**ECG Signal Analysis and Classification**

**A Project Report**

**Project by**

**Anirudh S Varrier CB.SC.U4AIE23113**

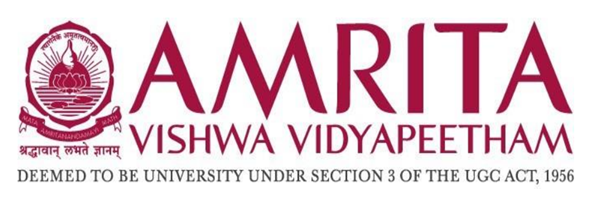
**B V Dharsini Sri CB.SC.U4AIE23115**

**Chanjhana Elango CB.SC.U4AIE23120**

**Kaniska M CB.SC.U4AIE23136**

*As a part of the subject*

**22PHY106- COMPUTATIONAL PHYSICS**

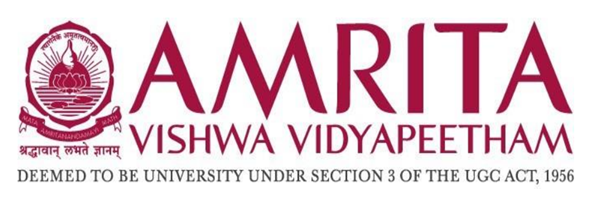


**Department of Artificial Intelligence**

AMRITA VISHWA VIDYAPEETHAM

COIMBATORE - 641 112 (INDIA)

**December 2023**



**AMRITA SCHOOL OF ENGINEERING**

**AMRITA VISHWA VIDYAPEETHAM**

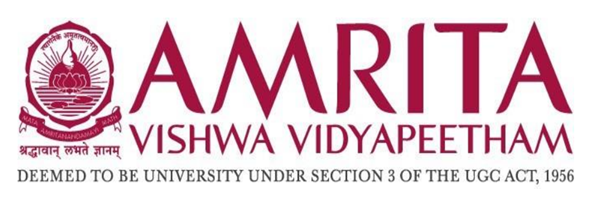
**COIMBATORE - 641 112**

**CERTIFICATE**

Amrita Vishwa Vidyapeetham Coimbatore CERTIFICATE This is to certify that the project entitled, “Drone Technology ” submitted by Abhishek Sankaramani , Vishal S , Meenakshi Sareesh , Yashwanth PT in fulfilments for the End Sem project of Computational Physics 1 st Semester, under Bachelor of Technology Degree in Computer Science Engineering and Artificial Intelligence at Amrita Vishwa Vidyapeetham, Coimbatore (Deemed University) is an authentic work carried out by him under our supervision and guidance. To the best of our knowledge, the matter embodied in the report has not been submitted to any other University / Institute for the award of any Degree or Diploma

Date: 22/12/2023 Jithin Velayudhan

Dept. of Artificial Engineering, Amrita Vishwa Vidyapeetham Coimbatore -641112, Tamil Nadu



**ACKNOWLEDGMENT**

We would like to extend our heartfelt gratitude to **Mr. Jithin Velayudhan (**Assistant Prof, Amrita Vishwa Vidhyapeedam) for his invaluable guidance and support throughout the development of our Computational Physics Project: **ECG Signal Analysis and Classification**. His insights and mentorship have been instrumental in shaping this project and enhancing my understanding of the subject. We would also like to thank Prof. K . P Soman (Head of AI, Amrita Vishwa Vidhyapeedam)

**TABLE OF CONTENTS**

**TITLE Page No**

1.Abstract 05

2.Introduction 06

3.Aim 09

4.Literature Review 10

5.MIT-BIH Arrhythmia Dataset 11

6.Methodology 13

7.Implementation, Code and Output 16

8.Limitations and conclusion 26

9.References 27

**1.0 ABSTRACT**

The project “ECG Signal Analysis and Classification” focuses on the analysis and classification of Electrocardiogram (ECG) signals using machine learning techniques. ECG signals, which represent the electrical activity of the heart, are crucial for diagnosing various heart diseases. The project employs the MIT-BIH Arrhythmia Database, a publicly available dataset containing ECG recordings exhibiting a variety of heart conditions.

The project used multiple machine learning algorithms such as SVM and Random Forest Classifiers to train and understand the ECG data which has been preprocessed. It will also apply algorithms to create synthetic data based on the original data to get a balanced dataset.

**2.0 INTRODUCTION**

The human body, a marvel of nature, is a complex and intricate system composed of billions of cells working in harmony from the moment of birth until the end of life. It is a symphony of multiple organ systems, each performing vital processes that are essential for our survival.

The Respiratory System, for instance, is responsible for providing us with the oxygen necessary for our cells to function and produce energy. It also removes carbon dioxide, a waste product that can be harmful if it accumulates in our blood. This system, comprising the lungs and associated structures, is a testament to the body’s efficiency and adaptability.

The Central Nervous System, often referred to as the body’s control center, regulates all the processes and actions our body performs. It ensures that each system works in harmony and efficiency, coordinating everything from our heartbeat to our ability to read and comprehend this text.

