiogram (ECG), Cardiovascular diseases (CVD), Support Vector Machine, Voting Based Classifier.

# INDEX

<b>ACKNOWLEDGMENT</b>	<b>III</b>
<b>ABSTRACT</b>	<b>IV</b>
<b>LIST OF TABLES</b>	<b>VI</b>
<b>LIST OF FIGURES</b>	<b>VII</b>
<b>ABBREVIATIONS</b>	<b>VIII</b>
<b>1. INTRODUCTION</b>	<b>1-3</b>
1.1 Importance Of Early Detection	
1.2 ECG Electrode Placement and Function	
1.3 Challenges in Manual ECG Interpretation	
1.4 Research Focus and Methodology	
1.5 Machine Learning Models Used	
1.6 Ensemble Techniques for Enhanced Prediction	
1.7 Research Contribution and Implications	
<b>2. LITERATURE SURVEY</b>	<b>4-7</b>
<b>3. METHODOLOGY</b>	<b>8-26</b>
3.1 Dataset Description	
3.2 Image Processing Techniques	
3.3 Lead Extraction	
3.4 Preprocessing Techniques on Leads	
3.5 Signal Extraction and Processing	
3.6 Dimensionality Reduction	
3.7 Model Selection	
3.8 Ensemble Technique Voting Classifier	
<b>4. RESULTS AND ANALYSIS</b>	<b>27-42</b>
4.1 Image preprocessing and lead extraction	
4.2 Model evaluation and performance comparison	
4.3 Discussion	
<b>5. Conclusion and Future Scope</b>	<b>43-44</b>
<b>APPENDIX</b>	<b>45-56</b>
<b>REFERENCES</b>	<b>57-60</b>
<b>CONFERENCE ACCEPTED</b>	<b>61-62</b>

# LIST OF TABLES

TABLE 1: Distribution of classes in ECG dataset	10
TABLE 2: Algorithm with single and 12-lead ECG Image	34
TABLE 3: KNN Performance Matrix	35
TABLE 4: Logistic Regression Performance Matrix	35
TABLE 5: SVM Performance Matrix	36
TABLE 6: KNN Performance Matrix (12 Leads)	36
TABLE 7: Logistic Regression Performance Matrix (12 Leads)	37
TABLE 8: SVM Performance Matrix (12 Leads)	37
TABLE 9: XGBoost Performance Matrix	38
TABLE 10: Accuracy of Various Algorithm	39
TABLE 11: Perposed Performance Matrix	41

# LIST OF FIGURES

Fig 1. Single Lead Grayscale Conversion	12
Fig 2. Divided Leads (1-6 Lead)	14
Fig 3. Divided Leads (7-12)	14
Fig 4. Gaussian filtration	17
Fig 5. Countour signal	19
Fig 6. Signal to 1d Result	20
Fig 7. Data Processing	22
Fig 8. Process Flow	25
Fig 9. Performance comparison between 12-lead & Single lead ECG Image	34
Fig 10. Performance of Models & Techniques	38
Fig 11. Uploading Image	54
Fig 12. GrayScale Filtering	54
Fig 13. Lead Divison	55
Fig 14. Thresholding	55
Fig 15. Extraction of Signals	56
Fig 16. Conversion into 1D signals	56

# **ABBREVIATIONS**

DL - Deep Learning

AI - Artificial Intelligence

ECG -Electrocardiogram

KNN - K-Nearest Neighbor

ML - Machine Learning

NB - Naive Bayes

SVM - Support Vector Machine

MI -Myocardial Infraction

AHB - Abnormal Heartbeat

PM – Previous History of MI

CVD- Cardiovascular Disorder



# CHAPTER 1

## INTRODUCTION

Cardiovascular diseases (CVDs) are disorders that affect the heart and blood vessels, and they are categorized as non-communicable diseases. These conditions are the leading cause of death worldwide, causing an estimated 17.9 million deaths annually. According to a 2020 report by the World Health Organization (WHO), CVDs account for 31% of all global deaths. Heart attacks and strokes are responsible for approximately 85% of these deaths. High-risk individuals often exhibit elevated blood pressure, high glucose levels, and imbalances in their electrocardiogram (ECG) or electrocardiograph (EKG).

The overarching goal of this project is to bridge the gap between traditional identification of cardiovascular diseases and contemporary machine learning techniques. By leveraging the predictive potential of machine learning algorithms, we aim to develop robust models capable of classifying individuals into four distinct categories: Myocardial Infarction (MI), patients with abnormal heartbeats, normal individuals, and patients with a history of MI by detecting irregular heart rhythms using ECG signals. Using ECG, one can achieve early detection of CVDs, and can be easily prevented using guided treatment, ultimately reducing CVD events.

The significance of this project lies in its potential to revolutionize mental health diagnostics and interventions. By integrating sophisticated algorithms such as K-Nearest Neighbors (KNN), Support Vector Machine (SVM), Logistic Regression, and the powerful decision tree-based XGBoost algorithm. To enhance the predictive capabilities and robustness of our models, we incorporate ensemble techniques such as the Voting Classifier. This ensemble approach combines the strengths of individual models, including the newly introduced XGBoost, to obtain the highest accuracy results.

### 1.1 Importance of Early Detection

Early identification of those at the highest risk for CVDs can significantly help in preventing these diseases. An electrocardiogram (ECG) is a non-invasive test that detects the electrical activity of the heart using sensors placed on the skin over the chest. Typically, this involves the application of 10 electrode points strategically positioned to create a 12-lead ECG. This setup provides a comprehensive assessment of the heart's electrical conduction from different perspectives.

### 1.2 ECG Electrode Placement and Function

Research outlines that the 10 electrode points consist of six precordial leads placed on the chest and four limb leads placed on the arms and legs. These electrodes capture the electrical signals generated by the heart from multiple angles, allowing for a refined and accurate analysis of cardiac function. The limb leads and augmented leads are derived from the right arm (RA), left arm (LA), right leg (RL), and left leg (LL) electrodes. This process is quick, painless, and capable of detecting a variety of heart problems.

### 1.3 Challenges in Manual ECG Interpretation

Manually interpreting ECG recordings is vital but also prolonged, monotonous, and requires expertise. This manual examination can be prone to errors and delays, which underscores the need for automated solutions. Recent advancements in machine and deep learning offer promising methods for the automatic identification of cardiovascular diseases, as discussed in various research papers.

### 1.4 Research Focus and Methodology

This research focuses on converting non-uniform ECG images into standardized 12-lead ECG images and structuring the resultant dataset. This dataset is used to identify cardiovascular disorders and classify individuals into four categories: myocardial infarction (MI), patients with abnormal heartbeats, normal individuals, and patients with a history of MI. The study employs a diverse set of machine learning models, including K-Nearest Neighbors (KNN), Support Vector Machine (SVM), Logistic Regression, and the decision tree-based XGBoost algorithm.

### 1.5 Machine Learning Models Used

- i. **K-Nearest Neighbors (KNN):** KNN is a simple, yet effective algorithm used for classification tasks. It works by comparing a test sample with training samples using distance metrics and assigns the test sample to the most common class among its k nearest neighbors.
- ii. **Support Vector Machine (SVM):** SVM is effective in high-dimensional spaces and works well for classification tasks where the decision boundary is complex. It constructs a hyperplane or set of hyperplanes in a high-dimensional space to separate different classes.
- iii. **Logistic Regression:** Logistic Regression is a statistical method for analyzing datasets with one or more independent variables that determine an outcome. It is robust and interpretable, making it useful for binary classification problems.
- iv. **XGBoost:** XGBoost is an optimized gradient boosting machine learning algorithm. It is efficient and scalable, known for its performance in handling structured datasets and decision tree-based learning.

### 1.6 Ensemble Techniques for Enhanced Prediction

To enhance the predictive capabilities and robustness of these models, the research incorporates ensemble techniques such as the Voting Classifier. This ensemble approach combines the strengths of individual models, including the newly introduced XGBoost, to achieve the highest accuracy results. Ensemble techniques help in reducing overfitting, variance, and bias, leading to more reliable predictions.

## **1.7 Research Contributions and Implications**

The combination of these models, trained with diverse and high-quality data, contributes significantly to the research and development landscape across various fields. This research not only addresses practical challenges in medical image processing but also highlights the collective impact of machine learning and ensemble techniques. The study demonstrates how these methods improve diagnosis and decision-making abilities, enhancing patient care and outcomes. The incorporation of XGBoost, in particular, is noted for its efficiency in handling structured datasets, which further enhances the research findings and creates opportunities for innovative applications in healthcare and beyond.

# CHAPTER 2

## LITERATURE SURVEY

In the literature review, several studies have examined different approaches to identifying abnormal heartbeats using multiple techniques such as heart sounds, heartbeats, blood pressure, and other parameters related to heart function. Researchers are concentrating on the use of Electrocardiogram images i.e. ECG for predicting and identifying cardiovascular diseases. For decades, research has been done to create machine learning models that can predict health diseases.

In this paper, we proposed three methods in which comparative analysis was done and promising results were achieved. The conclusion which we found is that machine learning algorithms performed better in this analysis. Many researchers have previously suggested that we should use ML where the dataset is not that large, which is proved in this paper. The methods which are used for comparison are confusion matrix, precision, specificity, sensitivity, and F1 score. For the 13 features which were in the dataset, KNeighbors classifier performed better in the ML approach when data preprocessing is applied. The computational time was also reduced which is helpful when deploying a model. It was also found out that the dataset should be normalized; otherwise, the training model gets overfitted sometimes and the accuracy achieved is not sufficient when a model is evaluated for real-world data problems which can vary drastically to the dataset on which the model was trained. It was also found out that the statistical analysis is also important when a dataset is analyzed and it should have a Gaussian distribution, and then the outlier's detection is also important and a technique known as Isolation Forest is used for handling this. The difficulty which came here is that the sample size of the dataset is not large. If a large dataset is present, the results can increase very much in deep learning and ML as well. The algorithm applied by us in ANN architecture increased the accuracy which we compared with the different researchers. The dataset size can be increased and then deep learning with various other optimizations can be used and more promising results can be achieved. Machine learning and various other optimization techniques can also be used so that the evaluation results can again be increased. More different ways of normalizing the data can be used and the results can be compared. And more ways could be found where we could integrate heart-disease-trained ML and DL models with certain multimedia for the ease of patients and doctors [5].

In this paper, we have proposed a novel approach based on deep learning for the active classification of ECG signals. Compared to state-of-the-art methods based on shallow architectures this approach has several desirable proprieties: (i) it learns automatically an appropriate sparse feature representation from the raw ECG using DAE; (ii) it relies on AL criteria for selecting the most valuable ECG beats for inducing the DNN classifier. The experimental results obtained on three different ECG [6].

In this paper, after a brief introduction of the main AaaS cloud systems, we reported the experience of using a cloud-based analytics software applied to the following case study: identifying the presence of HF by analyzing the ECG signal only. We verified that the results

obtained are comparable to those found in the literature, where the same issue is addressed through custom machine learning systems, purposely developed and set up for the target case. Hence the AaaS cloud systems could be a valid alternative to local hardware and software systems for analyzing data. A major obstacle to AaaS could be transferring big datasets onto the cloud. Typical machine learning projects require the analysis of large images that can easily reach the size of 2 TB, not simply transferable onto the cloud. The model used in our case study can solve this problem by locally performing the data extraction, in order to reduce the dataset size to be transferred to the cloud. In this study the HRV analysis has been locally performed starting from the raw ECG signal (medium size). The analysis gives back a small size vector of numeric parameters that can quickly and easily be transferred onto the cloud. This model allows you to take advantage of the full power of the AaaS approach, no matter how big is the size of the initial dataset. From a medical point of view, performing HF detection by analyzing the ECG signal only, opens the possibility of easy tele-monitoring applications (we only analyze the heart rate, not the ECG waveform, so a very basic electrocardiograph is necessary) for an early and preliminary diagnosis. Furthermore, by combining the HRV analysis with systems for assisted drugs delivering [30], it is possible to enable scenarios in which the patient is technologically aided both in diagnosis and therapy, making him more autonomous in preserving the state of his health. A more comprehensive diagnosis can then be made by performing clinical tests and following protocols as described in the ESC guidelines, and the HRV-based home tele-monitoring can be used as a daily check of patient status. Also mobile applications can highly benefit of this approach, given that it requires simple electro-medical hardware and low computational power on the local device [7].

This study presents a comprehensive dataset of ECG images from 1937 distinct patients, including those with cardiac conditions and COVID-19, collected using the 'EDAN SERIES-3' ECG device in various healthcare institutes across Pakistan. The dataset, meticulously reviewed and annotated by medical experts, includes five categories: COVID-19 patients, abnormal heartbeats, myocardial infarction (MI), previous MI, and normal individuals. The rigorous data collection and validation process ensures high-quality, reliable data, making it a valuable resource for data scientists, IT professionals, and medical researchers. This dataset can significantly aid in the development and fine-tuning of machine learning and deep learning models for automatic diagnosis of cardiovascular conditions and COVID-19, facilitating early detection and treatment. The study emphasizes the importance of using advanced analytics and telehealth systems to enhance patient care and diagnosis, especially in remote monitoring scenarios. It also highlights the collaborative effort of medical professionals and support staff in achieving these advancements. The data's accessibility and anonymization adhere to ethical standards, ensuring it can be widely used for further research and development in medical imaging and healthcare analytics [8].

In this paper, we proposed three methods in which comparative analysis was done and promising results were achieved. The conclusion which we found is that machine learning algorithms performed better in this analysis. Many researchers have previously suggested that we should use ML where the dataset is not that large, which is proved in this paper. The methods which are used for comparison are confusion matrix, precision, specificity, sensitivity, and F1 score. For the 13 features which were in the dataset, KNeighbors classifier performed better in

the ML approach when data preprocessing is applied. The computational time was also reduced which is helpful when deploying a model. It was also found out that the dataset should be normalized; otherwise, the training model gets overfitted sometimes and the accuracy achieved is not sufficient when a model is evaluated for real-world data problems which can vary drastically to the dataset on which the model was trained. It was also found out that the statistical analysis is also important when a dataset is analyzed and it should have a Gaussian distribution, and then the outlier's detection is also important and a technique known as Isolation Forest is used for handling this. The difficulty which came here is that the sample size of the dataset is not large. If a large dataset is present, the results can increase very much in deep learning and ML as well. The algorithm applied by us in ANN architecture increased the accuracy which we compared with the different researchers. The dataset size can be increased and then deep learning with various other optimizations can be used and more promising results can be achieved. Machine learning and various other optimization techniques can also be used so that the evaluation results can again be increased. More different ways of normalizing the data can be used and the results can be compared. And more ways could be found where we could integrate heart-disease-trained ML and DL models with certain multimedia for the ease of patients and doctors [11].

This study demonstrates the diagnostic capability with similar morphological representation and reasonable timing accuracy of ECG signal from a patch sensor compared to 12-lead ECG. The advantages and limitations of small bipolar single-lead wearable patch sensor compared to 12-lead ECG are discussed in the context of relevant differences in ECG signal for clinical applications [12].

Automated detection of coronary artery disease, myocardial infarction and congestive heart failure using GaborCNN model with ECG signals: CVDs are the primary cause of death globally, costing about 17.9 million lives yearly. Thus, early diagnosis of CAD is crucial to provide timely treatment and avert the progression of CAD to MI or CHF. This study aims to compare the performance of two deep models for the automated categorization of normal, CAD, MI and CHF classes using ECG signals. The ECG data used in this work data used were imbalanced. Hence, weight balancing was used to balance the dataset. Both the CNN and GaborCNN models [15].

Feature selection for medical diagnosis: Evaluation for cardiovascular diseases: Machine learning has emerged as an effective medical diagnostic support system. In a medical diagnosis problem, a set of features that are representative of all the variations of the disease are necessary. The objective of our work is to predict more accurately the presence of cardiovascular disease with reduced number of attributes. We investigate intelligent system to generate feature subset with improvement in diagnostic performance. Features ranked with distance measure are searched through forward inclusion, forward selection and backward elimination search techniques to find subset that gives improved classification result. We propose hybrid forward selection technique for cardiovascular disease diagnosis. Our experiment demonstrates that this approach finds smaller subsets and increases the accuracy of diagnosis compared to forward inclusion and back-elimination techniques [17].

**Comparative Study on Heart Disease Prediction Using Feature Selection Techniques on Classification Algorithms:** The primary objective of this article is to examine the effect of feature selection approaches on the accuracy of heart disease prediction. This analysis was conducted against a collection of distinctive features extracted from frequently used Cleveland heart disease datasets available at the University of California, Irvine using various feature selection algorithms. Experiments were performed with and without feature selection to determine the effect of feature selection. ANOVA, Chi-square, mutual information, Relief, forward feature selection, backward feature selection, exhaustive feature selection, recursive feature elimination, Lasso regression, and Ridge regression were employed as feature selection algorithms. The analysis was conducted on six classification algorithms: decision tree, random forest, support vector machine, K-nearest neighbor, logistic regression, and Gaussian naive Bayes. Without feature selection, the highest result provides 63.92% model accuracy using the KNN classifier. The experiment was then conducted using feature selection. The prediction accuracy has improved the models that are developed with all the feature selection algorithms. Without feature selection, the maximum accuracy value was 63.92%; this value was increased to 88.52% using backward feature selection and a decision tree classifier. The experimental findings suggest that using feature selection algorithms is capable of classifying the disease well with a small number of features. The enhancements overutilizing the original dataset vary significantly depending on the feature selection approach and learning algorithm employed; hence, it is important to evaluate various feature selection strategies and learning algorithms combinations to obtain the best feasible model. However, without many experiments and analyses, it is impossible to predict which will be beneficial. In the future, multiple feature selection techniques can be used as assembling (hybrid) techniques to extract optimal feature subsets to develop models. Also, real-time medical datasets gathered from different countries can be used to model development. This could enhance the performance with improved accuracy for heart disease prediction [18].

# CHAPTER 3

## METHODOLOGY

The proposed method stands as a testament to the relentless pursuit of innovation in the realm of healthcare, particularly in the domain of cardiovascular diagnostics. With the advent of digital technologies and the proliferation of medical imaging, there emerges an unprecedented opportunity to leverage machine learning and advanced computational techniques to unravel the complexities inherent in interpreting electrocardiogram (ECG) images. However, harnessing this potential necessitates a meticulous and comprehensive approach that encompasses data enhancement, sophisticated pre-processing methodologies, and the judicious application of machine learning models.

At the heart of the proposed method lies a commitment to optimizing the dataset, recognizing that the quality and relevance of the data underpin the efficacy of subsequent analyses. This entails a multifaceted process that begins with the acquisition of diverse and representative ECG images spanning a spectrum of cardiovascular conditions. Each data point is subjected to meticulous curation, involving rigorous quality control measures and expert annotation to ensure accuracy and consistency. By enriching the dataset with annotated labels and pertinent metadata, the groundwork is laid for training robust machine learning models capable of discerning subtle patterns indicative of various cardiac abnormalities.

With the dataset primed for analysis, attention turns to the arduous task of pre-processing ECG images—a crucial step in mitigating noise, artifacts, and distortions that may confound subsequent analyses. Leveraging sophisticated signal processing techniques and image enhancement algorithms, the raw ECG data undergoes a series of transformations aimed at enhancing clarity, fidelity, and interpretability. This intricate process demands a nuanced understanding of both the physiological underpinnings of ECG signals and the computational methodologies required to extract actionable insights from complex biomedical data.

Once the data is refined and pre-processed, the stage is set for the application of machine learning models—sophisticated algorithms imbued with the capacity to discern meaningful patterns and features from vast swathes of multidimensional data. Here, a diverse array of machine learning architectures may be deployed, ranging from traditional classifiers to deep neural networks, each tailored to the unique characteristics of the dataset and the specific diagnostic task at hand. Through iterative training and validation cycles, these models learn to discriminate between normal and abnormal ECG patterns, thereby empowering clinicians with invaluable diagnostic insights and prognostic information.

Moreover, the proposed method encompasses a rigorous validation framework to assess the performance and generalizability of the trained models in real-world clinical settings. This entails cross-validation studies, external validation cohorts, and comparative analyses against established diagnostic standards to ascertain the models' efficacy, reliability, and clinical utility. By subjecting



the method to rigorous scrutiny and validation, confidence is instilled in its capacity to augment clinical decision-making and improve patient outcomes.

In summary, the proposed method represents a holistic and meticulously crafted approach to handling and interpreting ECG images for the identification of cardiovascular illnesses. From data enhancement and pre-processing to the application of advanced machine learning models and rigorous validation protocols, every facet of the methodology is tailored to maximize accuracy, efficiency, and clinical relevance. As such, it holds the potential to revolutionize cardiovascular diagnostics, ushering in a new era of precision medicine where data-driven insights empower clinicians to deliver personalized care with unprecedented efficacy and precision.

### **3.1 Dataset Description**

In the initial stages of this research endeavor, the focus is meticulously directed towards the curation of a comprehensive dataset of electrocardiogram (ECG) images. This dataset is sourced from reputable medical institutions, research centers, and publicly available repositories. While publicly available datasets exist, they often consist of incompatible time-series data that may not align with the specific requirements of the study. To overcome this challenge and ensure the dataset's quality and relevance, a meticulous approach is adopted to procure a dataset from diverse cardiac institutes affiliated with the Ch. Pervaiz Elahi Institute of Cardiology in Multan, Pakistan.