The Digestive System plays a crucial role in breaking down the food we consume into simpler substances. These substances are then absorbed by the body, providing us with the necessary nutrients we need to stay energetic and perform our day-to-day tasks. This system, which includes the stomach, intestines, and associated organs, is a marvel of biological engineering.

Our Circulatory System, which includes the heart and blood vessels, is responsible for transporting oxygen-rich blood to every cell in our body. The blood flows through arteries, veins, and capillaries, reaching all parts of our body. The heart, a muscular organ, pumps the blood with extreme pressure, ensuring its delivery to even the most remote cells. It does this by sending electrical impulses to each of its muscles, causing them to contract and relax in a rhythmic pattern. This process, which continues throughout our lifetime without stopping, is vital for our survival.

Maintaining a healthy heart is therefore of utmost importance. This can be achieved by maintaining physical fitness and consuming a diet low in fat and cholesterol. Obesity, diabetes, high blood pressure, and stress are all factors that can lead to heart problems. Additionally, smoking and excessive alcohol consumption can also contribute to heart disease. A family history of heart disease can also increase one’s risk.

High cholesterol or fat levels can lead to the buildup of these substances in the arteries, which carry blood from the heart to other parts of the body. This can cause a blockage in the artery, leading to a heart attack.

Certain conditions, such as genetics or external factors, can cause our heart to beat irregularly, a condition known as arrhythmia. This can affect either the ventricles or the atria of the heart and can lead to serious complications.

Prevention or early detection of these diseases is of paramount importance. Heart diseases can often be recognized by physical symptoms such as chest pain or pressure, dizziness, and shortness of breath. In some cases, these symptoms may be followed by fainting. However, for some people, these symptoms may not be very pronounced, making it harder to detect heart diseases.

Medical tests such as electrocardiography (ECG) and angiograms can also be used to detect heart diseases. ECG is a test that records the electrical activity of the heart over a period of time. The recorded information is represented as a graph or series of waves on a monitor or paper. ECG is a valuable tool for diagnosing various heart conditions and assessing the heart’s rhythm and function. It is crucial in diagnosing conditions like atrial fibrillation, heart attacks, and heart rhythm disorders. Continuous or ambulatory ECG monitoring is also used to capture intermittent heart rhythm abnormalities that may not be evident during a standard ECG.

Electrocardiography (ECG) is a fundamental diagnostic procedure in the field of cardiology, providing valuable insights into the functioning of the heart. The process involves placing small electrodes on specific parts of the body, including the legs, wrists, and chest. These electrodes serve as sensors that detect the electrical signals generated by the heart during each cardiac cycle.

The heart’s electrical activity begins with the sinoatrial (SA) node, often referred to as the heart’s natural pacemaker. Located in the right atrium, the SA node generates an electrical signal that triggers the contraction of the heart muscles. This signal is first passed to the atria, causing them to contract and pump blood into the ventricles.

The electrical signal then travels to the atrioventricular (AV) node, which acts as a kind of electrical bridge allowing the signal to pass from the atria to the ventricles. Once the signal reaches the ventricles, it causes them to contract, pumping blood to the lungs and the rest of the body.

The electrodes placed on the skin detect these electrical signals as they travel through the heart. The ECG machine then records these signals, converting them into a visual format that can be analyzed by medical professionals.

The resulting graph, known as an electrocardiogram, consists of a series of waves and intervals that represent different phases of the cardiac cycle. The P wave corresponds to the contraction of the atria, the QRS complex reflects the contraction of the ventricles, and the T wave represents the relaxation of the ventricles.

Changes or abnormalities in these ECG waves or segments can indicate potential heart problems. For instance, alterations in the shape, duration, or timing of the P wave, QRS complex, or T wave can suggest conditions such as atrial fibrillation, ventricular tachycardia, or myocardial infarction, commonly known as a heart attack.

Specifically, Coronary Artery Disease (CAD), a condition characterized by the narrowing or blockage of the coronary arteries, can cause specific changes in the ST segment or T wave on an ECG. Similarly, arrhythmias, which are irregularities in the heart’s rhythm, can cause a wide range of ECG changes, depending on the type and severity of the arrhythmia.

In conclusion, ECG is a crucial tool in modern medicine, allowing for the non-invasive assessment of the heart’s electrical activity. By interpreting the patterns and anomalies in an electrocardiogram, healthcare professionals can diagnose and manage a wide range of heart conditions, potentially saving lives and improving patient outcomes.

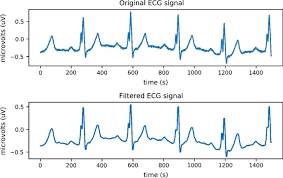
.

FIG 1: Examples of ECG signals (a) Original Signal (b) Filtered Signal

**3.0 Aim**

The detection of anomalies in Electrocardiogram (ECG) data is of paramount importance in the identification of various types of arrhythmias. Arrhythmias are conditions where the heartbeat is irregular, too slow, or too fast. They are often the result of disruptions in the electrical signaling that regulates the steady rhythm of the heart. The primary objective of this project is to develop a machine learning model that can classify ECG signals with a high degree of accuracy.

ECG signals, which are graphical representations of the electrical activity of the heart, can exhibit significant variations based on an individual’s heart condition. These signals are typically captured using electrodes placed on the skin, which detect the minute electrical changes on the skin that arise from the heart muscle’s electrophysiologic pattern of depolarizing and repolarizing during each heartbeat. The accurate classification of these signals is a critical step in diagnosing a wide range of heart diseases, including arrhythmias, coronary artery disease, and other conditions that affect the heart’s structure and function.

Moreover, the machine learning model developed in this project aims to predict potential anomalies in the ECG signals. These anomalies could be indicative of abnormal heart conditions, including but not limited to, myocardial infarction (heart attack), hypertrophy (enlargement of the heart), and various forms of heart disease. The ability to predict these anomalies accurately and in a timely manner could enable early detection and intervention, potentially saving lives and improving the quality of life for patients worldwide. Therefore, the development of such a model could have far-reaching implications in the field of healthcare, particularly in cardiology.

Cardiovascular diseases are currently the leading cause of death globally. The World Health Organization estimates that these diseases take the lives of 17.9 million people (about the population of New York) every year, accounting for 31% of all global deaths. In this context, the importance of a method or machine capable of analyzing ECG data in real-time and predicting the likelihood of a person having a cardiovascular issue cannot be overstated. This capability can contribute significantly to saving many lives by enabling prompt and appropriate medical intervention.

Doctors, despite their expertise and experience, can make mistakes when interpreting ECG data due to the complex nature of these signals and the subtle variations that can indicate different heart conditions. This underscores the need for a highly accurate and efficient machine that can analyze ECG signals with a higher degree of accuracy than is possible through manual interpretation. Such a machine could also play a crucial role in detecting cardiovascular diseases at an extremely early stage, thereby enabling early treatment and potentially preventing the progression of these diseases.

With a model such as the one being developed in this project, doctors and patients all over the world stand to benefit in numerous ways. For instance, quicker and more accurate prediction of signal types (ECG variations) can aid doctors in diagnosing heart conditions more efficiently. This, in turn, can expedite the process of devising an appropriate treatment plan for the patient, thereby improving the chances of a successful recovery. In conclusion, the development and implementation of machine learning models for ECG signal analysis and interpretation can revolutionize the field of cardiovascular healthcare, paving the way for improved patient outcomes and a healthier future for all.

**4.0 Literature Review**

The research papers based on ECG classification gave us a clear idea of how the project must be done. Since the ECG data is not simple text-based data, the pre-processing is complex. The papers we read helped us understand the methodology we must follow to process the given data. The papers also gave lots of important information on which machine learning algorithms can be used.

The research papers based on the MIT-BIH Arrhythmia Dataset gave a lot of information on the above-mentioned dataset. The papers gave information on the ECG data ie. From whom the data was taken, the length of the data etc. They also explained about the different abnormal classes and gave the list containing the names and symbols of the abnormalities.

**5.0 MIT-BIH Arrhythmia Data Set**

The MIT-BIH Arrhythmia Database, accessible on PhysioNet, is a well-known and widely used resource in biomedical research, particularly in the development and evaluation of automated systems for arrhythmia detection and heart rate variability. The database contains 48 half-hour two-channel ambulatory ECG recordings, a total of 24 hours of recordings. These were selected from a larger dataset comprising nearly 4,000 continuous ambulatory ECG recordings, collected from 47 subjects at Boston’s Beth Israel Hospital (now Beth Israel Deaconess Medical Center) and MIT between 1975 and 1979.

The subjects included in the database were 25 men aged between 32 and 89 years and 22 women aged between 23 and 89 years. They were a mix of inpatients and outpatients at the hospital. The recordings were chosen to include a wide variety of different types of arrhythmias and a range of common and rare morphologies.

The ECG signals were recorded using a Del Mar Avionics (now Spacelabs Healthcare) 445L series ECG recorder. The recordings were digitized at 360 samples per second per channel with 11-bit resolution over a 10-mV range. Two or more cardiologists independently annotated each record; disagreements were resolved to obtain the computer-readable reference annotations for each beat (approximately 110,000 annotations in all) included with the database.

The annotations include labels for both rhythm and beat annotations. Rhythm annotations mark changes in the rhythm and include rhythms like normal sinus rhythm, atrial fibrillation, atrial flutter, ventricular tachycardia, and others. Beat annotations label individual beats and include beat types like normal beat, left bundle branch block beat, right bundle branch block beat, premature ventricular contraction, fusion of ventricular and normal beat, atrial premature beat, and others.

This database has been used by researchers worldwide to develop and evaluate algorithms for arrhythmia and sudden cardiac death detection. It’s a benchmark for assessing the performance of these algorithms, contributing significantly to advancements in this field. The wide variety of ECG signals and the detailed annotations make the MIT-BIH Arrhythmia Database an invaluable resource for researchers in biomedical engineering and healthcare technology.

In conclusion, the MIT-BIH Arrhythmia Database is a comprehensive and meticulously annotated resource that continues to support the development and evaluation of automated ECG analysis systems, thereby contributing to improvements in cardiac care. The database’s enduring relevance, more than four decades after its creation, attests to the foresight of its creators and the ongoing need for publicly available, rigorously curated biomedical data. As researchers continue to develop and refine algorithms for arrhythmia detection, the MIT-BIH Arrhythmia Database will undoubtedly continue to play a crucial role in this important area of biomedical research.