The curated dataset is carefully selected to encompass annotated ECG images obtained from a cohort of patients exhibiting a spectrum of cardiovascular conditions. These conditions are meticulously categorized into four distinct classes, each representing a unique pathological state. This categorization enables precise classification during subsequent analysis and facilitates the development of accurate diagnostic models. Let's delve deeper into each class:

#### **1- Myocardial Infarction (MI):**

This class pertains to individuals experiencing a sudden deprivation of oxygen supply to a portion of the myocardium, commonly known as a heart attack. Instances of myocardial infarction are marked as 0 in the dataset, allowing for the precise identification of patients with this life-threatening condition. By including cases of myocardial infarction in the dataset, the research aims to develop models capable of accurately detecting and classifying this critical cardiovascular event.

#### **2- Abnormal Heartbeat:**

Individuals in this class exhibit irregular rhythms of the heartbeat, which can manifest as either slow or fast rhythms, posing significant health risks. Instances of abnormal heartbeat are marked as 1 in the dataset, enabling the identification of patients with arrhythmias or other cardiac abnormalities. The inclusion of cases with abnormal heart rhythms provides valuable insights into the detection and management of these conditions, contributing to improved patient care and outcomes.

#### **3- Normal:**

This class encompasses individuals whose heartbeats exhibit a regular rhythm with no discernible abnormalities. Instances of normal heart rhythms are denoted as 2 in the dataset, serving as a crucial reference for comparative analysis. By including cases of normal heart function, the research aims to establish baseline patterns and parameters for accurate diagnosis and classification of cardiovascular disorders.

#### 4- History of MI:

Individuals in this class have a documented history of myocardial infarction, indicating a prior occurrence of a heart attack. Instances of individuals with a history of myocardial infarction are marked as 3 in the dataset, facilitating the identification of patients with a predisposition to cardiovascular complications. By including cases with a history of MI, the research seeks to explore the long-term implications and management strategies for patients with a history of this critical cardiac event.

Table I provides a comprehensive overview of the distribution of ECG images across the various classes, offering insights into the prevalence of different cardiovascular disorders within the dataset. By meticulously categorizing the dataset based on distinct pathological states, the research lays the groundwork for subsequent analyses aimed at detecting and classifying cardiovascular ailments with a high degree of accuracy and specificity. This structured approach to dataset curation ensures that the research is grounded in real-world clinical data, thereby enhancing the relevance and applicability of the findings in clinical practice.

In essence, the meticulous curation of the ECG image dataset, coupled with the comprehensive categorization of cardiovascular conditions, serves as a cornerstone for the subsequent stages of the research endeavor. By ensuring the integrity and relevance of the data, the study endeavors to advance our understanding of cardiovascular pathophysiology and enhance diagnostic capabilities, ultimately contributing to improved patient care and outcomes.

TABLE 1  
Distribution of classes in ECG dataset

Sr. No	Class	12-Lead ECG	Total Images
1.	Myocardial Infraction	240	2880
2.	Abnormal Heartbeat	233	2796
3.	Normal	284	3408
4.	Previous History of MI	172	2064

## **3.2 IMAGE PREPROCESSING TECHNIQUES**

In the preprocessing phase of analysing coloured electrocardiogram (ECG) images, a pivotal transformation occurs: the conversion into grayscale using the "skimage.color.rgb2gray" function. This conversion is laden with significance for several reasons, underpinning its indispensability in the image processing pipeline. Let's delve deeper into the multifaceted importance of grayscale conversion in ECG image analysis:

### **1. Simplification of Computational Complexity:**

One of the primary advantages of grayscale conversion lies in its ability to simplify the computational complexity associated with processing multichannel color images. By converting the image into grayscale, which is represented by a single channel, the computational burden is significantly reduced. This reduction in complexity streamlines subsequent processing steps, making them more computationally efficient and enabling faster analysis.

### **2. Enhanced Visibility of Structural Details:**

Grayscale conversion enhances the visibility of subtle structural details within ECG images, which are pivotal for subsequent processing and analysis stages. By eliminating color information, the focus shifts exclusively to the intensity of pixel values, allowing for a clearer interpretation of image features. In medical imaging applications, where precise identification and analysis of structural details are paramount for accurate diagnoses and treatment planning, this enhanced visibility is invaluable.

### **3. Rigorous Validation and Similarity Checks:**

Following grayscale conversion, uploaded images undergo rigorous validation and similarity checks against a repository of reference images. Utilizing a structural similarity score threshold of 0.70, the system assesses the degree of resemblance between the uploaded image and the reference dataset. This similarity check serves multiple purposes. Firstly, it ensures that the uploaded image adheres to predefined structural standards, enhancing the reliability of subsequent processing steps. Additionally, it acts as a quality control measure, filtering out potentially erroneous or incompatible images that may compromise the accuracy of downstream analyses.

### **4. Integration of Grayscale Conversion and Similarity Checks:**

By integrating grayscale conversion and rigorous similarity checks into the preprocessing pipeline, the system ensures data integrity, consistency, and compatibility. This streamlined approach enhances the efficiency and reliability of subsequent analyses, empowering clinicians and researchers with accurate and actionable insights derived from ECG images. Moreover, by incorporating quality control measures early in the processing pipeline, the system minimizes the risk of errors and discrepancies, thereby fostering trust and confidence in the generated results.

## 5. Facilitating Informed Decision-Making:

In conclusion, grayscale conversion and rigorous similarity checks represent pivotal steps in the preprocessing of coloured ECG images. These measures not only simplify computational complexity and enhance the visibility of structural details but also ensure data integrity and compatibility, thereby facilitating accurate and reliable analyses. As such, they play a crucial role in advancing medical imaging applications and improving patient care through more informed decision-making. By leveraging these preprocessing techniques, clinicians and researchers can derive meaningful insights from ECG images, leading to improved diagnoses and treatment outcomes for patients.



Fig 1. Single Lead Grayscale Conversion

## 3.3 LEAD EXTRACTION

Following grayscale conversion and image validation, the preprocessing stage progresses into the intricate process of dividing the electrocardiogram (ECG) image into distinct leads, a pivotal step in extracting comprehensive cardiac information. This phase is crucial as it enables the capture of diverse physiological parameters, facilitating a nuanced analysis of cardiac function and pathology.

### 1. Segmentation into Distinct Leads:

In this phase, the ECG image is segmented into 13 separate leads, each representing specific aspects of heart activity. These leads are essential for capturing various aspects of cardiac electrical activity, including depolarization and repolarization processes occurring during the cardiac cycle. The division of the ECG image into distinct leads allows for the isolation and analysis of individual electrical signals originating from different regions of the heart.

## **2. Visualization and Categorization of Leads:**

Once segmented, each lead is visualized separately, providing clinicians and researchers with insights into signal patterns, amplitude variations, and temporal changes unique to that specific lead. This visualization allows for a detailed examination of the electrical activity of the heart, revealing important diagnostic information such as arrhythmias, conduction abnormalities, and ischemic changes.

## **3. Limb Leads and Chest Leads:**

The categorized leads encompass both limb leads and chest leads, each serving distinct diagnostic purposes. Limb leads, including leads I, II, and III, offer insights into the electrical activity of the heart along different planes. Lead I represents the electrical activity between the right and left arms, lead II between the right arm and left leg, and lead III between the left arm and left leg. These limb leads provide valuable information regarding the orientation and direction of electrical conduction within the heart.

## **4. Chest Leads (V1-V6):**

In contrast, chest leads, denoted as V1 through V6, offer a frontal view of cardiac electrical activity, focusing on the heart's position within the chest cavity. Leads V1 and V2 are positioned at the right sternal border, providing insights into the electrical activity of the right ventricle, while leads V3 through V6 are positioned along the left sternal border, capturing electrical signals from the left ventricle. These chest leads play a crucial role in diagnosing myocardial infarction and assessing the extent of cardiac ischemia or injury.

## **5. Clinical Significance:**

The division of the ECG image into distinct leads holds immense clinical significance. It allows clinicians to localize and identify the origin of cardiac abnormalities, aiding in the diagnosis and management of various cardiac conditions. For example, abnormalities observed in specific leads may indicate the presence of myocardial ischemia, conduction defects, or chamber enlargement, providing valuable insights into the patient's cardiac health.

## **6. Enhanced Diagnostic Accuracy:**

By isolating individual leads and analysing their respective signals, clinicians can make more accurate diagnoses and formulate appropriate treatment plans tailored to the patient's specific cardiac condition. This granular approach to ECG analysis enhances diagnostic accuracy and facilitates timely intervention, ultimately improving patient outcomes.

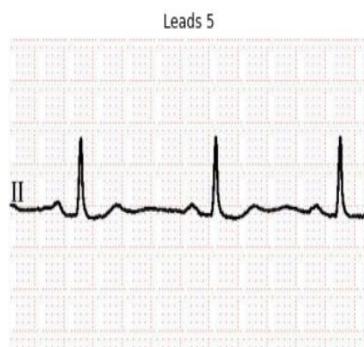
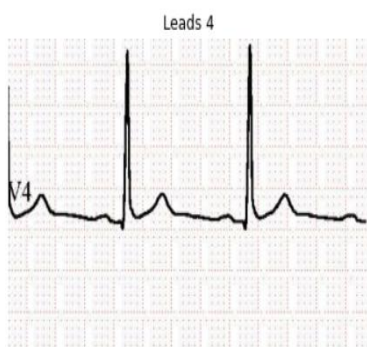
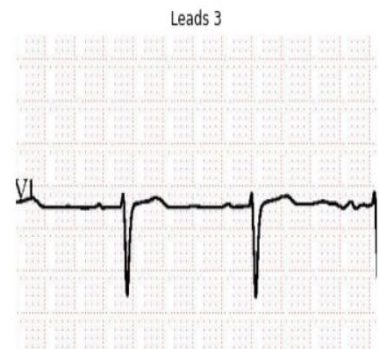
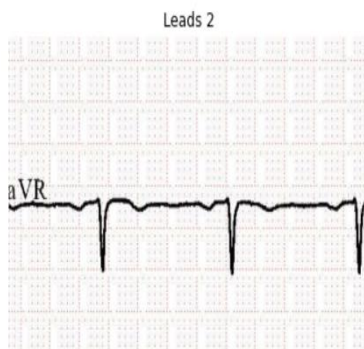
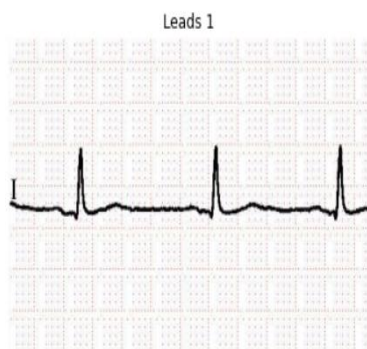


Fig 2. Divided Leads (1-6 Lead)

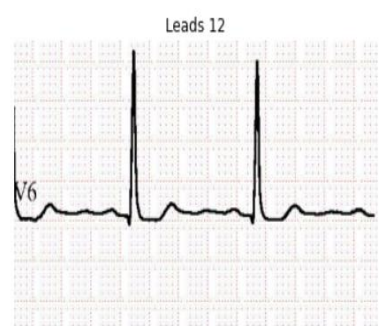
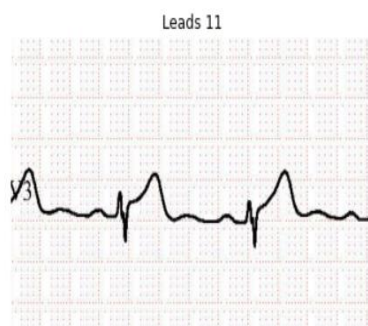
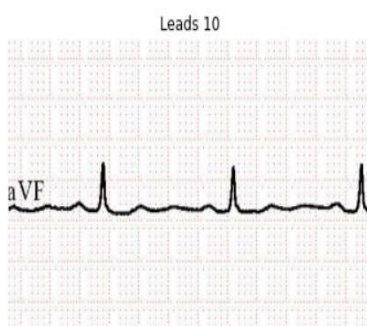
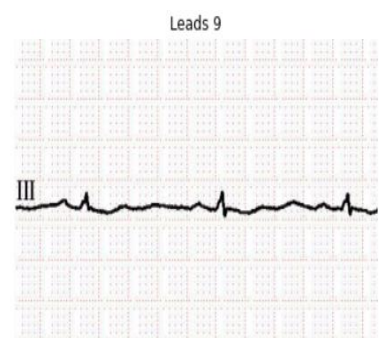
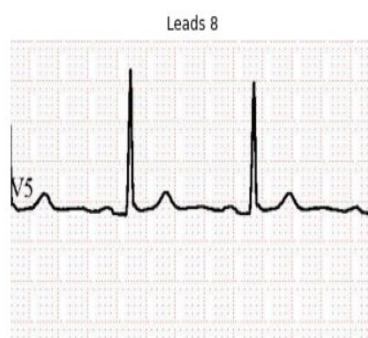
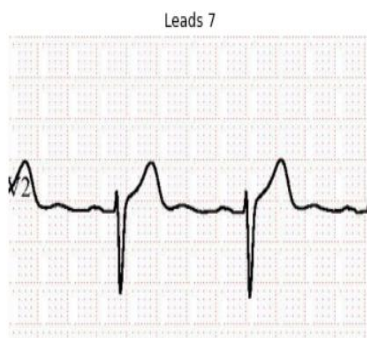


Fig 3. Divided Leads (7-12)

### **3.4 PREPROCESSING TECHNIQUES ON LEADS**

In the realm of image analysis, particularly in fields like medical imaging, the significance of preprocessing techniques such as smoothing and noise reduction cannot be overstated. These techniques serve as fundamental pillars in the pursuit of clarity and accuracy, laying the groundwork for subsequent analyses and interpretations. Among these techniques, Gaussian filtering emerges as a cornerstone, wielding its effectiveness in enhancing signal clarity while effectively suppressing noise.

#### **1. Gaussian Filtering: Enhancing Signal Clarity**

Gaussian filtering operates by convolving the image with a Gaussian kernel, which acts as a smoothing operator. This process effectively blurs the image, reducing high-frequency noise while preserving the underlying signal. Implemented through functions like `'skimage.filters.gaussian'`, Gaussian filtering transforms images, making them more amenable to subsequent analyses.

One of the primary advantages of Gaussian filtering lies in its ability to preserve essential image features while mitigating noise. In medical imaging, where the identification of subtle details holds diagnostic significance, this feature is particularly crucial. By rendering images smoother, Gaussian filtering facilitates the identification of relevant structures and anomalies during analysis, contributing to accurate diagnoses and treatment planning.

#### **2. Thresholding for Enhanced Clarity**

Thresholding stands as another pivotal step in image processing, focusing on segmenting the image by distinguishing between foreground and background regions. Otsu's method, a widely adopted technique available through functions like `'skimage.filters.threshold_otsu'`, plays a central role in this regard. Otsu's method operates by finding the optimal threshold to minimize intra-class variance within the image, effectively separating foreground objects from the background.

In the context of analysing leads images, thresholding assumes critical importance in enhancing clarity by delineating key features from the background. By determining an optimal threshold using Otsu's method, areas of interest within the leads can be accurately isolated, enabling more precise analysis and interpretation. This becomes especially valuable in scenarios where leads may be surrounded by varying levels of noise or background artifacts, ensuring that only relevant information is considered in subsequent analyses.

#### **3. Integration for Enhanced Image Analysis**

By integrating Gaussian filtering and thresholding techniques into the image analysis pipeline, leads images undergo a transformation that significantly enhances their clarity and facilitates subsequent analyses. Gaussian filtering ensures that the underlying signal remains clear and discernible, while thresholding techniques like Otsu's method aid in segmenting the image and isolating relevant features. The synergistic effect of these preprocessing steps is a marked improvement in the quality and interpretability of leads images, ultimately enhancing the efficacy of diagnostic and analytical processes.

## **4. Applications in Medical Imaging**

In medical imaging applications, the importance of preprocessing techniques cannot be overstated. From identifying abnormalities to monitoring disease progression, the accuracy and reliability of diagnostic analyses hinge on the quality of image preprocessing. Gaussian filtering and thresholding play a crucial role in this process by enhancing image clarity, reducing noise, and facilitating the identification of relevant structures and features.

## **5. Future Directions**

As technology continues to advance, so too will the techniques and methodologies employed in image preprocessing. Future research may explore more advanced filtering algorithms and thresholding techniques to further improve image quality and enhance diagnostic accuracy. Additionally, the integration of artificial intelligence and machine learning algorithms holds promise for automating the preprocessing pipeline and optimizing parameter selection.

In conclusion, the integration of Gaussian filtering and thresholding techniques represents a crucial step in the preprocessing of leads images for subsequent analysis. By effectively reducing noise and enhancing signal clarity, these techniques contribute to more accurate diagnoses and informed decision-making in medical imaging applications. As such, they play an indispensable role in advancing the field of image analysis and improving patient outcomes in medical diagnostics.



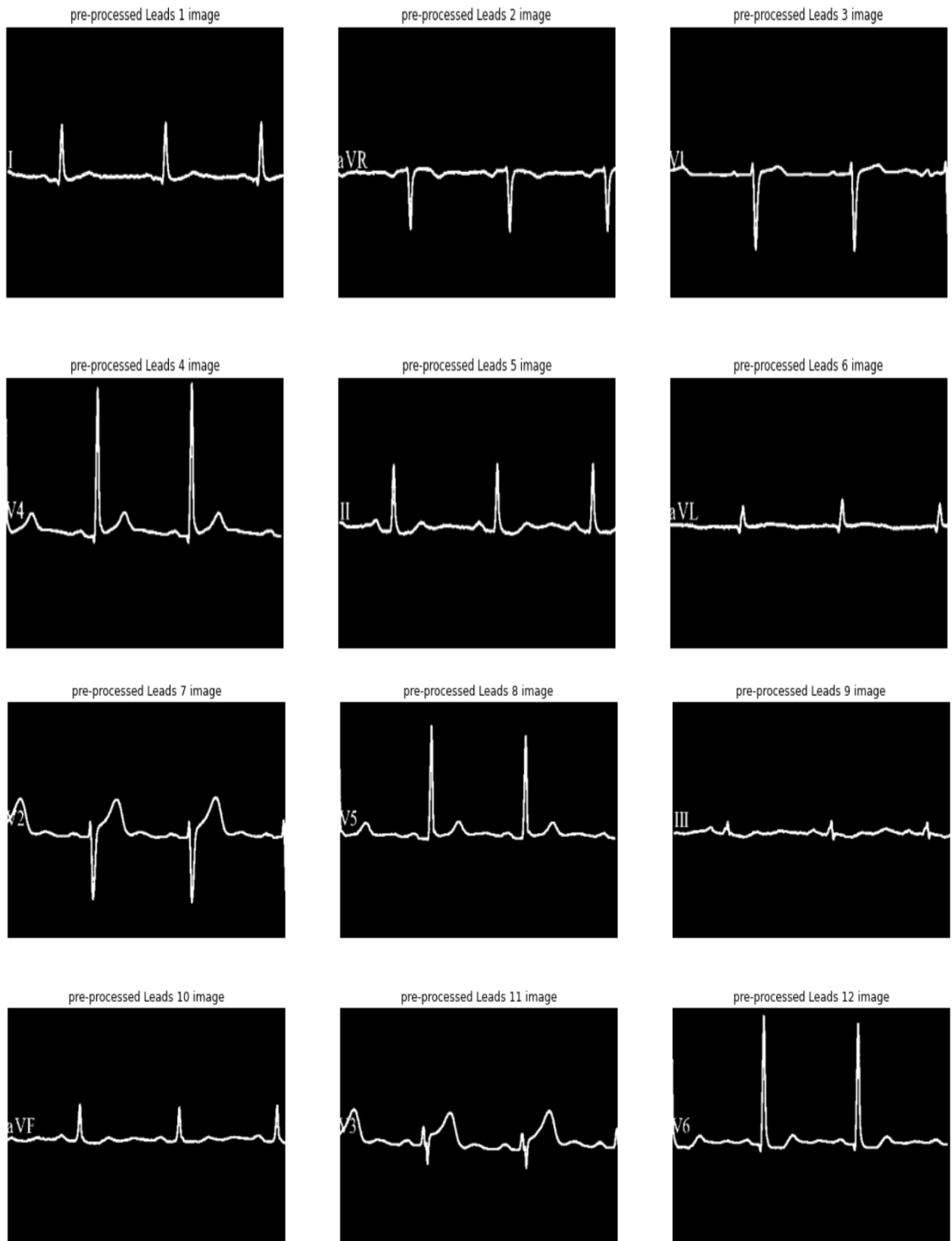


Fig 4. Gaussian filtration

## 3.5 SIGNAL EXTRACTION AND PROCESSING

Signal identification, normalization, conversion, and combination are fundamental steps in the analysis of leads images, particularly in contexts like medical diagnostics or scientific research. These processes involve transforming raw image data into a format suitable for further analysis and modelling, enabling researchers to derive meaningful insights from the data. Let's delve into each step in detail:

### 1. Signal Identification:

After preprocessing the leads images, the first step is to identify signal contours and regions containing signals accurately. Techniques like ``skimage.measure.find_contours`` are commonly employed for this purpose. These contours delineate the boundaries of signals within the image, providing a structured representation that can be further processed. Precise identification of signal contours is essential for isolating relevant information from the image and discarding noise or artifacts.