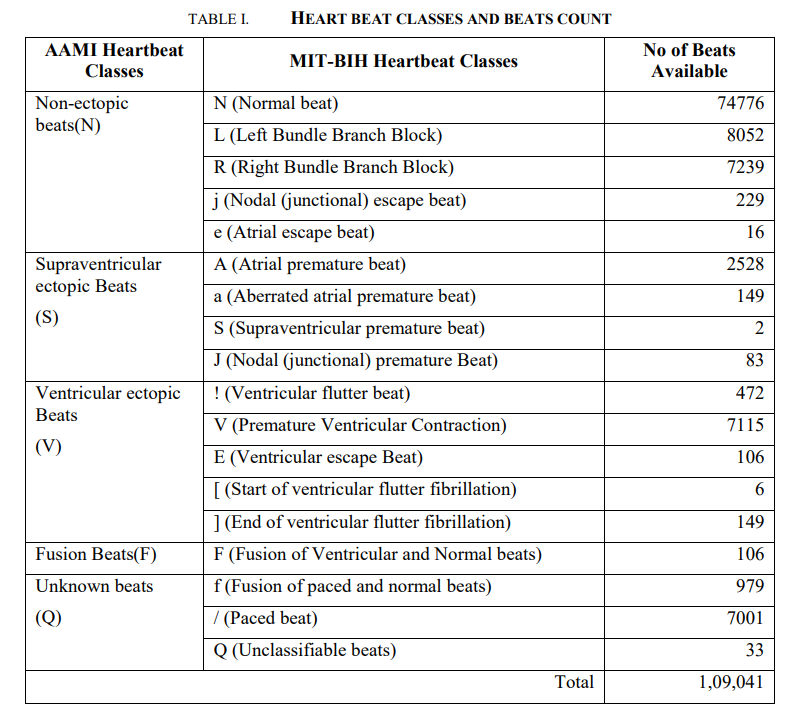


FIG 2: Figure shows the Classes and Beats there in the MIT-BIH Arrhythmia Dataset

**6.0 Methodology**

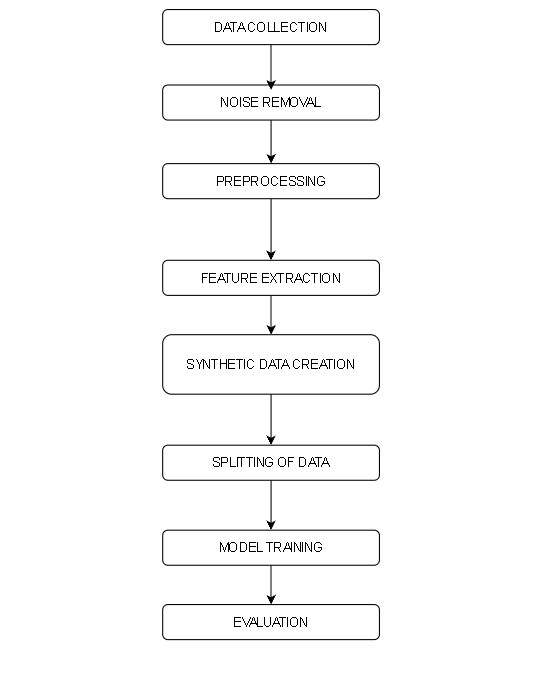


FIG 3: Figure represents a flow chart representing the methodology used in the project

The methodology used in the project involves several steps, including data acquisition, preprocessing, feature extraction, model training, and evaluation.

**6.1 Data Acquisition:**  
The initial phase of the project involves data acquisition. The dataset utilized in this project is the MIT-BIH Arrhythmia Database, which is publicly accessible on PhysioNet. This database comprises 48 half-hour excerpts of two-channel ambulatory ECG recordings, gathered from 47 subjects at Boston’s Beth Israel Hospital between 1975 and 1979. The ECG signals in the database display a variety of heart conditions, making it an invaluable resource for developing and testing automated arrhythmia detectors. The data acquisition process is crucial as the quality and quantity of data directly impact the subsequent steps of preprocessing, feature extraction, and model training.

**6.2 Noise Removal:**  
The Variational Mode Decomposition (VMD) technique is employed to eliminate any noise present in the data. Noise can significantly affect the readings and lead to inaccurate results. Therefore, it’s essential to apply noise removal techniques to ensure the data is clean and ready for further processing.

**6.3 Preprocessing:**  
The raw ECG data undergoes preprocessing to prepare it for further analysis. The data is processed through a VDM (Variational mode decomposition) algorithm which separates the original signal into multiple components, each representing a different frequency mode. Each Intrinsic Mode Function (IMF) generated by the VMD represents a distinct mode in the signal and together can reconstruct the signal. The data is then normalized, and the Signal to Noise Ratio (SNR) value is computed.

**6.4 Feature Extraction:**  
After preprocessing, features are extracted from the ECG signals using Variational Mode Decomposition (VMD). VMD is a signal processing technique that decomposes a signal into a set of intrinsic mode functions (IMFs). In this project, VMD is applied to each beat of the ECG signal, and statistical features are extracted from the resulting IMFs. The features include the mean, standard deviation, skewness, and kurtosis of each IMF segment.