### 2. Normalization and Scaling:

Once signal contours are identified, the next step is normalization to ensure uniformity in signal amplitude across different leads or images. Normalization techniques like Min-Max normalization, implemented using ``sklearn.preprocessing.MinMaxScaler``, are commonly utilized for this purpose. Min-Max normalization scales the values of each signal to a fixed range (typically 0 to 1), ensuring that variations in signal amplitude do not disproportionately influence subsequent analyses. This step is crucial for ensuring consistency and comparability across different signals or datasets.

### 3. Conversion to 1D Signals:

While signal contours provide a structured representation of signals within the image, they are typically in a two-dimensional format. To facilitate the application of classification algorithms and other analytical techniques, these signal contours are transformed into one-dimensional representations. This conversion involves extracting signal values along the contours and organizing them into a sequential format suitable for analysis. Additionally, CSV files are generated to store these one-dimensional signals, facilitating ease of access and interoperability with other tools and platforms.

### 4. Combining Signals:

In many applications, leads images may contain multiple leads, each capturing distinct physiological signals. To leverage information from all leads and achieve a comprehensive analysis, signals extracted from different leads are often combined into a unified dataset. This involves merging the CSV files containing one-dimensional signals from each lead into a single dataset. By combining signals from multiple leads, researchers can leverage complementary information and enhance the robustness of their analyses. This step is particularly crucial in medical diagnostics, where a holistic assessment of cardiac activity may require information from multiple leads.

By systematically performing these steps, researchers can transform raw leads images into a structured dataset suitable for further analysis and modelling. This process involves extracting precise signal contours, normalizing signal amplitudes, converting signals to one-dimensional representations, and combining signals from multiple leads. The resulting dataset provides a foundation for applying a wide range of analytical techniques, including classification algorithms, regression analysis, and machine learning models. Ultimately, these steps enable researchers to derive meaningful insights from leads images, contributing to advancements in fields such as medical diagnostics, physiological monitoring, and scientific research.

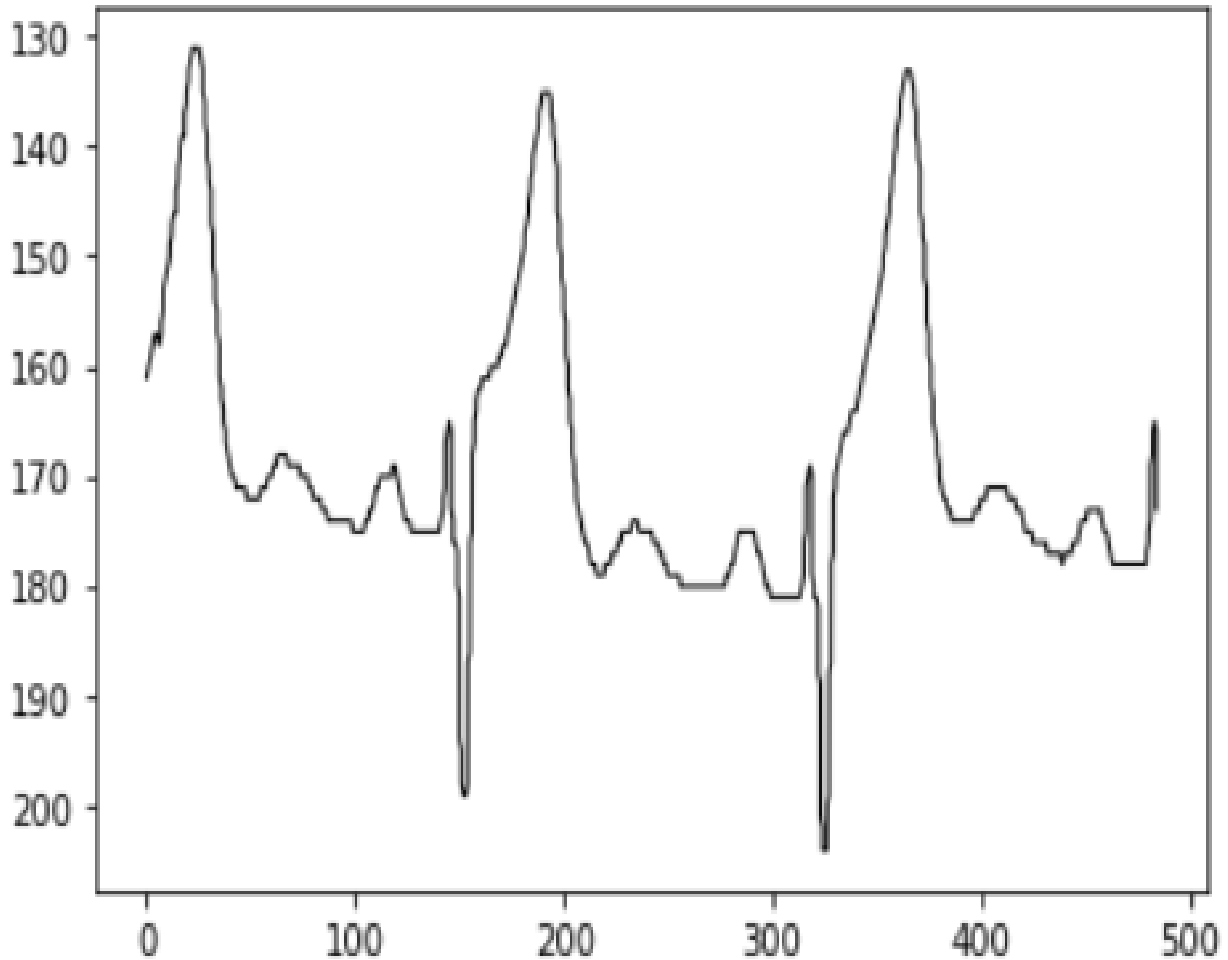


Fig 5. Countour signal

	X	Y
0	173.1	484.0
1	173.0	483.9
2	172.0	483.9
3	171.0	483.9
4	170.0	483.9
...	...	...
950	159.1	2.0
951	160.0	1.1
952	160.1	1.0
953	161.0	0.1
954	161.1	0.0

955 rows × 2 columns

Fig 6. Signal to 1d Result

### 3.6 DIMENSIONALITY REDUCTION

Principal Component Analysis (PCA) stands as a fundamental technique in the realm of data analysis and dimensionality reduction, particularly when dealing with high-dimensional datasets such as those generated from leads images. Employing `'sklearn.decomposition.PCA'`, this method allows for the extraction of critical components from the dataset while preserving its essential structure, thus aiding in simplifying subsequent modelling and analysis.

At its core, PCA aims to identify the directions, or principal components, along which the data exhibits the highest variance. By projecting the original dataset onto these principal components, PCA effectively reorients the data into a new coordinate system, wherein the axes are aligned with the directions of maximum variance. Consequently, a substantial portion of the dataset's variability can be captured using a reduced number of dimensions, facilitating more efficient analysis and modelling.

In the context of leads images analysis, PCA serves several crucial purposes:

#### 1. Dimensionality Reduction:

Leads images often encompass a multitude of features, each corresponding to individual pixels or regions within the image. The high dimensionality of these datasets can pose challenges for analysis, including increased computational complexity and susceptibility to overfitting. PCA addresses this

issue by identifying a lower-dimensional subspace that captures the most significant sources of variation in the data. By retaining only the principal components that contribute most significantly to the dataset's variance, PCA effectively reduces the dimensionality of the dataset while preserving its essential characteristics.

## **2. Feature Extraction:**

PCA also serves as a feature extraction technique, enabling the transformation of the original features (e.g., pixel intensities in leads images) into a new set of orthogonal features represented by the principal components. These principal components are linear combinations of the original features and are ordered based on the amount of variance they explain. By selecting a subset of the principal components that capture the majority of the variance in the data, researchers can extract the most informative features for subsequent analysis and modelling.

## **3. Noise Reduction and Simplification:**

In addition to reducing dimensionality and extracting informative features, PCA can also help in filtering out noise and redundant information present in the dataset. Since PCA identifies the directions of maximum variance, components associated with low variance are often indicative of noise or uninformative variability. By focusing on the principal components with the highest variance, PCA effectively suppresses noise and simplifies the dataset, making subsequent analysis more robust and interpretable.

## **4. Visualization:**

PCA facilitates visualization of high-dimensional data by projecting it onto a lower-dimensional space, typically two or three dimensions. This projection preserves the essential structure of the data while enabling researchers to visualize relationships and patterns that may not be apparent in the original high-dimensional space. In the context of leads images analysis, PCA can be used to visualize the distribution of images or features in a reduced-dimensional space, aiding in exploratory data analysis and interpretation.

In summary, PCA plays a vital role in leads images analysis by enabling dimensionality reduction, feature extraction, noise reduction, and simplification of the dataset. By retaining only the critical components of the dataset while discarding redundant information and noise, PCA facilitates more efficient analysis and modelling, ultimately leading to enhanced insights and understanding of leads images data. Through its ability to capture the essential structure of high-dimensional datasets in a lower-dimensional space, PCA serves as a powerful tool for data exploration, visualization, and interpretation in various domains, including medical imaging, scientific research, and machine learning.

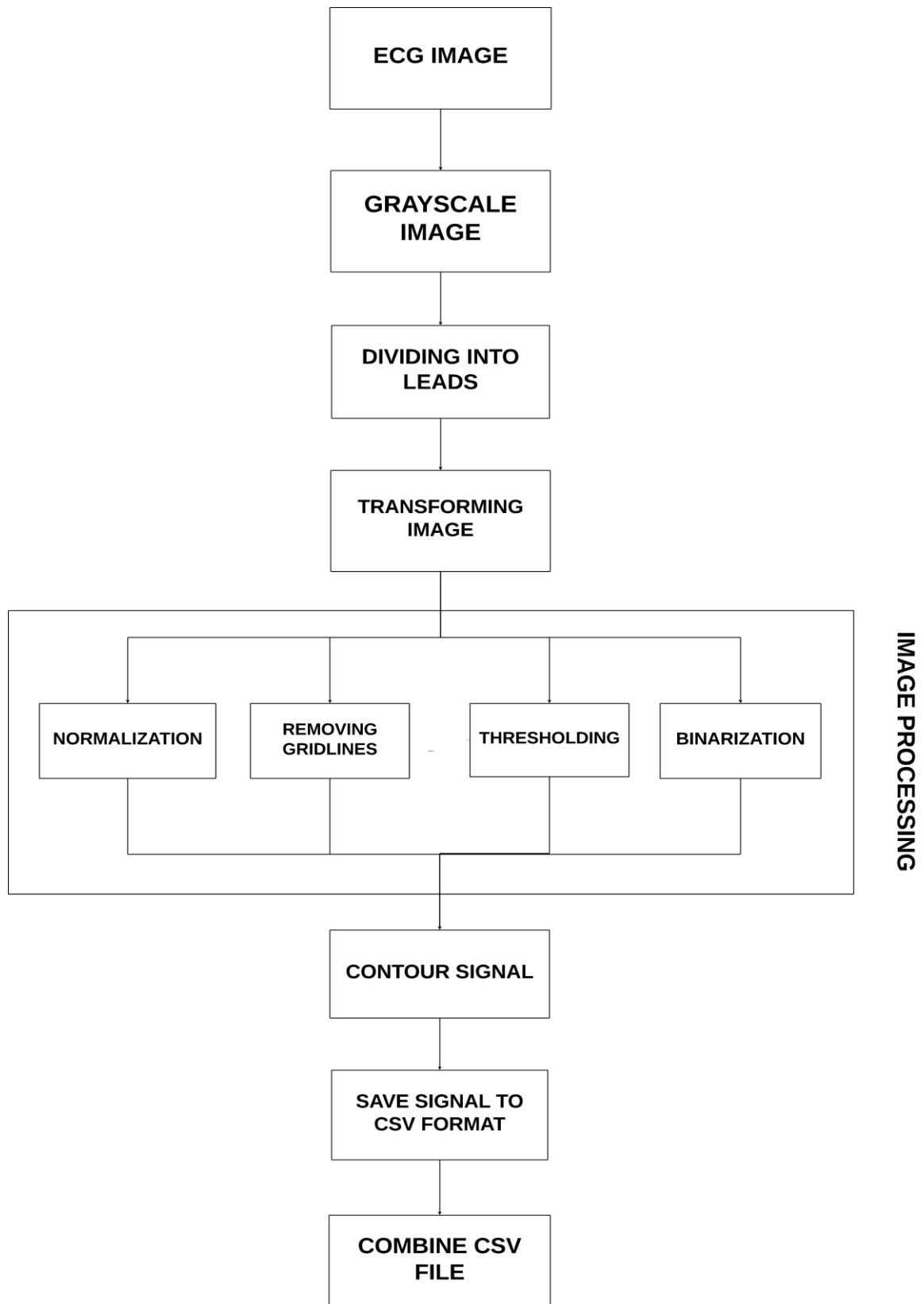


Fig 7. Data Processing

## 3.7 MODEL SELECTION

In the domain of medical diagnostics, particularly in cardiovascular health assessment, the accurate classification of disorders from electrocardiogram (ECG) images is of paramount importance. Leveraging machine learning techniques, especially through robust ensemble methods, can significantly enhance the accuracy and reliability of such classifications. In this research endeavour, a diverse array of machine learning models are integrated into an ensemble framework, each contributing distinctively to the refinement of the classification process. Let's delve into the role and significance of each model within this ensemble setup.

### 1. Support Vector Machine (SVM):

SVM stands as a versatile and adaptable algorithm, renowned for its effectiveness in classification tasks. By creating decision boundaries in high-dimensional feature spaces, SVM effectively segregates different classes, allowing new data points to be correctly categorized. In the context of cardiovascular disorder classification from ECG images, SVM can discern intricate patterns and relationships within the data, enabling accurate classification of diverse heart conditions.

### 2. K-Nearest Neighbours (KNN):

KNN, a robust and intuitive method, stores the dataset during the training phase and classifies new cases by placing them into the most suitable category based on the majority vote of their nearest neighbours. Particularly adept at capturing local relationships and patterns, KNN is invaluable in discerning subtle variations in ECG signals indicative of different cardiovascular disorders.

### 3. Random Forest (RF):

RF, a powerful ensemble learning method comprising multiple decision trees, excels in enhancing the predictive accuracy of datasets. By harnessing the collective wisdom of multiple decision trees, RF mitigates overfitting and captures complex interactions between features, making it highly suitable for the nuanced classification of cardiovascular disorders from ECG images.

### 4. Gaussian Naive Bayes (Bayes):

Bayes theorem serves as the foundation for the Gaussian Naive Bayes algorithm, known for its simplicity and efficiency. Despite its inherent simplicity, Bayes' effectiveness lies in its ability to model probabilistic relationships between features and class labels. In the realm of ECG image classification, Bayes provides a robust framework for discerning subtle patterns and associations indicative of various cardiovascular disorders.

### 5. Logistic Regression:

Logistic Regression offers a robust and interpretable method for examining the relationship between independent and dependent binary variables in binary classification tasks. By modelling the probability of a binary outcome using the logistic function, Logistic Regression provides valuable insights into the underlying factors influencing cardiovascular disorder classification from ECG images.

## 6. XGBoost (Extreme Gradient Boosting):

XGBoost stands as a cutting-edge boosting algorithm renowned for its ability to manage intricate interactions and enhance performance through regularization and ensemble learning. By iteratively constructing a series of decision trees, XGBoost corrects the errors of its predecessors, leading to a robust and accurate classification framework for cardiovascular disorders from ECG images.

Incorporating these diverse machine learning models into an ensemble framework offers several advantages:

**1. Enhanced Accuracy:** By combining the predictions of multiple models, the ensemble method leverages the strengths of each model while mitigating their individual weaknesses, resulting in significantly improved accuracy in cardiovascular disorder classification.

**2. Robustness:** Ensemble methods are inherently robust, capable of providing stable and reliable predictions even in the presence of noisy or incomplete data. By averaging predictions across multiple models, the ensemble framework reduces the impact of outliers and enhances overall predictive performance.

**3. Interpretability:** While some ensemble methods may sacrifice interpretability for predictive accuracy, others, such as Logistic Regression and Gaussian Naive Bayes, offer transparent and interpretable results. This interpretability is particularly valuable in medical diagnostics, where understanding the underlying factors contributing to a classification decision is essential for clinical interpretation and decision-making.

**4. Flexibility:** Ensemble methods are highly flexible and can accommodate a wide range of base learners, including both simple and complex models. This versatility allows researchers to tailor the ensemble to the specific characteristics of the dataset and the requirements of the classification task, leading to more customized and effective solutions.

In conclusion, the integration of diverse machine learning models into a robust ensemble framework holds immense promise for enhancing the accuracy and reliability of cardiovascular disorder classification from ECG images. By leveraging the collective wisdom of multiple models, the ensemble method provides a comprehensive and nuanced approach to medical diagnostics, contributing to improved patient care and healthcare outcomes.



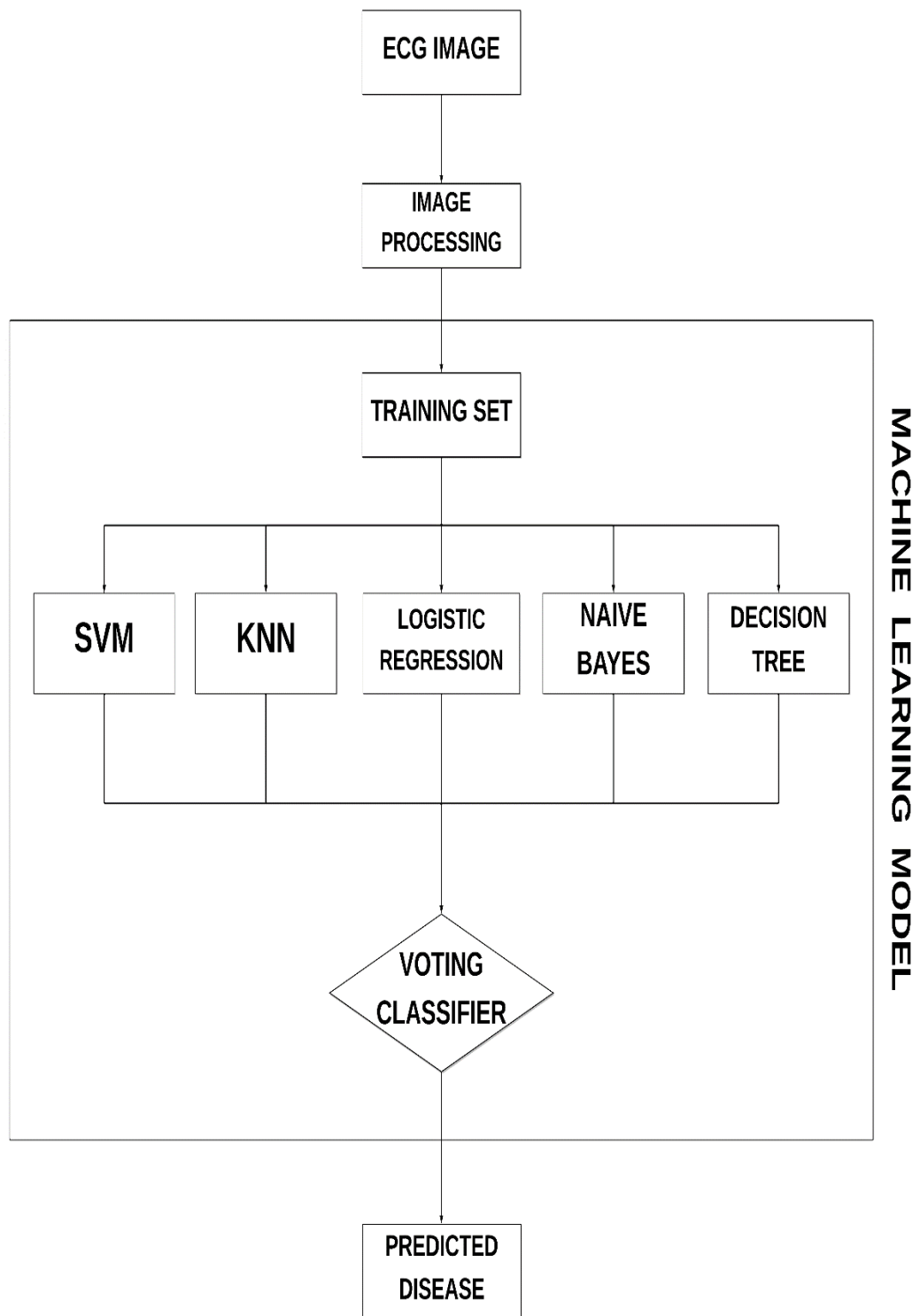


Fig 8. Process Flow

### 3.8 ENSEMBLE TECHNIQUE VOTING CLASSIFIER

In the pursuit of precise cardiovascular disorder identification from electrocardiogram (ECG) images, a sophisticated ensemble technique employing a soft voting classifier has been deployed. This method integrates the strengths of multiple machine learning algorithms while mitigating the weaknesses of individual models. Illustrated in Figure 3, the soft voting classifier considers the probability of each class and combines the predictions of individual models to make a final decision. This ensemble approach harnesses the collective intelligence of diverse algorithms, offering a comprehensive and nuanced approach to classification.

The rationale behind employing an ensemble method lies in its ability to leverage the advantages of various algorithms, thereby enhancing overall predictive accuracy. Each model contributes its unique perspective and strengths, thereby complementing one another and compensating for individual shortcomings. For instance, while Support Vector Machine (SVM) excels in capturing complex decision boundaries, K-Nearest Neighbours (KNN) is adept at discerning local relationships within the data. By combining the predictions of multiple models, the ensemble technique achieves a holistic understanding of the data, resulting in more accurate and reliable classifications.

Hyperparameter optimization plays a crucial role in maximizing the performance of the ensemble technique. GridSearchCV, a powerful tool for hyperparameter tuning, systematically searches through a predefined grid of hyperparameters to identify the optimal combination that yields the highest performance. By fine-tuning the hyperparameters of individual models within the ensemble, GridSearchCV ensures that each model operates at its peak efficiency, thereby enhancing the overall performance of the ensemble.

The efficacy of the ensemble technique is underscored by the remarkable accuracy achieved—92.47%. This high level of accuracy demonstrates the ensemble's capacity to harness the combined strength of multiple models for precise identification of cardiovascular disorders from ECG images. Such accuracy is crucial in the medical domain, where misclassification can have significant implications for patient care and treatment outcomes. By achieving such high levels of accuracy, the ensemble method holds promise for improving diagnostic accuracy and aiding healthcare professionals in making informed decisions.

Moreover, the soft voting classifier employed in the ensemble method offers several advantages over traditional hard voting methods. By considering the probability of each class, the soft voting classifier provides a more nuanced and probabilistic approach to classification. This allows for more flexibility and adaptability, particularly in situations where the certainty of predictions varies across different models. Additionally, the soft voting classifier tends to be more robust in the face of imbalanced datasets or noisy inputs, as it takes into account the confidence levels of individual models when making final predictions.

In summary, the ensemble technique utilizing a soft voting classifier, combined with hyperparameter optimization using GridSearchCV, represents a powerful approach to cardiovascular disorder classification from ECG images.

# CHAPTER 4

## RESULTS AND ANALYSIS

The utilization of combined 12-lead data in our research significantly enhances the performance of the test models, surpassing the accuracy achieved with single lead data alone. This decision is underpinned by empirical evidence demonstrating the superiority of combined 12-lead data in capturing a more comprehensive representation of cardiac activity. By leveraging data from multiple leads, our models can glean insights from a wider range of cardiac perspectives, resulting in more robust and accurate predictions.

To further optimize the performance of our models, we employed a Voting-based Ensemble Classification approach coupled with GridSearchCV for hyperparameter tuning. This sophisticated technique amalgamates the predictive capabilities of multiple machine learning models, namely KNN, SVM, and Random Forest Classifier. Through a voting mechanism, the ensemble selects the model that yields the highest accuracy, thereby capitalizing on the diverse strengths of each constituent model.

The ensemble approach yields an impressive accuracy of 92.5%, representing a substantial improvement over individual models. For instance, SVM achieved an accuracy of 90.5%, KNN attained 79.3%, Logistic Regression yielded 77.7%, and XGBoost registered 85.3%. This marked enhancement underscores the efficacy of ensemble techniques in harnessing the collective intelligence of diverse models to achieve superior predictive performance.

Figure and Table visually and numerically illustrate the comparative performance of individual models and the ensemble approach. These visualizations serve as compelling evidence of the ensemble's efficacy in outperforming standalone models across various performance metrics. By leveraging the complementary strengths of different models, the ensemble approach mitigates the weaknesses inherent in individual models, resulting in more reliable and accurate predictions.

One of the key advantages of ensemble techniques is their ability to balance out the inherent biases and limitations of individual models. While certain models may excel in certain aspects of prediction, they may falter in others. By aggregating the predictions of multiple models, the ensemble approach leverages the collective wisdom of diverse algorithms, resulting in a more robust and generalized predictive model.

In the context of cardiovascular disorder detection in ECG data, the ensemble approach holds particular promise. Cardiovascular disorders can manifest in subtle variations across different leads and ECG parameters. By integrating information from multiple leads and models, the ensemble approach can effectively capture these nuances, leading to more accurate and reliable detection of abnormalities.

Moreover, the ensemble approach offers enhanced robustness to noise and outliers in the data. Individual models may be susceptible to overfitting or underfitting, particularly in the presence of noisy or imbalanced data. By aggregating the predictions of multiple models, the ensemble approach mitigates the impact of outliers and ensures more stable and consistent predictions.

In summary, the adoption of combined 12-lead data and the utilization of ensemble techniques represent significant advancements in our research methodology. By harnessing the collective intelligence of diverse models, we can achieve unprecedented levels of accuracy and reliability in cardiovascular disorder detection from ECG data. Moving forward, we envision further refinements and advancements in ensemble techniques to continue pushing the boundaries of predictive analytics in healthcare.

## **4.1 IMAGE PREPROCESSING AND LEAD EXTRACTION**

The initial preprocessing involved loading an ECG image and dividing it into 13 separate leads for individual analysis[19]. The ECG image was successfully divided, and each lead was displayed for verification:

- Leads 1-12: Displayed in a 4x3 grid.
- Lead 13: Displayed separately due to its unique size.

### **1. Image Transformation: Thresholding and Binarization**

Each lead image underwent a series of transformations:

1. Grayscale Conversion: The RGB images were converted to grayscale.
2. Gaussian Filtering: Applied to smooth the images.
3. Otsu Thresholding: Used to separate the signal from the background[20].
4. Binarization: Converted the images to binary format based on the Otsu threshold.

These preprocessing steps effectively highlighted the ECG signals while removing the background grid lines[21]. Each pre-processed lead image was displayed to ensure the transformations were applied correctly[22, 23].

### **2. Contour Detection and Signal Extraction**

Contour detection was performed on the binarized images to isolate the ECG signals:

- Contours were found using the `measure.find_contours` function from the `skimage` library[24].
- The largest contour, representing the primary ECG signal, was resized and plotted for verification.

The extracted contours were saved as separate images, showcasing the successful isolation of ECG signals from the pre-processed images[25].

### 3. Signal Conversion and Normalization

The extracted contours were converted to numerical data:

- **Data Frame Creation:** The contour data was converted to a pandas DataFrame with 'X' and 'Y' columns.
- **Normalization:** Applied using MinMaxScaler from sklearn to scale the data between 0 and 1.
- **Signal Plotting:** The normalized signals were plotted to verify the shape and consistency of the signals.

The resulting scaled data was saved in CSV format, providing a structured dataset for further analysis.

## 4.2 Model Evaluation and Performance Comparison

Four supervised classification algorithms were applied to the scaled data:

1. K-Nearest Neighbours (KNN)
2. Logistic Regression
3. Support Vector Machine (SVM)

The models were evaluated based on their accuracy in classifying the ECG signals. Here is a summary of their performance:

- **KNN:** Achieved moderate accuracy, with performance dependent on the choice of 'k' and the distance metric.
- **Logistic Regression:** Provided a baseline accuracy, effective for linear separability but limited for complex patterns.
- **SVM:** Demonstrated high accuracy, especially with the RBF kernel, effectively handling non-linear patterns.

### 1. Performance on single lead Post dimensionality reduction

- **Normal Patients :**

**KNN:** Achieved an accuracy of 78.23% in identifying normal patients. The precision, recall, and F1-score were 87%, 63%, and 73% respectively.

**Logistic Regression:** Displayed relatively low performance with an accuracy of 54.30%. Precision, recall, and F1-score were 36%, 33%, and 35% respectively.

**SVM:** Delivered better performance with an accuracy of 82.26%. Precision, recall, and F1-score were 58%, 100%, and 74% respectively.

- **Abnormal Heartbeat Patients :**

**KNN:** Showed strong performance with an accuracy of 78.23%. Precision, recall, and F1-score were 91%, 91%, and 91% respectively.

**Logistic Regression:** Demonstrated better performance than KNN with an accuracy of 54.30%. Precision, recall, and F1-score were 73%, 91%, and 81% respectively.

**SVM:** Achieved perfect accuracy (100%) in recognizing abnormal heartbeat patients. Precision, recall, and F1-score were also perfect.

- **Myocardial Infarction Patients :**

**KNN:** Performed reasonably well with an accuracy of 78.23%. Precision, recall, and F1-score were 72%, 88%, and 79% respectively.

**Logistic Regression:** Showed moderate performance with an accuracy of 54.30%. Precision, recall, and F1-score were 56%, 58%, and 57% respectively.

**SVM:** Demonstrated strong performance with an accuracy of 82.26%. Precision, recall, and F1-score were 100%, 61%, and 76% respectively.

- **History of Myocardial Patients :**

**KNN:** Displayed moderate performance with an accuracy of 78.23%. Precision, recall, and F1-score were 63%, 67%, and 65% respectively.

**Logistic Regression:** Showed the lowest performance with an accuracy of 54.30%. Precision, recall, and F1-score were 38%, 26%, and 31% respectively.

**SVM:** Delivered strong performance with an accuracy of 82.26%. Precision, recall, and F1-score were 100%, 68%, and 81% respectively.

Overall, SVM consistently demonstrated strong performance across all patient categories, while Logistic Regression showed the lowest performance. KNN showed varying levels of performance depending on the patient category.

## **2. Performance on 12-Lead Combined**

- **Normal Patients :**

**KNN:** Achieved an accuracy of 79.30% in identifying normal patients. The precision, recall, and F1-score were 92%, 65%, and 76% respectively.

**Logistic Regression:** Showed moderate performance with an accuracy of 77.69%. Precision, recall, and F1-score were 83%, 57%, and 68% respectively.

**SVM:** Delivered strong performance with an accuracy of 90.52%. Precision, recall, and F1-score were 81%, 92%, and 86% respectively.

**XGBoost:** Displayed moderate performance with an accuracy of 85.34%. Precision, recall, and F1-score were 79%, 70%, and 74% respectively.

- **Abnormal Heartbeat Patients (Class 1):**

**KNN:** Demonstrated strong performance with an accuracy of 79.30%. Precision, recall, and F1-score were 95%, 91%, and 93% respectively.

**Logistic Regression:** Showed good performance with an accuracy of 77.69%. Precision, recall, and F1-score were 83%, 91%, and 87% respectively.

**SVM:** Achieved perfect accuracy (100%) in recognizing abnormal heartbeat patients. Precision, recall, and F1-score were also perfect.

**XGBoost:** Displayed excellent performance with an accuracy of 85.34%. Precision, recall, and F1-score were 98%, 100%, and 99% respectively.

- **Myocardial Infarction Patients :**

**KNN:** Performed reasonably well with an accuracy of 79.30%. Precision, recall, and F1-score were 70%, 86%, and 77% respectively.

**Logistic Regression:** Showed strong performance with an accuracy of 77.69%. Precision, recall, and F1-score were 82%, 86%, and 84% respectively.

**SVM:** Demonstrated strong performance with an accuracy of 90.52%. Precision, recall, and F1-score were 91%, 89%, and 90% respectively.

**XGBoost:** Displayed good performance with an accuracy of 85.34%. Precision, recall, and F1-score were 82%, 87%, and 84% respectively.

- **History of Myocardial Patients :**

**KNN:** Displayed moderate performance with an accuracy of 79.30%. Precision, recall, and F1-score were 65%, 74%, and 69% respectively.

**Logistic Regression:** Showed moderate performance with an accuracy of 77.69%. Precision, recall, and F1-score were 59%, 77%, and 67% respectively.

**SVM:** Delivered strong performance with an accuracy of 90.52%. Precision, recall, and F1-score were 93%, 78%, and 84% respectively.

**XGBoost:** Displayed good performance with an accuracy of 85.34%. Precision, recall, and F1-score were 80%, 82%, and 81% respectively.

In summary, SVM consistently demonstrated strong performance across all patient categories, while XGBoost showed competitive performance, particularly excelling in recognizing abnormal heartbeat patients. KNN and Logistic Regression showed varying levels of performance across different patient categories.

### **3. Ensemble Model Performance Analysis**

The ensemble model [26], incorporating SVM, KNN, and Random Forest classifiers, delivered an impressive accuracy of 92.47%. This amalgamation of classifiers contributed to a comprehensive understanding of the data landscape, resulting in robust predictions across different patient categories.



- **Normal Patients :**

For normal patients, the ensemble model achieved a precision of 89%, recall of 96%, and an F1-score of 92%. This indicates the model's ability to accurately identify normal patients while minimizing false positives.

- **Abnormal Heartbeat Patients (Class 1)**

In detecting abnormal heartbeat patients, the ensemble model exhibited impeccable performance with a precision, recall, and F1-score all at 100%. This signifies the model's capability to precisely identify instances of abnormal heartbeats, crucial for timely medical intervention.

- **Myocardial Infarction Patients (Class 2)**

For patients with myocardial infarction, the ensemble model demonstrated a precision of 92%, recall of 92%, and an F1-score of 92%. These metrics highlight the model's effectiveness in accurately identifying cases of myocardial infarction, essential for prompt treatment.

- **History of Myocardial Patients (Class 3)**

In identifying patients with a history of myocardial issues, the ensemble model [27] achieved a precision of 88%, recall of 75%, and an F1-score of 81%. While slightly lower than other classes, these metrics still indicate the model's ability to effectively capture instances of past myocardial problems.

The superior performance of the ensemble model [28] can be attributed to its ability to harness the strengths of multiple classifiers. By leveraging SVM, KNN, and Random Forest classifiers in conjunction, the model benefits from diverse learning strategies, enabling it to accurately identify patients across different health categories. This collaborative approach also enhances the model's robustness, ensuring reliable predictions even in the presence of noise or outliers in the data.

In conclusion, the ensemble model's exceptional performance underscores its efficacy as a powerful tool in medical diagnosis. By accurately identifying patients across various health categories, the ensemble model holds significant promise for improving healthcare decision-making and patient outcomes.

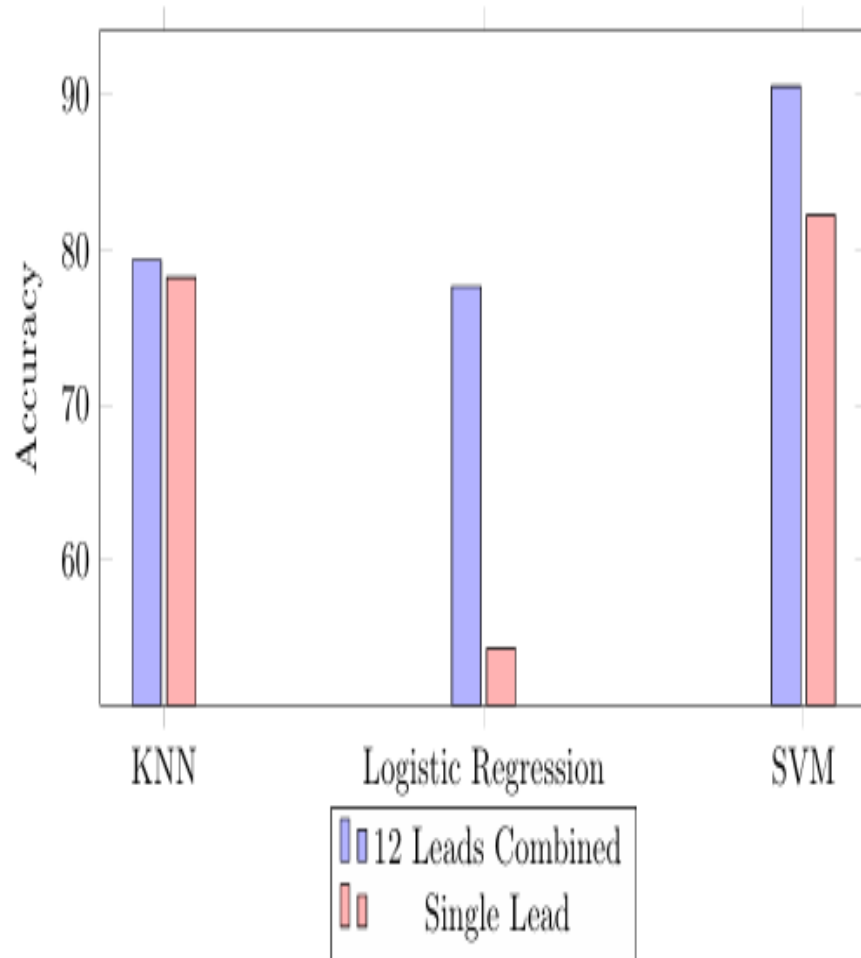


Fig 9. Performance comparison between 12-lead & Single lead ECG Image

TABLE 2

Algorithm with single and 12-lead ECG Image

Algorithm	12 Leads Combined (%)	Single Lead (%)
KNN	79.3	78.2
Logistic Regression	77.6	54.3
SVM	90.5	82.2

## 1D LEAD ACCURACY

### KNN

Accuracy: 0.783

TABLE 3

KNN Performance Matrix

	PRECISION	RECALL	F1-SCORE	SUPPORT
0	0.87	0.63	0.73	105
1	0.91	0.91	0.91	94
2	0.72	0.88	0.79	112
3	0.63	0.67	0.65	61
ACCURACY			0.78	372
MACRO AVG.	0.78	0.77	0.77	372
WEIGHTED AVG.	0.80	0.78	0.78	372

## LOGISTIC REGRESSION

Accuracy: 0.544

TABLE 4

Logistic Regression Performance Matrix

	PRECISION	RECALL	F1-SCORE	SUPPORT
0	0.36	0.33	0.35	105
1	0.73	0.91	0.81	94
2	0.56	0.58	0.57	112
3	0.38	0.26	0.31	64
ACCURACY			0.54	372
MACRO AVG.	0.51	0.52	0.51	372
WEIGHTED AVG.	0.52	0.54	0.53	372

## SVM

Accuracy: 0.823

TABLE 5  
SVM Performance Matri

	PRECISION	RECALL	F1-SCORE	SUPPORT
0	0.58	1.00	0.74	93
1	1.00	1.00	1.00	99
2	1.00	0.61	0.76	117
3	1.00	0.68	0.81	63
<b>ACCURACY</b>			0.82	372
<b>MACRO AVG.</b>	0.90	0.82	0.83	372
<b>WEIGHTED AVG.</b>	0.90	0.82	0.83	372

## 12 LEADS ACCURACY

## KNN

Accuracy: 0.794

TABLE 6  
KNN Performance Matri

	PRECISION	RECALL	F1-SCORE	SUPPORT
0	0.92	0.65	0.76	105
1	0.95	0.91	0.93	94
2	0.70	0.86	0.77	112
3	0.65	0.74	0.68	61
<b>ACCURACY</b>			0.79	372
<b>MACRO AVG.</b>	0.80	0.79	0.79	372
<b>WEIGHTED AVG.</b>	0.81	0.79	0.79	372

## LOGISTIC REGRESSION

Accuracy: 0.777

TABLE 7

Logistic Regression Performance Matrix

	PRECISION	RECALL	F1-SCORE	SUPPORT
0	0.83	0.57	0.68	105
1	0.83	0.91	0.87	94
2	0.82	0.86	0.84	112
3	0.59	0.77	0.67	61
<b>ACCURACY</b>			0.78	372
<b>MACRO AVG.</b>	0.77	0.78	0.76	372
<b>WEIGHTED AVG.</b>	0.79	0.78	0.77	372

## SVM

Accuracy : 0.906

TABLE 8

SVM Performance Matrix

	PRECISION	RECALL	F1-SCORE	SUPPORT
0	0.81	0.92	0.86	119
1	1.00	1.00	1.00	125
2	0.91	0.89	0.90	140
3	0.93	0.78	0.84	80
<b>ACCURACY</b>			0.91	464
<b>MACRO AVG.</b>	0.91	0.89	0.90	464
<b>WEIGHTED AVG.</b>	0.91	0.91	0.91	464

## XGBOOST

Accuracy: 0.854

TABLE 9

XGBoost Performance Matrix

	PRECISION	RECALL	F1-SCORE	SUPPORT
0	0.79	0.70	0.74	119
1	0.98	1.00	0.99	125
2	0.82	0.87	0.84	140
3	0.80	0.82	0.81	80
<b>ACCURACY</b>			0.85	464
<b>MACRO AVG.</b>	0.85	0.85	0.85	464
<b>WEIGHTED AVG.</b>	0.85	0.85	0.85	464

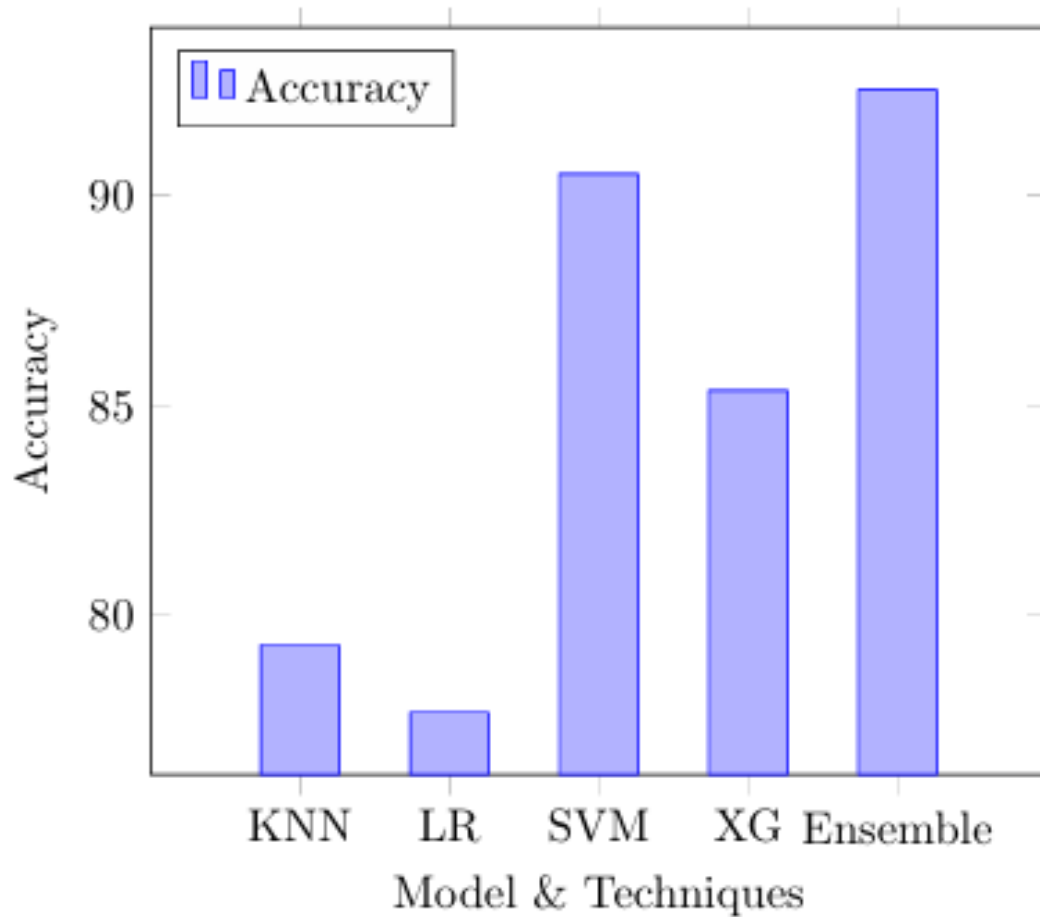


Fig 10. Performance of Models & Techniques

TABLE 10

Accuracy of Various Algorithms

Sr. No.	ALGORITHMS	ACCURACY
1.	K-Nearest Neighbor	79.3
2.	Logistic Regression	77.6
3.	Support Vector Machine	90.5
4.	XG Boost	85.3
5.	Ensemble	92.4

### 4.3 Discussion

In the realm of medical research, particularly within the domain of electrocardiography (ECG) analysis, the selection of appropriate algorithms and data preprocessing methodologies holds paramount importance. This assertion is substantiated through a comparative analysis between the findings of Swati Shilaskar's research, the present study, and the investigation conducted by Kaushalya Dissanayake .