**6.5 Synthetic Data Creation:**  
The features extracted from the group when corresponded to the labels will result in a severe data imbalance. To address this, Synthetic Minority Over-sampling Technique (SMOTE) algorithms are used to create more data with respect to labels and the features. The Synthetic sample is created where each label has an equal number of features.

**6.6 Splitting of Data:**  
The new synthetic data is now split into the testing and training set where the training set has 80% of the values of the synthetic set where the random state value is 42. This ensures that the model has a good amount of data to learn from, while also having a separate set of data to validate its performance.

**6.7 Model Training:**  
Once the features are extracted, they are used to train machine learning models. In this project, three different models are trained: a HistGradientBoostingClassifier, a Support Vector Machine (SVM), DecisionTreeClassifier and a RandomForestClassifier. These models are trained separately with the same training data. It is also tested with the testing data and predicted value is gotten from each model.

**6.8 Evaluation:**  
Each of the predicted values is now compared with the labels of the testing data. It is then run through a code to check the accuracy, F1 scores, Recall scores and Precision. A Confusion matrix is then plotted for the data. This step is crucial in understanding the performance of the models and identifying areas of improvement.

Four different Machine learning models were used to compare the effectiveness of each model with the ECG data. This is necessary to implement the code if used in hospitals or clinics.

**7.0 CODE and IMPLEMENTATION**

The following code is the base of the project. The entire code is made using python with many modules.

**# Importing the modules needed for the project**

import numpy as np

import matplotlib.pyplot as plt

import wfdb

from scipy import stats

from vmdpy import VMD

from sklearn import svm

from sklearn.model\_selection import train\_test\_split

from sklearn.preprocessing import StandardScaler

from sklearn.metrics import classification\_report

from scipy.stats import skew, kurtosis

from sklearn.experimental import enable\_hist\_gradient\_boosting

from sklearn.ensemble import HistGradientBoostingClassifier

from sklearn.ensemble import RandomForestClassifier

from sklearn.ensemble import GradientBoostingClassifier

from sklearn.metrics import precision\_score, recall\_score, f1\_score

from imblearn.over\_sampling import SMOTE

from sklearn.metrics import confusion\_matrix

import seaborn as sns

from sklearn.tree import DecisionTreeClassifier

import matplotlib.pyplot as plt

**# Define the VMD parameters**

alpha = 2000 **# Bandwidth constraint**

tau = 0. **# Noise-tolerance (no strict fidelity enforcement)**

K = 3 **# 3 modes**

DC = 0 **# No DC part imposed**

init = 1  **# Initialize omegas uniformly**

tol = 1e-7 **# Tolerance**

L\_data=[] **# Add the file number into the list**

**# Initialize a list to store the results**

u\_chunks = []

ecg\_data=[]

for j in L\_data:

**# Load the MIT-BIH Arrhythmia dataset**

record = wfdb.rdrecord(f'/content/{j}')

**# Get the ECG data**

ecg\_data1 = record.p\_signal.flatten()

ecg\_data.append(ecg\_data1)

**# Define the size of each chunk**

chunk\_size = 10000 # Adjust this value based on your resources

**# Calculate the number of chunks**

num\_chunks = len(ecg\_data1) // chunk\_size

**# Loop over each chunk**

for i in range(num\_chunks):

# Get the start and end indices of the chunk

start = i \* chunk\_size

end = start + chunk\_size

**# Get the chunk of data**

chunk = ecg\_data1[start:end]

**# Apply VMD to the chunk**

u\_chunk, u\_hat\_chunk, omega\_chunk = VMD(chunk, alpha, tau, K, DC, init, tol)

**# Append the result to the list**

u\_chunks.append(u\_chunk)

print("Done VDM",j, "chunk",i+1)

**# Concatenate the chunks along the second axis**

u = np.concatenate(u\_chunks, axis=1)

ecg\_data=np.concatenate(ecg\_data)

def normalize(data):

mean = np.mean(data)

std = np.std(data)

return (data - mean) / std

def signal\_to\_noise\_ratio(signal, noise):

signal\_power = np.mean(np.square(signal))

noise\_power = np.mean(np.square(noise))

return 10 \* np.log10(signal\_power / noise\_power)

u=normalize(u)

**# Plot the original ECG data and the IMFs**

plt.figure(figsize=(12, 8))

plt.subplot(len(u) + 1, 1, 1)

plt.plot(ecg\_data)

plt.title('Original ECG data')

for i, IMF in enumerate(u, start=2):

plt.subplot(len(u) + 1, 1, i)

plt.plot(IMF)

plt.title(f'IMF {i - 1}')

plt.tight\_layout()

plt.show()

snrval = signal\_to\_noise\_ratio(u, ecg\_data - u)

print(f'Signal-to-Noise Ratio after VMD: {snrval}')

**# Initialize an empty list to store the features**

features = []

**# Initialize an empty list to store the labels**

labels = []

for j in L\_data:

**# Read the annotation**

annotation = wfdb.rdann(f'/content/{j}', 'atr')

**# Loop over each beat location**

for i in range(1, len(annotation.sample)):

if annotation.symbol[i] in ['N','L','R','j','e','A','S','J','a','!','V','E','[',']','F','f','/','Q']:

**# Get the start and end of the beat**

start = annotation.sample[i-1]

end = annotation.sample[i]

**# Initialize a list to store the features for this beat**

beat\_features = []

**# Loop over each IMF**

for imf in u:

# Segment the IMF at the beat location

segment = imf[start:end]

**# Calculate the mean, standard deviation, skewness, and kurtosis of the segment**

mean = np.mean(segment)

std = np.std(segment)

skewness = skew(segment)

kurt = kurtosis(segment)

**# Append the features to the beat\_features list**

beat\_features.extend([mean, std, skewness, kurt])

**# Append the beat\_features to the features list**

features.append(beat\_features)

**# Append the label to the labels list**

labels.append(annotation.symbol[i])

print("Feature extraction done for data set : ", j)

**# Convert the lists to NumPy arrays**

features = np.array(features)

labels = np.array(labels)

**# Scale the features**

scaler = StandardScaler()

val = scaler.fit\_transform(features)

smote = SMOTE(k\_neighbors=1)

**# Fit the SMOTE transformer to the data**

smote.fit(val, labels)

**# Resample the data**

val\_resampled, labels\_resampled = smote.fit\_resample(val, labels)

**# Split the data into training and test sets**

X\_train, X\_test, y\_train, y\_test = train\_test\_split(val\_resampled, labels\_resampled, test\_size=0.2, random\_state=42)

unique\_labels=[]

for i in labels\_resampled:

if i not in unique\_labels:

unique\_labels.append(i)

Predicted\_labels=[]

for i in unique\_labels:

j= "predicted " + i

Predicted\_labels.append(j)

**# Model-1 Histogram Gradient Boosting Classifier**

model1 = HistGradientBoostingClassifier()

model1.fit(X\_train, y\_train)

print("Model 1 trained")

pred1 = model1.predict(X\_test)

accuracy = np.mean(pred1 == y\_test)

print(f'Accuracy of model 1: {accuracy}')

**# Calculate precision, recall, and F1 score**

precision = precision\_score(y\_test, pred1, average='weighted')

recall = recall\_score(y\_test, pred1, average='weighted')

f1 = f1\_score(y\_test, pred1, average='weighted')

**# Print the scores**

print(f'Precision: {precision}')

print(f'Recall: {recall}')

print(f'F1 Score: {f1}')

**# Generate confusion matrix**

matrix = confusion\_matrix(y\_test, pred1)

**# Create a heatmap**

sns.heatmap(matrix, annot=True, fmt='d', cmap='Blues',

xticklabels=Predicted\_labels,

yticklabels=unique\_labels)

plt.title('Confusion Matrix')

plt.show()

**# Model-2 SVM**

model2 = svm.SVC()

model2.fit(X\_train,y\_train)

print("Model 2 trained")

pred2 = model2.predict(X\_test)

accuracy = np.mean(pred2 == y\_test)

print(f'Accuracy of model 2: {accuracy}')

**# Calculate precision, recall, and F1 score**

precision = precision\_score(y\_test, pred2, average='weighted')

recall = recall\_score(y\_test, pred2, average='weighted')

f1 = f1\_score(y\_test, pred2, average='weighted')

**# Print the scores**

print(f'Precision: {precision}')

print(f'Recall: {recall}')

print(f'F1 Score: {f1}')

**# Generate confusion matrix**

matrix = confusion\_matrix(y\_test, pred2)

**# Create a heatmap**

sns.heatmap(matrix, annot=True, fmt='d', cmap='Blues',

xticklabels=Predicted\_labels,

yticklabels=unique\_labels)

plt.title('Confusion Matrix')

plt.show()

**#Model-3 Random Forest Classifiers**

model3 = RandomForestClassifier()

model3.fit(X\_train,y\_train)

print("Model 3 trained")

pred3 = model3.predict(X\_test)

accuracy = np.mean(pred3 == y\_test)

print(f'Accuracy of model 3: {accuracy}')

**# Calculate precision, recall, and F1 score**

precision = precision\_score(y\_test, pred3, average='weighted')

recall = recall\_score(y\_test, pred3, average='weighted')

f1 = f1\_score(y\_test, pred3, average='weighted')

**# Print the scores**

print(f'Precision: {precision}')

print(f'Recall: {recall}')

print(f'F1 Score: {f1}')

**# Generate confusion matrix**

matrix = confusion\_matrix(y\_test, pred3)

**# Create a heatmap**

sns.heatmap(matrix, annot=True, fmt='d', cmap='Blues',

xticklabels=Predicted\_labels,

yticklabels=unique\_labels)

plt.title('Confusion Matrix')

plt.show()

**# Model-4 Decision Tree Classifier**

model4 = DecisionTreeClassifier()

model4.fit(X\_train, y\_train)

pred4 = model4.predict(X\_test)

accuracy = np.mean(pred4 == y\_test)

print(f'Accuracy of Decision Tree model: {accuracy}')

**# Calculate precision, recall, and F1 score**

precision = precision\_score(y\_test, pred4, average='weighted')

recall = recall\_score(y\_test, pred4, average='weighted')

f1 = f1\_score(y\_test, pred4, average='weighted')