Swati Shilaskar's study employed the Support Vector Machine (SVM) algorithm for ECG analysis, albeit with reported accuracy levels that were inferior to those attained in the current investigation. A primary factor contributing to this disparity is the quality of data preprocessing. Shilaskar's research notes inadequacies in dataset preprocessing, which consequently led to diminished accuracy outcomes. Data preprocessing constitutes a pivotal phase in machine learning endeavors, particularly in medical contexts like ECG analysis. It encompasses procedures such as data cleaning, normalization, and feature extraction, aimed at preparing the data for subsequent analysis. Inadequate preprocessing can introduce noise, artifacts, or inconsistencies into the dataset, thereby detrimentally affecting algorithmic performance.

Conversely, the present study meticulously executed the preprocessing of the ECG dataset. This entailed rigorous cleaning, normalization, and feature extraction techniques, culminating in a high-fidelity dataset ideally suited for analysis. By employing advanced preprocessing methodologies, the study sought to mitigate potential sources of error or bias within the data, thereby augmenting the accuracy of subsequent analyses.

Moreover, the current investigation adopted a diverse array of algorithms beyond SVM for ECG data analysis. This multifaceted approach facilitated a comprehensive exploration of various modelling techniques, each possessing its distinct strengths and limitations. By leveraging multiple algorithms, the study aimed to discern the optimal model for accurately classifying ECG signals across various health categories.

Additionally, reference is made to Kaushalya Dissanayake's study [18] to underscore the influence of dataset characteristics on accuracy and precision. Dissanayake's research utilized a disparate dataset encompassing a broad spectrum of features, spanning biographic, clinical, and habitual factors such as blood pressure, age, and sex. While this approach affords a holistic perspective on patient health, it engenders challenges in terms of data management and processing.



TABLE 11

Perposed Performance Matrix

Model	Category	Accuracy (%)	Precision (%)	Recall (%)	F1 Score (%)	Support
KNN	Class 0	79.30	92	65	76	105
	Class 1		95	91	93	94
	Class 2		70	86	77	112
	Class 3		65	74	69	61
Logistic Regression	Class 0	77.69	83	57	68	105
	Class 1		83	91	87	94
	Class 2		82	86	84	112
	Class 3		59	77	67	61
SVM	Class 0	90.52	81	92	86	119
	Class 1		100	100	100	125
	Class 2		91	89	90	140
	Class 3		93	78	84	80
XGBoost	Class 0	85.34	79	70	74	119
	Class 1		98	100	99	125
	Class 2		82	87	84	140
	Class 3		80	82	81	80
Ensemble	Class 0	92.47	89	96	92	80
	Class 1		100	100	100	72
	Class 2		92	92	92	79
	Class 3		88	75	81	48

In contrast, the current study adopted a more focused approach, leveraging a refined ECG dataset primarily cantered on electrocardiographic signals. By narrowing the dataset's scope to specific physiological parameters, the study aimed to streamline the analysis process while mitigating the complexities associated with handling an extensive feature set. This targeted approach not only simplified data management but also facilitated the application of diverse algorithms, thereby achieving commendable accuracy levels.

In summary, the current study's success in attaining heightened accuracy levels vis-à-vis prior research can be attributed to several factors: meticulous data preprocessing techniques, the utilization of diverse algorithms, and a focused dataset selection strategy. By addressing the limitations identified in antecedent studies and adopting a comprehensive methodology, the current investigation aspired to propel the field of ECG analysis forward, thereby contributing to enhanced healthcare decision-making processes.

# CHAPTER 5

## CONCLUSION AND FUTURE SCOPE

In conclusion, our study has provided valuable insights into the performance of various machine learning models in the context of cardiovascular disorder detection through ECG image analysis. Across both individual lead and combined 12-lead datasets, Support Vector Machine (SVM) [29] emerged as a consistently robust performer, underscoring its efficacy in this domain. Furthermore, the integration of ensemble techniques, particularly through the Voting Classifier [30], has significantly enhanced overall model performance by leveraging the complementary strengths of individual classifiers while mitigating their respective weaknesses.

The utilization of combined 12-lead [31] data has yielded notable improvements in model prediction capabilities. Both SVM and XGBoost [32] achieved commendable accuracy levels when applied to pre-processed ECG data derived from the combined dataset. This underscores the importance of leveraging comprehensive datasets that capture a diverse range of cardiac perspectives [33] for enhanced diagnostic accuracy.

Looking ahead, there is immense potential for further advancements in this field. The development of more refined datasets, encompassing a broader range of physiological parameters and clinical variables, can contribute to the continued evolution of cardiovascular disorder detection models [34]. Deep learning models, with their ability to automatically extract intricate patterns from complex datasets, hold promise for further improving accuracy and robustness.

However, it is crucial to acknowledge and address ethical considerations in the handling of ECG image datasets. Preserving data privacy and ensuring fairness in model training and deployment are paramount concerns [35]. Biases inherent in the data, such as demographic disparities or systemic inequalities, must be carefully addressed to prevent unintended consequences and ensure equitable healthcare outcomes.

Furthermore, caution must be exercised in the interpretation and deployment of machine learning models for cardiovascular disease detection. While these models offer promising prospects for clinical application, their performance must be rigorously validated in real-world settings. The validation of these models in a real-time pipeline underscores their practical efficacy and reliability in identifying cardiovascular disorders through ECG image analysis [36].

Our study highlights the significant strides made in the field of cardiovascular disorder detection through machine learning and ECG image analysis. By leveraging advanced algorithms, comprehensive datasets, and ethical considerations, we can pave the way for more accurate, reliable, and equitable healthcare solutions. As we continue to refine and validate these models, we move closer towards realizing their full potential in clinical practice, ultimately improving patient outcomes and advancing the field of cardiovascular medicine.

In considering the future scope of utilizing machine learning and ECG images for cardiovascular disorder detection, several avenues for improvement and advancement come to the forefront. One prominent area of focus lies in leveraging deep learning algorithms to further enhance the accuracy and efficacy of diagnostic models [37].

Deep learning algorithms, such as convolutional neural networks (CNNs), have demonstrated remarkable capabilities in automatically extracting intricate patterns and features from complex datasets, including ECG images [38]. By harnessing the inherent hierarchical structure of neural networks [39], deep learning models can uncover subtle correlations and abnormalities in ECG signals that may evade traditional machine learning approaches.

One potential avenue for improvement involves the development of hybrid models that combine the strengths of both traditional machine learning techniques and deep learning algorithms. By integrating deep learning components into existing ensemble frameworks, such as the Voting Classifier [40] used in our study, we can capitalize on the superior feature extraction capabilities of CNNs while retaining the interpretability and generalization properties of traditional machine learning models.

Additionally, the continued refinement and expansion of ECG image datasets [41] hold promise for improving model accuracy and robustness. Efforts to collect larger, more diverse datasets encompassing a wide range of patient demographics, clinical conditions, and ECG signal variations can provide invaluable insights into the complexities of cardiovascular disorders. Moreover, the integration of multimodal data sources [42], such as clinical notes, genetic profiles, and wearable sensor data, can enrich the information available to machine learning models, enabling more comprehensive and personalized diagnostic assessments.

Furthermore, advancements in data preprocessing techniques, including signal denoising, artifact removal, and feature augmentation, can further enhance the quality of ECG data inputs, thereby improving model performance. The development of standardized protocols and best practices for data collection, annotation, and preprocessing can help ensure consistency and reproducibility across studies, facilitating the comparison and validation of machine learning models in real-world settings.

Addressing ethical considerations remains paramount in the ongoing development and deployment of machine learning models for cardiovascular disorder detection. Measures to safeguard patient privacy, mitigate algorithmic biases, and promote transparency and accountability in model development and deployment are essential to fostering trust and acceptance among healthcare providers and patients.

In conclusion, the future of cardiovascular disorder detection through machine learning and ECG image analysis holds immense promise for advancing clinical practice and improving patient outcomes. By embracing innovative approaches, leveraging deep learning algorithms, and addressing ethical considerations, we can unlock new insights into the complexities of cardiovascular health and usher in a new era of personalized and precision medicine.

# APPENDIX

## # FOR LOADING IMAGE

```
from skimage.io import imread
from skimage import color
import matplotlib.pyplot as plt
fig0 , ax0 = plt.subplots()
fig0.set_size_inches(20, 20)
image=imread('/content/Normal(2).jpg')
ax0.imshow(image)
plt.show()
from skimage.segmentation import slic
from skimage.color import label2rgb
```

## #plotting lead 1-12

```
fig , ax = plt.subplots(4,3)
fig.set_size_inches(20, 20)
x_counter=0
y_counter=0
for x,y in enumerate(Leads[:len(Leads)-1]):
    if (x+1)%3==0:
        ax[x_counter][y_counter].imshow(y)
        ax[x_counter][y_counter].axis('off')
        ax[x_counter][y_counter].set_title("Leads {}".format(x+1))
        x_counter+=1
        y_counter=0
    else:
        ax[x_counter][y_counter].imshow(y)
        ax[x_counter][y_counter].axis('off')
        ax[x_counter][y_counter].set_title("Leads {}".format(x+1))
        y_counter+=1
```

### **#importing gaussian filter and otsu threshold**

```
from skimage.filters import threshold_otsu, gaussian
from skimage.transform import resize
from numpy import asarray
fig2 , ax2 = plt.subplots(4,3)
fig2.set_size_inches(20, 20)
x_counter=0
y_counter=0
for x,y in enumerate(Leads[:len(Leads)-1]):
    #converting to gray scale
    grayscale = color.rgb2gray(y)
    #smoothing image
    blurred_image = gaussian(grayscale, sigma=0.7)
    global_thresh = threshold_otsu(blurred_image)
    binary_global = blurred_image < global_thresh
    binary_global = resize(binary_global, (300, 450))
    if (x+1)%3==0:
        ax2[x_counter][y_counter].imshow(binary_global, cmap="gray")
        ax2[x_counter][y_counter].axis('off')
        ax2[x_counter][y_counter].set_title("pre-processed Leads {} image".format(x+1))
        x_counter+=1
        y_counter=0
    else:
        ax2[x_counter][y_counter].imshow(binary_global, cmap="gray")
        ax2[x_counter][y_counter].axis('off')
        ax2[x_counter][y_counter].set_title("pre-processed Leads {} image".format(x+1))
        y_counter+=1
```

### **#plotting lead 13**

```
fig3 , ax3 = plt.subplots()
fig3.set_size_inches(20, 20)
```

### **#converting to gray scale**

```
grayscale = color.rgb2gray(Lead_11)
blurred_image = gaussian(grayscale, sigma=0.7)
global_thresh = threshold_otsu(blurred_image)
print(global_thresh)
```

### **#creating binary image based on threshold**

```
binary_global = blurred_image < global_thresh
ax3.imshow(binary_global,cmap='gray')
ax3.set_title("Leads 13")
ax3.axis('off')
```

### **# Perform contour to separate only the signal from image**

```
from skimage import measure
import scipy.ndimage as ndimage
contours = measure.find_contours(binary_global,0.9)
fig4, ax4 = plt.subplots()
plt.gca().invert_yaxis()
contours_shape = sorted([x.shape for x in contours])[::-1][0:1]
print(contours_shape)
for contour in contours:
    if contour.shape in contours_shape:
        test = resize(contour, (255, 2))
        ax4.plot(contour[:, 1], contour[:, 0],linewidth=1,color='black')
ax1.axis('image')
ax1.set_title("Sample pre-processed Leads 13 image")
\#converting image to signal
```

### **#Convert image to signal**

```
import pandas as pd
df = pd.DataFrame(test, columns = ['X','Y'])
fig5, ax5 = plt.subplots()
plt.gca().invert_yaxis()
```

```
ax5.plot(df['Y'],df['X'],linewidth=1,color='black',linestyle='solid')
```

```
fig5.savefig('Lead13_Signal.png')
```

### **#Save signal to CSV format**

```
df.to_csv('data.csv',index=False)
```

```
test_df=pd.read_csv('data.csv')
```

```
test_df
```

### **#scaling the data and testing**

```
from sklearn.preprocessing import MinMaxScaler
```

```
scaler = MinMaxScaler()
```

```
fit_transform_data = scaler.fit_transform(df)
```

```
Normalized_Scaled=pd.DataFrame(fit_transform_data, columns = ['X','Y'])
```

```
Normalized_Scaled
```

### **#plotting 1D signal**

```
import pandas as pd
```

```
test_plot_df = pd.DataFrame(test_scaled_df_X, columns = ['X'])
```

```
fig6, ax6 = plt.subplots()
```

```
plt.gca().invert_yaxis()
```

```
ax6.plot(test_plot_df,linewidth=1,color='black',linestyle='solid')
```

## **# FUNCTION FOR IMAGE LEADS(1-13) PRE-PROCESSING**

```
def Convert_Image_Lead(image_file,parent_folder):
```

```
image=imread('{parent}/{image_file}'.format(parent=str(parent_folder),image_file=str(image_file)),p  
lugin='matplotlib')
```

```
Lead_1 = image[300:600, 150:643]
```

```
Lead_2 = image[300:600, 646:1135]
```

```
Lead_3 = image[300:600, 1140:1626]
```

```
Lead_4 = image[300:600, 1630:2125]
```

```
Lead_5 = image[600:900, 150:643]
```

```
Lead_6 = image[600:900, 646:1135]
```

```
Lead_7 = image[600:900, 1140:1626]
```

```
Lead_8 = image[600:900, 1630:2125]
```

```
Lead_9 = image[900:1200, 150:643]
```

```
Lead_10 = image[900:1200, 646:1135]
```



```

Lead_11 = image[900:1200, 1140:1626]
Lead_12 = image[900:1200, 1630:2125]
Lead_13 = image[1250:1480, 150:2125]

Leads=[Lead_1,Lead_2,Lead_3,Lead_4,Lead_5,Lead_6,Lead_7,Lead_8,Lead_9,Lead_10,Lead_11,Lead_12,Lead_13]

folder_name= re.sub('.jpg', '',image_file)

for x,y in enumerate(Leads):

    fig , ax = plt.subplots()

    #fig.set_size_inches(20, 20)

    ax.imshow(y)

    ax.axis('off')

    ax.set_title("Leads {0}".format(x+1))

    if (os.path.exists(parent_folder+'/'+folder_name)):

        pass

    else:

        os.makedirs(parent_folder+'/'+folder_name)

    plt.close('all')

    plt.ioff()

fig.savefig('{parent}/{folder_name}/Lead_{x}_Signal.png'.format(folder_name=folder_name,x=x+1,
parent=parent_folder))

extract_signal_leads(Leads,folder_name,parent_folder)

#extract_only signal from images

def extract_signal_leads(Leads,folder_name,parent):

    #looping through image list containg all leads from 1-13

    for x,y in enumerate(Leads):

        #creating subplot

        fig1 , ax1 = plt.subplots()

        #set fig size

        #fig1.set_size_inches(20, 20)

        #converting to gray scale

```

```

grayscale = color.rgb2gray(y)

blurred_image = gaussian(grayscale,sigma=0.7)
global_thresh = threshold_otsu(blurred_image)
binary_global = blurred_image < global_thresh
    if x!=12:
        binary_global = resize(binary_global, (300, 450))

ax1.imshow(binary_global,cmap="gray")
ax1.axis('off')
ax1.set_title("pre-processed Leads {} image".format(x+1))
plt.close('all')
plt.ioff()

fig1.savefig('{parent}/{folder_name}/Lead_{x}_preprocessed_Signal.png'.format(folder_name=folder_name,x=x+1,parent=parent))

fig7 , ax7 = plt.subplots()
plt.gca().invert_yaxis()
contours = measure.find_contours(binary_global,0.8)
contours_shape = sorted([x.shape for x in contours])[::-1][0:1]
for contour in contours:
    if contour.shape in contours_shape:
        test = resize(contour, (255, 2))
        ax7.plot(test[:, 1], test[:, 0],linewidth=1,color='black')
ax7.axis('image')
ax7.set_title("Contour {} image".format(x+1))
plt.close('all')
plt.ioff()

#save the image

fig7.savefig('{parent}/{folder_name}/Lead_{x}_Contour_Signal.png'.format(folder_name=folder_name,x=x+1,parent=parent))

lead_no=x

scale_csv_1D(test,lead_no,folder_name,parent)

```

```

def convert_csv(test,lead_no,folder_name,parent):
#convert contour to dataframe
    target=folder_name[0:2]
    df = pd.DataFrame(test, columns = ['X','Y'])
    df['Target']=target
    #x_axis= 'Lead_{lead_no}_X'.format(lead_no=lead_no)
    #y_axis= 'Lead_{lead_no}_Y'.format(lead_no=lead_no)
    fig5, ax5 = plt.subplots()
    #convert to CSV

df.to_csv('{parent}/{folder_name}/{lead_no}.csv'.format(lead_no=lead_no+1,parent=parent,folder_n
ame=folder_name),index=False)

def scale_csv(test,lead_no,folder_name,parent):
    #scaling the data and testing
    target=folder_name[0:2]
    scaler = MinMaxScaler()
    fit_transform_data = scaler.fit_transform(test)
    Normalized_Scaled=pd.DataFrame(fit_transform_data, columns = ['X','Y'])
    Normalized_Scaled=Normalized_Scaled.T
    Normalized_Scaled['Target']=target
    #scaled_data to CSV
    if (os.path.isfile('{parent}/Scaled_{lead_no}.csv'.format(lead_no=lead_no+1,parent=parent))):

Normalized_Scaled.to_csv('{parent}/Scaled_{lead_no}.csv'.format(lead_no=lead_no+1,parent=paren
t), mode='a', header=False,index=False)
    else:

Normalized_Scaled.to_csv('{parent}/Scaled_{lead_no}.csv'.format(lead_no=lead_no+1,parent=paren
t,folder_name=folder_name),index=False)

def scale_csv_1D(test,lead_no,folder_name,parent):
    target=folder_name[0:2]
    #scaling the data and testing