**# Print the scores**

print(f'Precision: {precision}')

print(f'Recall: {recall}')

print(f'F1 Score: {f1}')

**# Generate confusion matrix**

matrix4 = confusion\_matrix(y\_test, pred4)

**# Create a heatmap**

sns.heatmap(matrix4, annot=True, fmt='d', cmap='Blues',

xticklabels=Predicted\_labels,

yticklabels=unique\_labels)

plt.title('Confusion Matrix for Decision Tree')

plt.show()

**7.1 CODE EXPLAINATION**

The code you’ve shared is for a machine learning task involving ECG data. It starts by defining parameters for Variational Mode Decomposition (VMD), a technique used to decompose a signal into intrinsic mode functions. The ECG data is loaded, divided into chunks, and VMD is applied to each chunk. The results are concatenated and normalized.

Next, features are extracted from the data. For each beat in the ECG data, statistical measures (mean, standard deviation, skewness, and kurtosis) are calculated for each IMF segment corresponding to that beat. These form the features for the machine learning models.

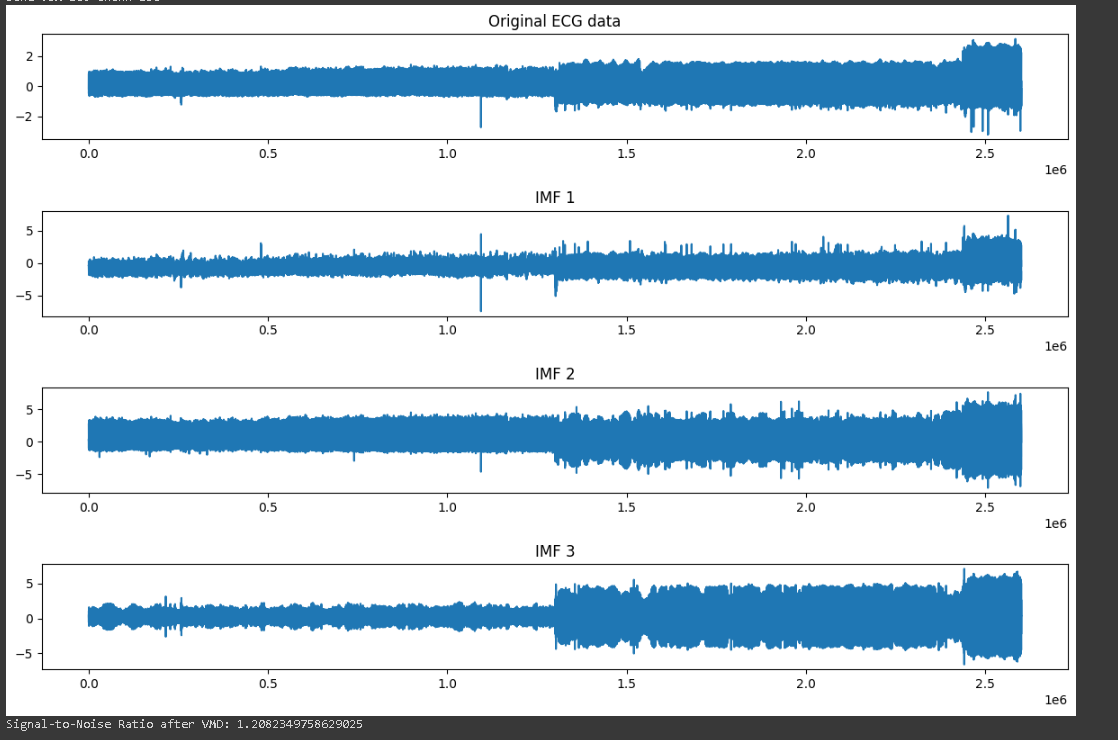
The features are then scaled using StandardScaler, and the Synthetic Minority Over-sampling Technique (SMOTE) is used to handle class imbalance in the data. The resampled data is split into training and test sets.

Four different machine learning models are trained on this data: Histogram Gradient Boosting Classifier, Support Vector Machine (SVM), Random Forest Classifier, and Decision Tree Classifier. For each model, predictions are made on the test set and performance metrics (accuracy, precision, recall, F1 score) are calculated. A confusion matrix is also generated for each model and displayed as a heatmap.

This code is a comprehensive example of applying machine learning to ECG data, from preprocessing and feature extraction to model training, prediction, and evaluation. It demonstrates the use of various techniques and models in tackling a complex classification problem.

**7.2 CODE OUPUT**

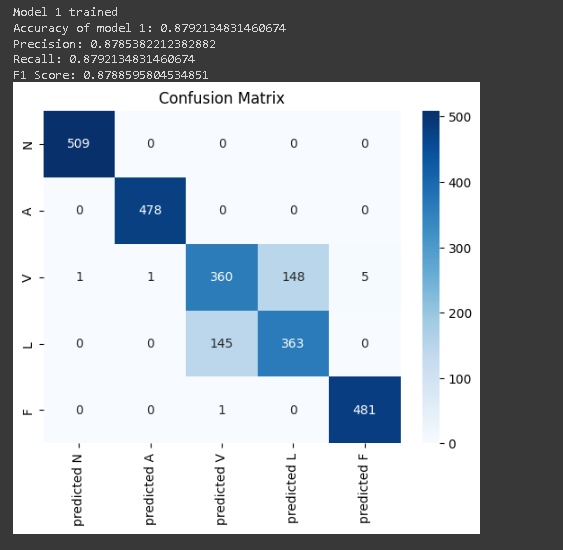
The above code is run with files 100 and 109.