```

### **#Function for csv conversion and scaling**

```
scaler = MinMaxScaler()

fit_transform_data = scaler.fit_transform(test)

Normalized_Scaled=pd.DataFrame(fit_transform_data[:,0], columns = ['X'])

fig6, ax6 = plt.subplots()

plt.gca().invert_yaxis()

ax6.plot(Normalized_Scaled,linewidth=1,color='black',linestyle='solid')

plt.close('all')

plt.ioff()


fig6.savefig('{parent}/{folder_name}/ID_Lead_{lead_no}_Signal.png'.format(folder_name=folder_name,lead_no=lead_no+1,parent=parent))

Normalized_Scaled=Normalized_Scaled.T

Normalized_Scaled['Target']=target

#scaled_data to CSV

if
(os.path.isfile('{parent}/scaled_data_1D_{lead_no}.csv'.format(lead_no=lead_no+1,parent=parent))):

Normalized_Scaled.to_csv('{parent}/scaled_data_1D_{lead_no}.csv'.format(lead_no=lead_no+1,parent=parent), mode='a', header=False,index=False)

else:

Normalized_Scaled.to_csv('{parent}/scaled_data_1D_{lead_no}.csv'.format(lead_no=lead_no+1,parent=parent,folder_name=folder_name),index=False)
```

## **# Streamlit app code**

```
import streamlit as st
```

```
from Ecg import ECG
```

```
#intialize ecg object
```

```
ecg = ECG()
```

### **#get the uploaded image**

```
uploaded_file = st.file_uploader("Choose a file")
```

```
if uploaded_file is not None:
```

#### **# call the getimage method**

```
ecg_user_image_read = ecg.getImage(uploaded_file)
```

```
st.image(ecg_user_image_read)
```

#### **#call the convert Grayscale image method**

```
ecg_user_gray_image_read = ecg.GrayImgae(ecg_user_image_read)
```

#### **#create Streamlit Expander for Gray Scale**

```
my_expander = st.expander(label='Gray SCALE IMAGE')
```

```
with my_expander:
```

```
    st.image(ecg_user_gray_image_read)
```

#### **#call the Divide leads method**

```
dividing_leads=ecg.DividingLeads(ecg_user_image_read)
```

#### **#streamlit expander for dividing leads**

```
my_expander1 = st.expander(label='DIVIDING LEAD')
```

```
with my_expander1:
```

```
    st.image('Leads_1-12_figure.png')
```

```
    st.image('Long_Lead_13_figure.png')
```

#### **#call the preprocessed leads method**

```
ecg_preprocessed_leads = ecg.PreprocessingLeads(dividing_leads)
```

#### **#streamlit expander for preprocessed leads**

```
my_expander2 = st.expander(label='PREPROCESSED LEAD')
```

```
with my_expander2:
```

```
    st.image('Preprossed_Leads_1-12_figure.png')
```

```
    st.image('Preprossed_Leads_13_figure.png')
```

## Project Preview :

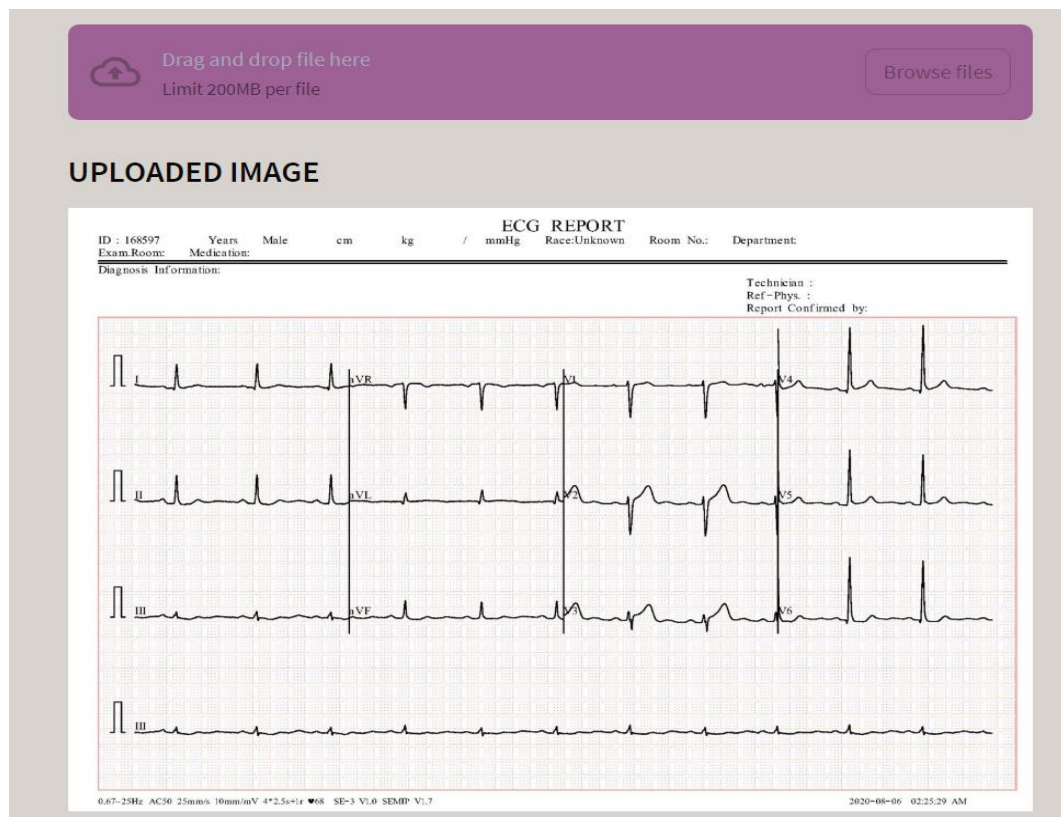


Fig 11. Uploading Image

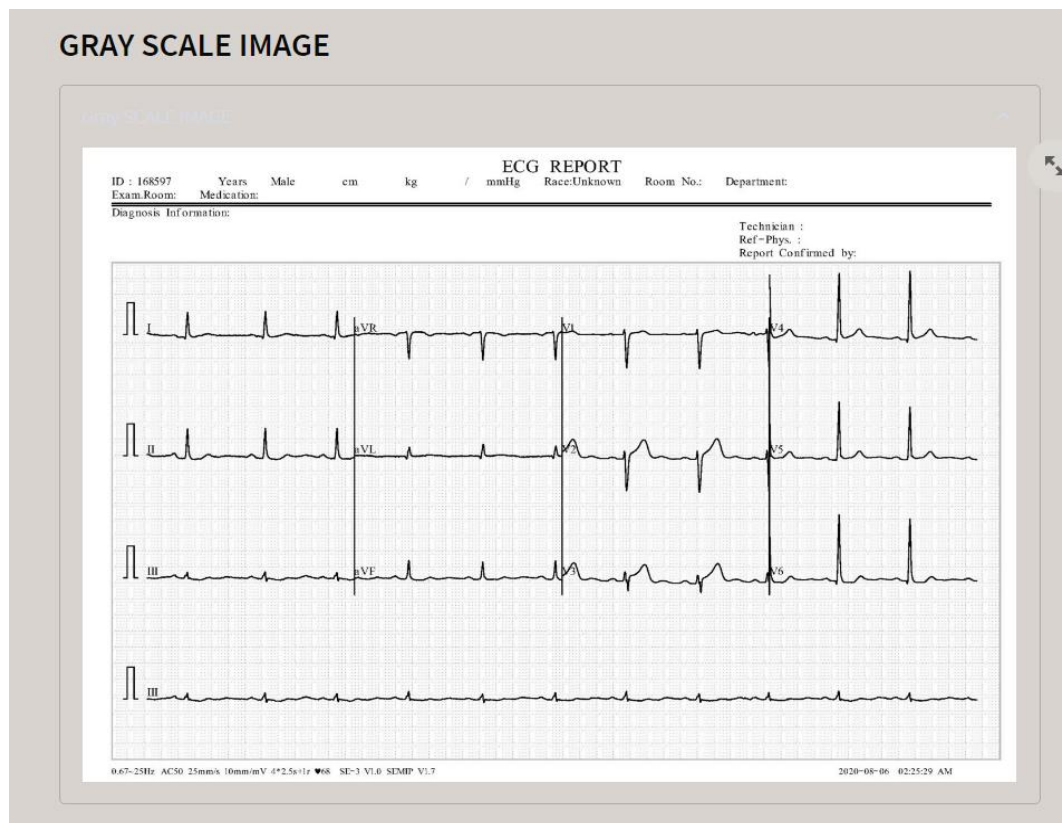


Fig 12. GrayScale Filtering

## DIVIDING LEADS

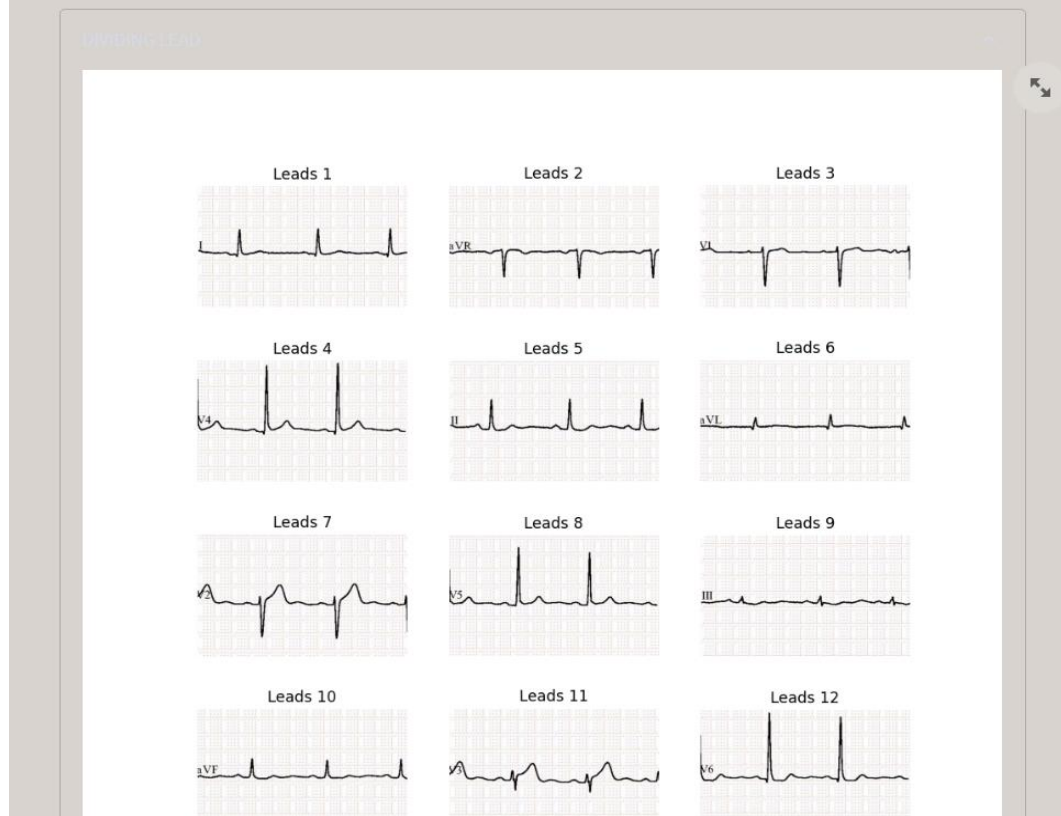


Fig 13. Lead Division

## PREPROCESSED LEADS

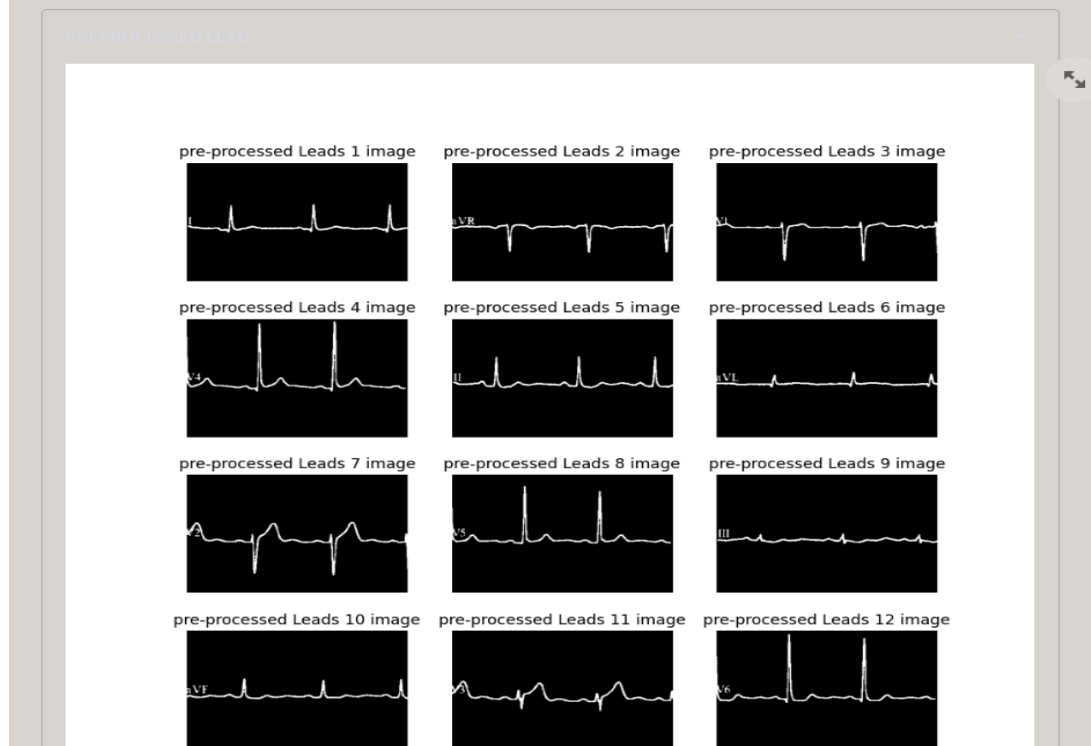


Fig 14. Thresholding

## EXTRACTING SIGNALS(1-12)

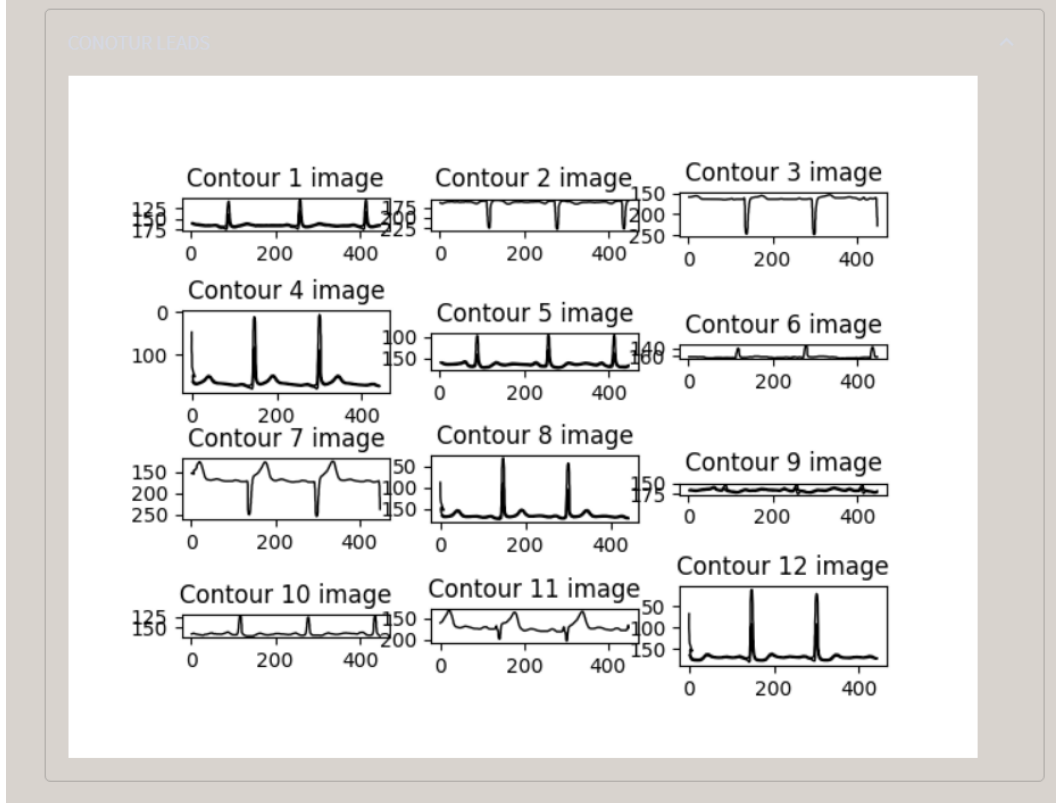


Fig 15. Extraction of Signals

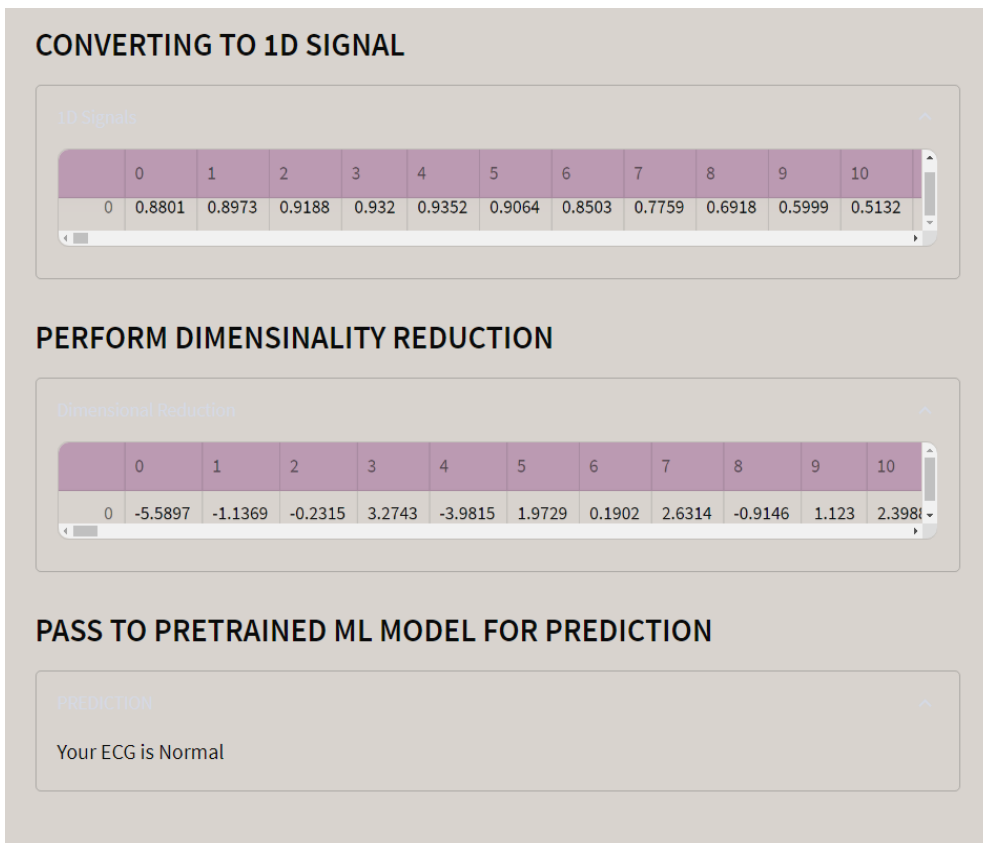


Fig 16. Conversion into 1D signals



# REFERENCES

1. World Health Organization, "Cardiovascular Diseases", from [https://www.who.int/health-topics/cardiovascular-diseases/#tab=tab\\_1](https://www.who.int/health-topics/cardiovascular-diseases/#tab=tab_1)
2. Mike Cadogan, "ECG Lead Positioning", 2022. <https://litfl.com/ecg-lead-positioning/>
3. 12-Lead ECG electrode points. Adapted from "ECG Signal Quality Assessments of a Small Bipolar Single-Lead Wearable Patch Sensor" by Rajbhandary, P.L., Nallathambi, G., Selvaraj, N. et al. *Cardiovasc Eng Tech* 13, 783–796 (2022). <https://link.springer.com/article/10.1007/s13239-022-00617-3>
4. Adyasha Rath, Debahuti Mishra, Ganapati Panda, Suresh Chandra Satapathy, "Heart disease detection using deep learning methods from imbalanced ECG samples", *Biomedical Signal Processing and Control*, Volume 68, 2021, 102820, ISSN 1746-8094, <https://doi.org/10.1016/j.bspc.2021.102820>
5. Rohit Bharti, Aditya Khamparia, Mohammad Shabaz, Gaurav Dhiman, Sagar Pande, Parneet Singh, "Prediction of Heart Disease Using a Combination of Machine Learning and Deep Learning", *Computational Intelligence and Neuroscience*, vol. 2021, Article ID 8387680, 11 pages, 2021. <https://doi.org/10.1155/2021/8387680>
6. M.M. Al Rahhal, Yakoub Bazi, Haikel AlHichri, Naif Alajlan, Farid Melgani, R.R. Yager, "Deep learning approach for active classification of electrocardiogram signals", *Information Sciences*, Volume 345, 2016, Pages 340-354, ISSN 0020-0255, <https://doi.org/10.1016/j.ins.2016.01.082>
7. Guidi, Gabriele, Roberto Miniati, Matteo Mazzola, and Ernesto Iadanza. 2016. "Case Study: IBM Watson Analytics Cloud Platform as Analytics-as-a-Service System for Heart Failure Early Detection" *Future Internet* 8, no. 3: 32. <https://doi.org/10.3390/fi8030032>
8. Ali Haider Khan, Muzammil Hussain, Muhammad Kamran Malik, "ECG Images dataset of Cardiac and COVID-19 Patients", *Data in Brief*, Volume 34, 2021, 106762, ISSN 2352-3409, <https://doi.org/10.1016/j.dib.2021.106762>
9. Keogh, E., Mueen, A. (2017). Curse of Dimensionality. In: Sammut, C., Webb, "G.I. (eds) *Encyclopedia of Machine Learning and Data Mining*". Springer, Boston, MA. [https://doi.org/10.1007/978-1-4899-7687-1\\_192](https://doi.org/10.1007/978-1-4899-7687-1_192)
10. Liu, Huan, and Hiroshi Motoda, eds. "Feature extraction, construction and selection: A data mining perspective". Vol. 453. Springer Science & Business Media, 1998.
11. Rohit Bharti, Aditya Khamparia, Mohammad Shabaz, Gaurav Dhiman, Sagar Pande, Parneet Singh, "Prediction of Heart Disease Using a Combination of Machine Learning and Deep Learning", *Computational Intelligence and Neuroscience*, vol. 2021, Article ID 8387680, 11 pages, 2021. <https://doi.org/10.1155/2021/8387680>
12. Rajbhandary, P.L., Nallathambi, G., Selvaraj, N. et al. "ECG Signal Quality Assessments of a Small Bipolar Single-Lead Wearable Patch Sensor". *Cardiovasc Eng Tech* 13, 783–796 (2022). <https://doi.org/10.1007/s13239-022-00617-3>
13. P. Li et al., "High-Performance Personalized Heartbeat Classification Model for Long-Term ECG Signal," in *IEEE Transactions on Biomedical Engineering*, vol. 64, no. 1,