Denoising and combining of both data files 

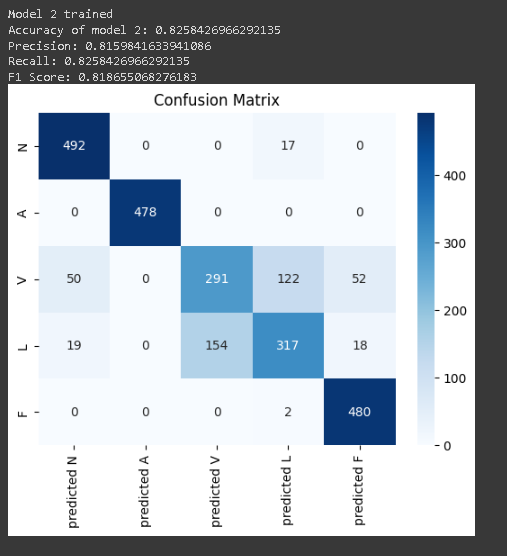
Feature extraction from both files



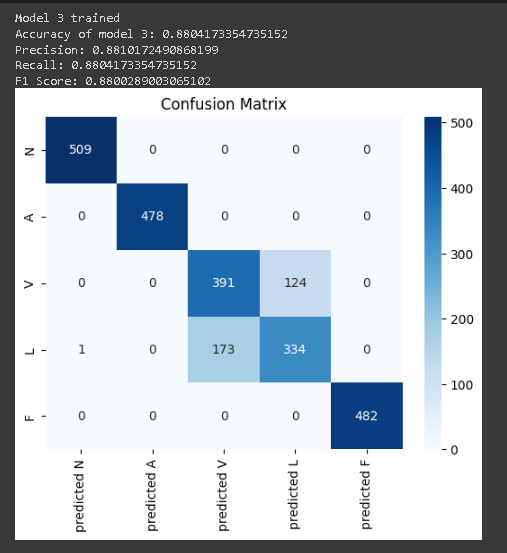
Model – 1



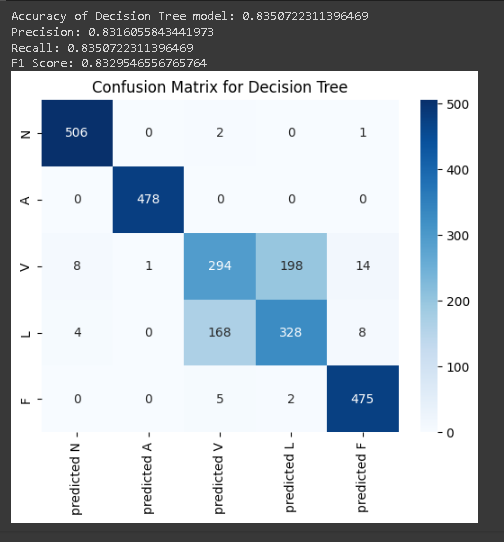
Model – 2



Model-3



Model-4



**8.0 LIMITATIONS**

The project has a lot of uses but it also has some limitations:

1. **Data Quality**:

The performance of the models heavily depends on the quality of the ECG data. Any noise or inaccuracies in the data can significantly affect the results.

1. **Model Selection**:

The code uses specific machine learning models. Other models might perform better on this task.

1. **Class Imbalance**:

The code uses SMOTE to handle class imbalance. However, this method can introduce noise by creating synthetic examples.

1. **Computational Resources**:

The code might require significant computational resources (memory and processing power) for large datasets.

1. **Real-Time Application**:

The code is not designed for real-time prediction, which would be necessary for a live patient monitoring system.

**8.1 CONCLUSION**

The project “ECG signal analysis and classification” successfully demonstrates the application of machine learning techniques to ECG data. The code effectively preprocesses the data, extracts meaningful features, and applies various machine learning models for classification. The use of Histogram Gradient Boosting, SVM, Random Forest, and Decision Tree classifiers provides a comprehensive comparison of different approaches. However, the code has limitations in terms of data quality, feature and model selection, handling of class imbalance, computational resources, real-time application, validation, interpretability, and generalizability. Despite these, the project provides a solid foundation for further exploration and improvement in the field of ECG signal analysis and classification.

**9.0 REFERENCES**

9.1 <https://www.researchgate.net/publication/328460579_Arrhythmia_Detection_Using_MIT-BIH_Dataset_A_Review>

9.2 <https://arxiv.org/pdf/1805.00794.pdf>

9.3 <https://www.researchgate.net/publication/339910895_Arrhythmia_Detection_-_A_Machine_Learning_based_Comparative_Analysis_with_MIT-BIH_ECG_Data>

9.4 <https://www.frontiersin.org/articles/10.3389/fncom.2020.564015/full>

9.5 <https://core.ac.uk/download/pdf/270138548.pdf>