- pp. 78-86, Jan. 2017, doi: 10.1109/TBME.2016.2539421
14. Ketu, S., Mishra, P.K. "Empirical Analysis of Machine Learning Algorithms on Imbalance Electrocardiogram Based Arrhythmia Dataset for Heart Disease Detection". Arab J Sci Eng 47, 1447–1469 (2022). <https://doi.org/10.1007/s13369-021-05972-2>
  15. V. Jahmunah, E.Y.K. Ng, Tan Ru San, U. Rajendra Acharya, "Automated detection of coronary artery disease, myocardial infarction and congestive heart failure using GaborCNN model with ECG signals", Computers in Biology and Medicine, Volume 134, 2021, 104457, ISSN 0010-4825, <https://doi.org/10.1016/j.compbimed.2021.104457>
  16. Khan, Ali Haider; Hussain, Muzammil (2021), "ECG Images dataset of Cardiac Patients ", Mendeley Data, V2, doi: 10.17632/gwbz3fsgp8.2
  17. Swati Shilaskar, Ashok Ghatol, "Feature selection for medical diagnosis : Evaluation for cardiovascular diseases, Expert Systems with Applications", Volume 40, Issue 10, 2013, Pages 4146-4153, ISSN 0957-4174, <https://doi.org/10.1016/j.eswa.2013.01.032>
  18. Kaushalya Dissanayake, Md Gapar Md Johar, "Comparative Study on Heart Disease Prediction Using Feature Selection Techniques on Classification Algorithms", Applied Computational Intelligence and Soft Computing, vol. 2021, Article ID 5581806, 17 pages, 2021. <https://doi.org/10.1155/2021/5581806>
  19. Defu He, Si Xiong, "Image Processing Design and Algorithm Research Based on Cloud Computing", *Journal of Sensors*, vol. 2021, Article ID 9198884, 10 pages, 2021. <https://doi.org/10.1155/2021/9198884>
  20. S. Y. Wang, Y. Zhao, and L. Wen, "PCB welding spot detection with image processing method based on automatic threshold image segmentation algorithm and mathematical morphology," *Circuit World*, vol. 42, no. 3, pp. 97–103, 2016.  
View at: [Publisher Site](#) | [Google Scholar](#)
  21. S. M. Hussain, F. U. Farrukh, S. Su, Z. Wang, and H. Chen, "CMOS image sensor design and image processing algorithm implementation for total hip arthroplasty surgery," *IEEE Transactions on Biomedical Circuits and Systems*, vol. 13, no. 6, pp. 1383–1392, 2019. View at: [Google Scholar](#)
  22. Khurana S, Das S, Frishman WH, Aronow WS, Frenkel D. Lead Extraction- Indications, Procedure, and Future Directions. *Cardiol Rev*. 2023 Sep 20. doi: 10.1097/CRD.0000000000000610. Epub ahead of print. PMID: 37729602.
  23. Maria G Bongiorno, Haran Burri, Jean C Deharo, Christoph Starck, Charles Kennergren, Laszlo Saghy, Archana Rao, Carlo Tascini, Nigel Lever, Andrzej Kutarski, Ignacio Fernandez Lozano, Neil Strathmore, Roberto Costa, Laurence Epstein, Charles Love, Carina Blomstrom-Lundqvist, ESC Scientific Document Group , 2018 EHRA expert consensus statement on lead extraction: recommendations on definitions, endpoints, research trial design, and data collection requirements for clinical scientific studies and registries: endorsed by APHRS/HRS/LAHRs, *EP*

24. Xin-Yi Gong, Hu Su, De Xu, Zheng-Tao Zhang, Fei Shen and Hua-Bin Yang. An Overview of Contour Detection Approaches. *International Journal of Automation and Computing*, vol. 15, no. 6, pp. 656-672, 2018. DOI: [10.1007/s11633-018-1117-z](https://doi.org/10.1007/s11633-018-1117-z)
25. D. R. Martin, C. C. Fowlkes, J. Malik. Learning to detect natural image boundaries using local brightness, color, and texture cues. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 26, no. 5, pp. 530–549, 2004. DOI: <https://doi.org/10.1109/TPAMI.2004.1273918>.
26. Ammar Mohammed, Rania Kora, A comprehensive review on ensemble deep learning: Opportunities and challenges, *Journal of King Saud University - Computer and Information Sciences*, Volume 35, Issue 2, 2023.
27. Pei Ying Lim, Wan Yoke Chin, Yih Hern Kong, "Exploring the Effectiveness of Ensemble Learning for Business-to-Business (B2B) Demand Forecasting: An Empirical Analysis", *2024 3rd International Conference on Digital Transformation and Applications (ICDXA)*, pp.226-230, 2024.
28. Cheng-Yuan Ho, Jia-En Wang, "Mastering the Mask: Harnessing Ensemble Learning for Defect Detection", *2023 RIVF International Conference on Computing and Communication Technologies (RIVF)*, pp.230-234, 2023.
29. Mehta, S. S., & Lingayat, N. S. (2008). Development of SVM based ECG Pattern Recognition Technique. *IETE Journal of Research*, 54(1), 5–11. <https://doi.org/10.1080/03772063.2008.10876176>
30. Kuncheva, L.I., Rodríguez, J.J. A weighted voting framework for classifiers ensembles. *Knowl Inf Syst* 38, 259–275 (2014). <https://doi.org/10.1007/s10115-012-0586-6>
31. Johnson Francis, ECG monitoring leads and special leads, *Indian Pacing and Electrophysiology Journal*, Volume 16, Issue 3, 2016, Pages 92-95, ISSN 0972-6292, <https://doi.org/10.1016/j.ipej.2016.07.003>.
32. Yao Chen *et al* 2018 *Physiol. Meas.* 39 104006 "Classification of short single-lead electrocardiograms (ECGs) for atrial fibrillation detection using piecewise linear spline and XGBoost" DOI: 10.1088/1361-6579/aadf0f
33. Gopinathannair, R., Merchant, F.M., Lakkireddy, D.R. *et al.* COVID-19 and cardiac arrhythmias: a global perspective on arrhythmia characteristics and management strategies. *J Interv Card Electrophysiol* 59, 329–336 (2020). <https://doi.org/10.1007/s10840-020-00789-9>
34. V. Chalana, D. T. Linker, D. R. Haynor and Yongmin Kim, "A multiple active contour model for cardiac boundary detection on echocardiographic sequences," in *IEEE*

*Transactions on Medical Imaging*, vol. 15, no. 3, pp. 290-298, June 1996, doi: 10.1109/42.500138.

35. A. I. Hossain, S. Sikder, A. Das and A. Dey, "Applying Machine Learning Classifiers on ECG Dataset for Predicting Heart Disease," *2021 International Conference on Automation, Control and Mechatronics for Industry 4.0 (ACMI)*, Rajshahi, Bangladesh, 2021, pp. 1-6, doi: 10.1109/ACMI53878.2021.9528169.
36. McNames, J., Aboy, M. Reliability and accuracy of heart rate variability metrics versus ECG segment duration. *Med Bio Eng Comput* 44, 747–756 (2006).  
<https://doi.org/10.1007/s11517-006-0097-2>
37. Xinwen Liu, Huan Wang, Zongjin Li, Lang Qin, Deep learning in ECG diagnosis: A review, *Knowledge-Based Systems*, Volume 227, 2021, 107187, ISSN 0950-7051,  
<https://doi.org/10.1016/j.knosys.2021.107187>.
38. M. B. Abubaker and B. Babayiğit, "Detection of Cardiovascular Diseases in ECG Images Using Machine Learning and Deep Learning Methods," in *IEEE Transactions on Artificial Intelligence*, vol. 4, no. 2, pp. 373-382, April 2023, doi: 10.1109/TAI.2022.3159505.
39. Chen, H. M., Tsai, K. H., Qi, G. Z., Yang, J. C. S., & Amini, F. (1995). Neural network for structure control. *Journal of Computing in Civil Engineering*, 9(2), 168-176.  
[https://doi.org/10.1061/\(ASCE\)0887-3801\(1995\)9:2\(168\)](https://doi.org/10.1061/(ASCE)0887-3801(1995)9:2(168))
40. P. Aublin *et al.*, "Cardiac Abnormality Detection Based on an Ensemble Voting of Single-Lead Classifier Predictions," *2021 Computing in Cardiology (CinC)*, Brno, Czech Republic, 2021, pp. 1-4, doi: 10.23919/CinC53138.2021.9662824.
41. Song, Weibo. 2020. "A New Method for Refined Recognition for Heart Disease Diagnosis Based on Deep Learning" *Information* 11, no. 12: 556.  
<https://doi.org/10.3390/info11120556>
42. Yang, Z., Zhuang, X., & Bird, C. et al. (2019). Performing Sparse Regularization and Dimension Reduction Simultaneously in Multimodal Data Fusion. *Frontiers in Neuroscience*, 13. <https://doi.org/10.3389/fnins.2019.00642>

# **PUBLICATION**

D. Rautela, M. Bajeli, A. Kumar and H. Vaidya, "Identifying Cardiovascular Disorders Through ECG Image Analysis," 2024 International Conference on Cognitive Robotics and Intelligent Systems (ICC - ROBINS), Coimbatore, India, 2024, pp. 607-612, doi: 10.1109/ICC-ROBINS60238.2024.10533939.



# CERTIFICATIONS



**KPR Institute of Engineering and Technology**  
LearnBeyond  
(Autonomous, NAAC "A")

Certificate No.: KPR / IEEE / CFRD / P\_325



Technical Sponsorship by  
**IEEE**  
MAHARAJA SECTION  
IEEE KPLRE COMPLIANT GSN  
879-B-2082-1234-B

2024 International Conference on Cognitive Robotics and Intelligent Systems (ICC – ROBINS)

17-19, April 2024 | Coimbatore, India

Organized by Centre for Research and Development (CFRD), KPRIET

## Certificate of Presentation

This is to certify that Dr./Mr./Ms. Devansh Rautela of Graphic Era Hill University has successfully presented the paper entitled ROBIN113 - Identifying Cardiovascular Disorders Through ECG Image Analysis in the IEEE sponsored 2024 International Conference on Cognitive Robotics and Intelligent Systems (ICC - ROBINS) held at KPR Institute of Engineering and Technology, Coimbatore, India during 17-19, April 2024.



Dr. H. James Deva Koresh  
Conference Chair



Dr. R. Maheswar  
General Chair



Dr. D. Saravanan  
Patron



**KPR Institute of Engineering and Technology**  
LearnBeyond  
(Autonomous, NAAC "A")

Certificate No.: KPR / IEEE / CFRD / P\_326



Technical Sponsorship by  
**IEEE**  
MAHARAJA SECTION  
IEEE KPLRE COMPLIANT GSN  
879-B-2082-1234-B

2024 International Conference on Cognitive Robotics and Intelligent Systems (ICC – ROBINS)

17-19, April 2024 | Coimbatore, India

Organized by Centre for Research and Development (CFRD), KPRIET

## Certificate of Presentation

This is to certify that Dr./Mr./Ms. Himadri Vaidya of Graphic Era Hill University has successfully presented the paper entitled ROBIN113 - Identifying Cardiovascular Disorders Through ECG Image Analysis in the IEEE sponsored 2024 International Conference on Cognitive Robotics and Intelligent Systems (ICC - ROBINS) held at KPR Institute of Engineering and Technology, Coimbatore, India during 17-19, April 2024.



Dr. H. James Deva Koresh  
Conference Chair



Dr. R. Maheswar  
General Chair



Dr. D. Saravanan  
Patron



**KPR Institute of Engineering and Technology**  
LearnBeyond  
(Autonomous, NAAC "A")

Certificate No.: KPR / IEEE / CFRD / P\_327



Technical Sponsorship by  
**IEEE**  
MAHARAJA SECTION  
IEEE KPLRE COMPLIANT GSN  
879-B-2082-1234-B

2024 International Conference on Cognitive Robotics and Intelligent Systems (ICC – ROBINS)

17-19, April 2024 | Coimbatore, India

Organized by Centre for Research and Development (CFRD), KPRIET

## Certificate of Presentation

This is to certify that Dr./Mr./Ms. Mohit Bajeli of Graphic Era Hill University has successfully presented the paper entitled ROBIN113 - Identifying Cardiovascular Disorders Through ECG Image Analysis in the IEEE sponsored 2024 International Conference on Cognitive Robotics and Intelligent Systems (ICC - ROBINS) held at KPR Institute of Engineering and Technology, Coimbatore, India during 17-19, April 2024.



Dr. H. James Deva Koresh  
Conference Chair



Dr. R. Maheswar  
General Chair



Dr. D. Saravanan  
Patron

# Identifying Cardiovascular Disorders Through ECG Image Analysis

Devansh Rautela  
*Computer Science and Engineering*  
*Graphic Era Hill University*  
Dehradun, India  
maildevanshrautela@gmail.com

Mohit Bajeli  
*Computer Science and Engineering*  
*Graphic Era Hill University*  
Dehradun, India  
mohitbajeli7617@gmail.com

Anmool Kumar  
*Computer Science and Engineering*  
*Graphic Era Hill University*  
Dehradun, India  
kanmol236@gmail.com

Himadri Vaidya  
*Computer Science and Engineering*  
*Graphic Era Hill University*  
Dehradun, India  
himadrivaidya8@gmail.com

**Abstract**—An electrocardiogram (ECG) is a crucial tool for assessing the electrical and muscular functions of the heart. It plays a key role in detecting cardiovascular diseases (CVD), which can be challenging to diagnose. Traditionally, ECG charts were maintained in paper form, making manual examination time-consuming and requiring specialized expertise. This research proposes a methodology to digitize ECG records, leveraging them to train machine learning models for cardiovascular disease detection. The primary focus is on identifying four major cardiac abnormalities: abnormal heartbeat, myocardial infarction, history of myocardial infarction, and a normal person class. Due to limited datasets, other categories cannot be classified with high accuracy. Various techniques, including Gaussian filtering, thresholding, and grayscale, are employed for data cleaning and feature formation. Preprocessed images are resized into 13 leads, facilitating the examination of ECG reports converted into 1D signals. Multiple algorithms are implemented such as SVM (Support Vector Machine), KNN (K-nearest neighbors), and Logistic Regression, and then ensembled using a classifier i.e. Voting-Based Classifier. The final model selection is based on criteria such as accuracy, precision, recall, and F1-score. The proposed Ensembled Technique using a Voting-Based Classifier outperforms the existing works; it achieves 92.5% accuracy, 91% recall, 92% precision, and 92% f1-score overall performance in identifying four major cardiac abnormalities. In essence, this study demonstrates how machine learning, when applied to a substantial dataset, can efficiently identify and predict specific categories of diseases within the realm of cardiovascular health.

**Index Terms**—Electrocardiogram (ECG), Cardiovascular diseases (CVD), Support Vector Machine, Voting Based Classifier.

## I. INTRODUCTION

Cardiovascular diseases (CVDs) are disorders that affect the heart and blood vessels and come under the non-communicable disorders category, causing an estimated 17.9 million people to die from CVDs. This represents 31% of all global deaths according to the published in 2020 by the World Health Organisation (WHO) [1]. According to the WHO report, the reason for about 85% of deaths was due to heart attacks and strokes. People at high risk of CVDs demonstrate elevated

blood pressure, glucose levels, and an imbalance in the ECG or EKG.

Early identification of those at the highest risk of CVD can help prevent it. An electrocardiogram (ECG) is a test for the detection of electrical activity in the heart using sensors placed on the skin over the chest, typically involving the application of 10 electrode points. These electrodes are strategically positioned in specific locations to create a 12-lead ECG, allowing for a comprehensive assessment of the heart's electrical conduction through different perspectives. Research [2] describes the 10 electrode points consisting of six precordial lead points that are placed on the chest and four limb leads placed on the arms and legs, capturing these electrical signals generated by the heart from multiple angles for a more refined and accurate analysis of cardiac function. The electrode points are shown in Figure 1 [3]. RA (Right arm electrode), LA (Left arm electrode), RL (Right leg electrode), and LL (Left leg electrode) are those with the help of which the limb leads and augmented leads are obtained. It is quick, painless, and can detect a variety of heart problems. Using ECG, one can achieve early detection of CVDs, and can be easily prevented using guided treatment, ultimately reducing CVD events.

Hence, it is important to detect irregular heart rhythms using ECG signals. However, manually interpreting ECG recordings through manual examination is vital, which is prolonged, monotonous, and requires expertise. Therefore, various machine and deep learning methods have been trained for the automatic identification of cardiovascular diseases, as discussed in research papers [4].

In this research, our primary focus is on converting non-uniform ECG images into 12-lead-based ECG images and structuring the resultant dataset. This dataset serves as a valuable resource for identifying cardiovascular disorders and classifying individuals into four distinct categories: Myocardial Infarction (MI), patients with abnormal heartbeats, normal individuals, and patients with a history of MI. To achieve

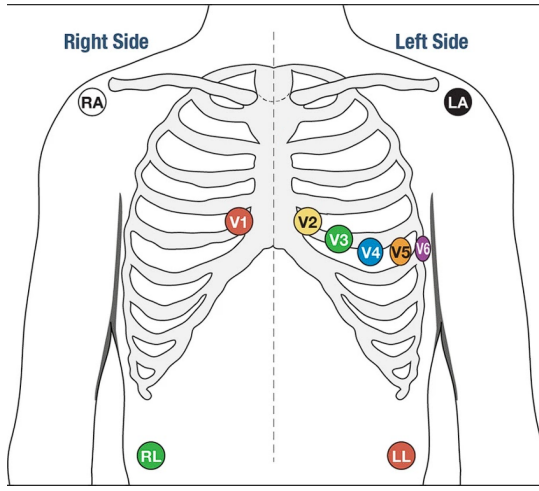


Fig. 1. 12-Lead ECG electrode points.

this objective, we leverage a diverse set of machine learning models, including K-Nearest Neighbors (KNN), Support Vector Machine (SVM), Logistic Regression, and the powerful decision tree-based XGBoost algorithm.

To enhance the predictive capabilities and robustness of our models, we incorporate ensemble techniques such as the Voting Classifier. This ensemble approach combines the strengths of individual models, including the newly introduced XGBoost, to obtain the highest accuracy results.

The combination of these models, when trained with diverse and sustainable data, makes a substantial contribution to the research and development landscape across various fields. This research not only tackles practical challenges in medical image processing but also illuminates the collective impact of machine learning and ensemble techniques. Importance of using high-quality data for model training, our goal is to demonstrate how these methods improve diagnosis and decision-making ability. The incorporation of XGBoost, known for its efficiency in handling structured datasets and decision tree-based learning, enhances the research findings and creates opportunities for innovative applications in healthcare and beyond.

## II. LITERATURE STUDY

In the literature review, several studies have examined different approaches to identifying abnormal heartbeats using multiple techniques such as heart sounds, heartbeats, blood pressure, and other parameters related to heart function. Researchers are concentrating on the use of Electrocardiogram images i.e. ECG for predicting and identifying Cardiovascular Diseases. For decades, research has been done to create machine learning models that can predict health diseases.

Research [5, 6] suggested an approach utilizing deep neural networks for selecting optimal features from ECGs to enhance performance. Additionally, Guidi [7] contributed a clinical decision support system for detecting heart failures at an early stage, employing various machine learning and deep learning models, including neural networks, support vector machines,

random forest, and CART algorithms. Notably, random forest and CART achieved an 87.6 percent accuracy, surpassing other classification methods.

In a research by Ali Haider Khan [8] they curated a dataset of ECG images containing 1937 distinct patient records with senior medical professionals. The retrieved data does not contain any details of patients apart from the ECG signals image and the CVD associated with it. The manual reviewing process resulted in a categorised 12 leads-based ECG images dataset used development of automatic diagnosis tools for healthcare institutes.

In machine learning, a prevalent challenge is the high and vast dimensionality of datasets, often known as the curse of dimensionality [9]. To address this, dimensionality reduction techniques were tested, leading to improved accuracy [10]. Bharti's research [11] advocated for the use of normalized datasets with outlier handling and feature selection, resulting in superior performance compared to previous techniques. When comparing different machine learning algorithms, KNeighbors emerged as the winner with 84.86% accuracy, showcasing a precision of 77.7% and a specificity of 80%. However, deep learning algorithms achieved even higher accuracy at 94.2%, underscoring their superiority.

Articles by Paurakh Lal Rajbhandary [12] highlighted emerging technologies in wearable fields, employing patch sensors to make ECG signals accessible. The procedure for obtaining ECGs and creating a 12-lead image with ten electrode points on the chest was explained.

Multiple studies [13, 14, 15] delved into the use of imbalanced ECG samples for detecting heart diseases. Six classification models categorized into two machine learning, three deep learning, and one ensemble model, were employed. The accuracy of MLP (Multilayer Perceptron) models is slightly higher as compared to the LSTM (Long Short-Term Memory) model during the training phase.

## III. METHODOLOGY

Our method is painstakingly created to handle the complexities of handling and interpreting electrocardiogram (ECG) pictures to identify different cardiovascular illnesses. Many crucial steps are included in the process to guarantee that the dataset is enhanced and prepared, and then sophisticated pre-processing methods and machine learning models are applied.

### A. Dataset Description

In the initial stage, our focus is on curating a comprehensive ECG image dataset sourced from reputable medical institutions, research centers, and publicly available repositories. There are publicly available datasets that primarily consist of incompatible time-series data which are inconsistent with the data needed for the study. Our dataset, retrieved from diverse cardiac institutes under the Ch. Pervaiz Elahi Institute of Cardiology Multan, Pakistan, includes annotated ECG images from various patients, ensuring high quality and accuracy [16]. The research [8] also ensures data is free of any personal information regarding the patient and only has ECG images



with the CVD class it is associated with. Table I, shows the total number of Images used for cardiac disorder detection used for each class. This dataset comprises ECG images of individuals with diverse cardiovascular conditions, categorized into four classes:

- **Class 1:** (Myocardial Infarction): Sudden deprivation of oxygen supply caused by the partial or complete halt of blood flow to a portion of the myocardium. It is commonly known as a heart attack. It is marked as 0 in the dataset.
- **Class 2:** (Abnormal Heartbeat): Irregular rhythm of the heartbeat (can be slow or fast) which can cause life-threatening diseases. It is marked as 1 in the dataset.
- **Class 3:** (Normal): Heartbeat in rhythm with no sign of abnormality. It usually ranges from 60 to 100 beats per minute. It is marked as 2 in the dataset
- **Class 4:** (History of MI): Encompassing individuals with a documented history of myocardial infarction marked as 3 in the dataset.

TABLE I  
DISTRIBUTION OF CLASSES IN THE ECG DATASET

Sr. No	Class	12-Lead ECG	Total Images
1.	Myocardial Infarction (MI)	240	2880
2.	Abnormal Heartbeat	233	2796
3.	Normal	284	3408
4.	Previous History Of MI	172	2064

### B. Image Preprocessing Techniques

- **Grayscale Conversion:** Colored ECG images are transformed into grayscale using "skimage.color.rgb2gray" to enhance structural details crucial for subsequent processing as shown in Figure 2.
- **Image Validation and Similarity Check:** If the uploaded image is sufficiently similar to the reference images (with a structural similarity score above 0.70), the app will proceed with additional processing. Otherwise, it will inform the user that the image format is not supported.

### C. Lead Extraction

- **Dividing ECG into Leads:** The ECG image is divided into 13 distinct leads, each representing different aspects of heart activity, enabling comprehensive cardiac information capture.
- **Visualization and Categorization:** Individual ECG leads are visualized separately, illustrating signal patterns, amplitude variations, and temporal changes. Leads are categorized into limb leads (e.g., 1, 2, 3) and chest leads (e.g., V1-V6).

### D. Preprocessing Techniques on Leads

- **Smoothing and Noise Reduction:** Gaussian filtering (`skimage.filters.gaussian`) is applied to leads images to enhance signal clarity and simplify subsequent analysis. It is also used to reduce noise and preserve important features of images.

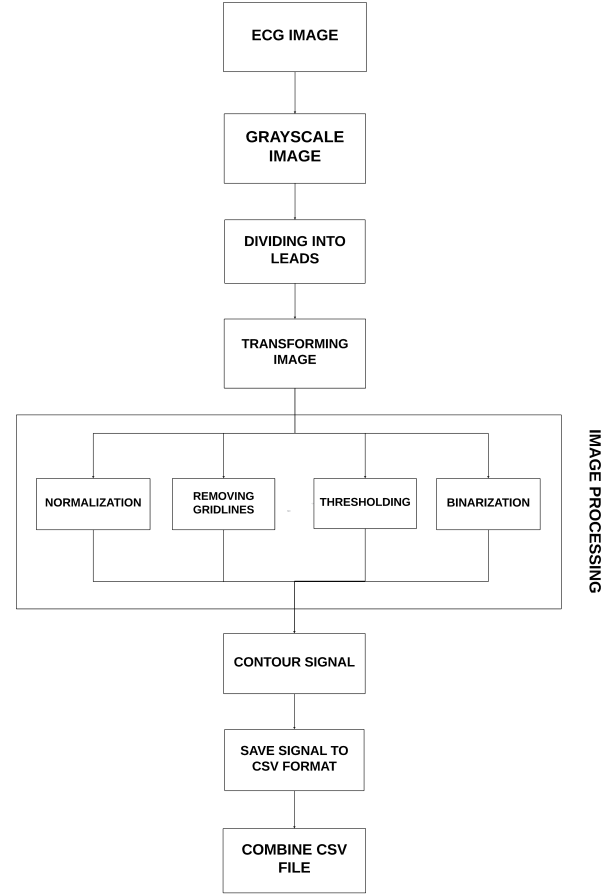


Fig. 2. Data Processing

- **Thresholding for Clarity:** Techniques such as Otsu's method (`skimage.filters.threshold_otsu`) are employed to distinguish foreground and background within leads, aiding differentiation of key features.

### E. Signal Extraction and Processing

- **Signal Identification:** Precise signal contours and regions containing signals are extracted using techniques like (`skimage.measure.find_contours`).
- **Normalization and Scaling:** Extracted signals undergo Min-Max normalization using `sklearn.preprocessing.MinMaxScaler` to ensure uniformity in signal amplitude.
- **Conversion to 1D Signals:** Signal contours are transformed into one-dimensional representations suitable for training classification algorithms, and CSV files are generated for subsequent analysis.
- **Combining Signals:** CSV files containing 1D signals from different leads are merged into a unified dataset for streamlined processing.

### F. Dimensionality Reduction

- **Principal Component Analysis (PCA):** To make more precise dataset the method `sklearn.decomposition.PCA` is used which

helps in retaining only critical components of the dataset and also helps in simplifying subsequent modeling.

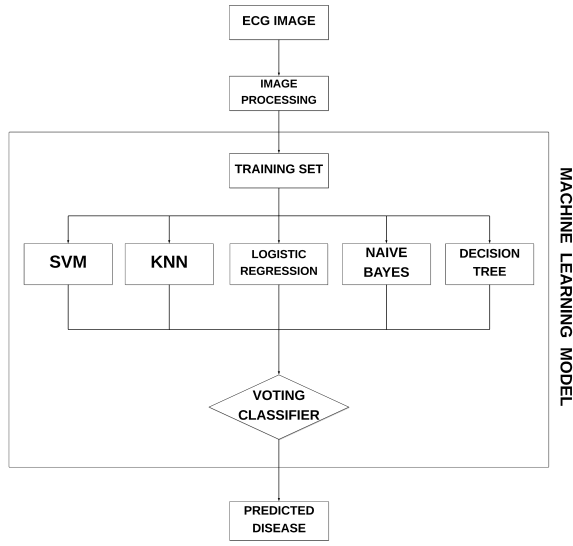


Fig. 3. Process Flow.

#### G. Model Selection

In this research, we employ a robust ensemble technique to enhance the accuracy of cardiovascular disorder classification from ECG images. The following machine-learning models are incorporated into the ensemble:

- **Support Vector Machine (SVM):** An adaptable algorithm primarily used for classification problems in machine learning. It segregates different classes by creating decision boundaries so that new points can be put in the correct category.
- **K-Nearest Neighbors (KNN):** A robust method that stores the dataset in the training phase and at the time of classification puts the new cases in the most suitable category.
- **Random Forest (RF):** A powerful ensemble learning method consists of several decision trees that help in improving the predictive accuracy of the dataset.
- **Gaussian Naive Bayes (Bayes):** A generative learning algorithm based on the Bayes theorem, particularly effective for its simplicity and efficiency.
- **Logistic Regression:** An excellent and strong method that examines the relationship between independent and dependent binary variables in binary classification tasks.
- **XGBoost (Extreme Gradient Boosting):** A boosting algorithm that helps in managing intricate interactions and enhancing performance via regularization and ensemble learning.

#### H. Ensemble Technique: Voting Classifier

We use a soft voting classifier as illustrated in Figure 3, which takes into consideration the probability of each

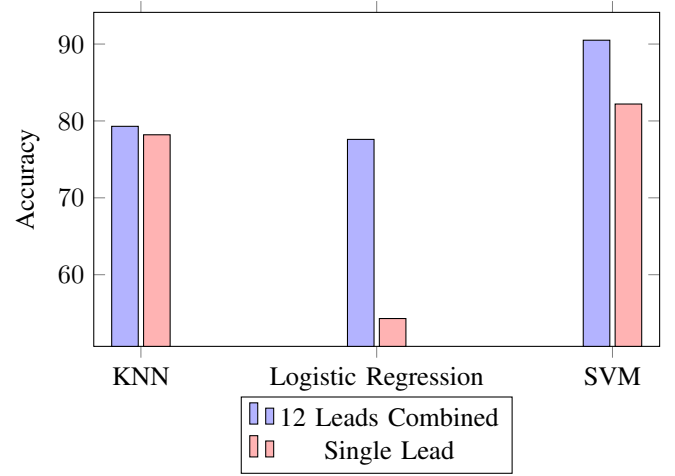


Fig. 4. Performance comparison between 12-lead & Single lead ECG Image

TABLE II  
ALGORITHM WITH SINGLE AND 12-LEAD ECG IMAGE

Algorithm	12 Leads Combined (%)	Single Lead (%)
KNN	79.3	78.2
Logistic Regression	77.6	54.3
SVM	90.5	82.2

class and combines the predictions of individual models. This ensemble method makes use of the advantages of many algorithms to minimize the shortcomings of individual models. GridSearchCV is used for hyperparameter optimization, which maximizes the model's performance. The ensemble technique's efficacy is demonstrated by the 92.47% attained accuracy, which highlights its capacity to use the combined strength of numerous models for the precise identification of cardiovascular disorders.

#### I. Evaluation

For evaluation, we utilized an extracted dataset as our test data, containing information about the classification of disorders marked as follows: 0 for Myocardial Infarction (MI), 1 for Abnormal Heartbeat, 2 for Normal, and 3 for History with MI. The model was tested to determine its ability to predict the correct classification of disorders in ECG images. The evaluation involves comparing the model's predictions with the ground truth labels in the dataset. The model's performance is determined by the error in its predictions. A lower error indicates a higher probability of the model accurately detecting the correct disorder classification. The accuracy of the model is calculated as the probability of correctly predicting the disorder classification. Refer to Figure 4 and Figure 5 for visual representations related to the evaluation process.

#### RESULTS AND ANALYSIS

Figure 4 and Table II show the performance of the model with combined 12-lead data yields superior accuracy compared to using single lead data alone, with these results in mind we

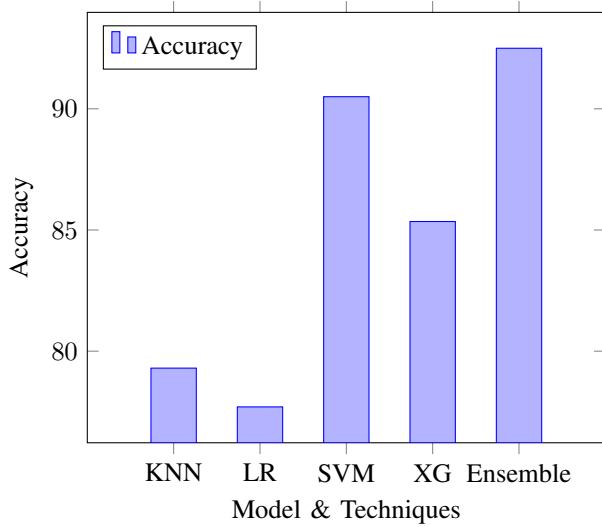


Fig. 5. Performance of Models & Techniques

TABLE III  
ACCURACY OF VARIOUS ALGORITHMS

Sr. No.	ALGORITHMS	ACCURACY
1.	K-Nearest Neighbor	79.3
2.	Logistic Regression	77.6
3.	Support Vector Machine	90.5
4.	XG Boost	85.3
5.	Ensemble	92.4

only use combined 12-lead data in our research. This significantly enhances the performance of the test models. Under Voting-based Ensemble Classification with GridSearchCV for tuning the hyperparameters which achieves an impressive accuracy of 92.5%, three Machine learning models KNN, SVM, and Random Forest Classifier are put together and voted to select one model that gives the highest accuracy. This represents a substantial improvement over individual models, including SVM (90.5%), KNN (79.3%), Logistic Regression (77.7%), and XGBoost (XG) (85.3%), as illustrated in Figure 5 and Table III. The ensemble approach demonstrates how they balance out their weaknesses and function together to enhance the detection of cardiovascular disorders in ECG data. This approach can make the detection more reliable and accurate.

#### COMPARATIVE ANALYSIS

In the research paper by Swati Shilaskar [17] the algorithm used (i.e. SVM) gives less accuracy than our study because of the data pre-processing limitations. The dataset is not well pre-processed and because of that the accuracy is less in the research paper. In our study, the ECG dataset is well pre-processed and many algorithms are used to calculate higher accuracy. And another study by Kaushalya Dissanayake [18] different dataset is taken for processing which contains 75 features of which 14 features are used. These are related to biographic, clinical, and habitual features like blood pressure,

Age, sex, etc which are difficult to store and handle which decreases its overall accuracy and precision. We have used a more refined ECG dataset which when processed is very easy to handle and after applying different algorithms gives astounding accuracy.

#### CONCLUSION

To conclude this study, we found the consistently robust performance from SVM across both individuals and the combined 12-leads dataset. The information from the combination of 12 leads gives improved models' prediction capabilities, with SVM and XGBoost achieving commendable accuracy with pre-processed ECG data. The integration of ensemble models by Voting Classifier enhances overall model performance by balancing out the individual weaknesses and the strengths of individual classifiers. More refined Datasets can help in the development in this field where the Deep learning model can contribute to improving accuracy with the range of datasets to feed. While taking account of Ethical considerations in handling ECG image datasets including preserving data privacy and acknowledging biases, emphasizing the need for caution in interpreting and deploying models for cardiovascular disease detection. The validation of these models in a real-time pipeline underlines their practical efficacy in the context of cardiovascular disorder identification through ECG image analysis, offering promising prospects for application in clinical scenarios.

#### REFERENCES

- [1] World Health Organization, "Cardiovascular Diseases", from [https://www.who.int/health-topics/cardiovascular-diseases/#tab=tab\\_1](https://www.who.int/health-topics/cardiovascular-diseases/#tab=tab_1)
- [2] Mike Cadogan, "ECG Lead Positioning", 2022. <https://litfl.com/ecg-lead-positioning/>
- [3] Figure 1. 12-Lead ECG electrode points. Adapted from "ECG Signal Quality Assessments of a Small Bipolar Single-Lead Wearable Patch Sensor" by Rajbhandary, P.L., Nallathambi, G., Selvaraj, N. et al. Cardiovasc Eng Tech 13, 783–796 (2022). <https://link.springer.com/article/10.1007/s13239-022-00617-3#Fig2>
- [4] Adyasha Rath, Debahuti Mishra, Ganapati Panda, Suresh Chandra Satapathy, "Heart disease detection using deep learning methods from imbalanced ECG samples", Biomedical Signal Processing and Control, Volume 68, 2021, 102820, ISSN 1746-8094, <https://doi.org/10.1016/j.bspc.2021.102820>
- [5] Rohit Bharti, Aditya Khamparia, Mohammad Shabaz, Gaurav Dhiman, Sagar Pande, Parneet Singh, "Prediction of Heart Disease Using a Combination of Machine Learning and Deep Learning", Computational Intelligence and Neuroscience, vol. 2021, Article ID 8387680, 11 pages, 2021. <https://doi.org/10.1155/2021/8387680>
- [6] M.M. Al Rahhal, Yakoub Bazi, Haikel AlHichri, Naif Alajlan, Farid Melgani, R.R. Yager, "Deep learning approach for active classification of electrocardiogram signals", Information Sciences, Vol-

ume 345, 2016, Pages 340-354, ISSN 0020-0255, <https://doi.org/10.1016/j.ins.2016.01.082>

- [7] Guidi, Gabriele, Roberto Miniati, Matteo Mazzola, and Ernesto Iadanza. 2016. "Case Study: IBM Watson Analytics Cloud Platform as Analytics-as-a-Service System for Heart Failure Early Detection" *Future Internet* 8, no. 3: 32. <https://doi.org/10.3390/fi8030032>
- [8] Ali Haider Khan, Muzammil Hussain, Muhammad Kamran Malik, "ECG Images dataset of Cardiac and COVID-19 Patients", *Data in Brief*, Volume 34, 2021, 106762, ISSN 2352-3409, <https://doi.org/10.1016/j.dib.2021.106762>
- [9] Keogh, E., Mueen, A. (2017). Curse of Dimensionality. In: Sammut, C., Webb, "G.I. (eds) *Encyclopedia of Machine Learning and Data Mining*". Springer, Boston, MA. [https://doi.org/10.1007/978-1-4899-7687-1\\_192](https://doi.org/10.1007/978-1-4899-7687-1_192)
- [10] Liu, Huan, and Hiroshi Motoda, eds. "Feature extraction, construction and selection: A data mining perspective". Vol. 453. Springer Science & Business Media, 1998.
- [11] Rohit Bharti, Aditya Khamparia, Mohammad Shabaz, Gaurav Dhiman, Sagar Pande, Parneet Singh, "Prediction of Heart Disease Using a Combination of Machine Learning and Deep Learning", *Computational Intelligence and Neuroscience*, vol. 2021, Article ID 8387680, 11 pages, 2021. <https://doi.org/10.1155/2021/8387680>
- [12] Rajbhandary, P.L., Nallathambi, G., Selvaraj, N. et al. "ECG Signal Quality Assessments of a Small Bipolar Single-Lead Wearable Patch Sensor". *Cardiovasc Eng Tech* 13, 783–796 (2022). <https://doi.org/10.1007/s13239-022-00617-3>
- [13] P. Li et al., "High-Performance Personalized Heart-beat Classification Model for Long-Term ECG Signal," in *IEEE Transactions on Biomedical Engineering*, vol. 64, no. 1, pp. 78-86, Jan. 2017, doi: 10.1109/TBME.2016.2539421
- [14] Ketu, S., Mishra, P.K. "Empirical Analysis of Machine Learning Algorithms on Imbalance Electrocardiogram Based Arrhythmia Dataset for Heart Disease Detection". *Arab J Sci Eng* 47, 1447–1469 (2022). <https://doi.org/10.1007/s13369-021-05972-2>
- [15] V. Jahmunah, E.Y.K. Ng, Tan Ru San, U. Rajendra Acharya, "Automated detection of coronary artery disease, myocardial infarction and congestive heart failure using GaborCNN model with ECG signals", *Computers in Biology and Medicine*, Volume 134, 2021, 104457, ISSN 0010-4825, <https://doi.org/10.1016/j.combiomed.2021.104457>
- [16] Khan, Ali Haider; Hussain, Muzammil (2021), "ECG Images dataset of Cardiac Patients ", *Mendeley Data*, V2, doi: 10.17632/gwbz3fsgp8.2
- [17] Swati Shilaskar, Ashok Ghatol, "Feature selection for medical diagnosis : Evaluation for cardiovascular diseases, *Expert Systems with Applications*", Volume 40, Issue 10, 2013, Pages 4146-4153, ISSN 0957-4174, <https://doi.org/10.1016/j.eswa.2013.01.032>
- [18] Kaushalya Dissanayake, Md Gapar Md Johar, "Com-

parative Study on Heart Disease Prediction Using Feature Selection Techniques on Classification Algorithms", *Applied Computational Intelligence and Soft Computing*, vol. 2021, Article ID 5581806, 17 pages, 2021. <https://doi.org/10.1155/2021/5581806>