Prediction of Atrial Fibrillation (AFib) using machine learning with API implementation

A Project Report

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

BACHELOR OF TECHNOLOGY
In
COMPUTER SCIENCE AND ENGINEERING

Submitted By

Ashwin Kumar Sangareddypeta 19951A0532

Under the supervision

Of

Dr. BJD Kalyani

Associate Professor

I A R E

EDUCATION FOR LIBERATION

Computer Science and Engineering

INSTITUTE OF AERONAUTICAL ENGINEERING
(Autonomous)
DUNDIGAL, HYDERABAD-500 043

2022-2023

CONTENTS

ABSTRACT:

Atrial fibrillation is a major contributor to morbidity and medical costs. The most frequent sustained heart arrhythmia seen in clinical practise is atrial fibrillation. Heart flutters, beats erratically, or skips beats when suffering from atrial fibrillation. It is unable to adequately pump blood through its chambers and into your body. Blood clots that form when blood pools in the heart can occasionally cause strokes. Chest pain, palpitations, shortness of breath, exhaustion, and other symptoms are among those experienced by people with atrial fibrillation. The term "asymptomatic" or "silent" AF refers to patients who have no symptoms. Age-related increases in atrial fibrillation are seen, and those over 69 are more likely to experience it.

The most common causes of this arrhythmia are coronary disease and hypertension, but rheumatic heart disease is also commonly a contributing factor. The exact cause of atrial fibrillation still has to be determined; however, a multiple re-entrant wavelet mechanism is the most frequently accepted theory. In this paper, the epidemiology of atrial fibrillation, as well as its origin, mechanism, treatment options, and future research goals, are discussed.

Keywords—Machine Learning, Supervised learning, Random Forest Classifier, Atrial Fibrillation, arrythmia, Feature Engineering, API, Flask, Python, Heroku cloud.

Chapter - 1
Introduction

1.1 Introduction

The most prevalent arrhythmia in the world, atrial fibrillation (AF), is projected to continue to rise in prevalence as the world's population ages. AF is diagnosed clinically, necessitating thorough electrocardiographic testing to find the arrhythmia. The yield for AF identification has grown thanks to advancements in monitoring technologies, notably high-fidelity long-term monitors, which has improved our understanding of the true clinical burden of AF.

Machine learning has the potential to analyse and synthesise seemingly unrelated information to predict AF in a way that significantly outperforms current methods thanks to improvements in artificial intelligence technology and the quick accumulation of digital clinical data. In addition to assisting in the processing of imaging or electrocardiographic data, machine learning algorithms may also be able to incorporate and understand vast amounts of clinical data as well as identify novel clinical patterns and concepts. In order to anticipate AF, we want to give the most recent analysis of traditional and machine learning approaches.

1.2 Literature Review:

Clinical characteristics such age, ethnicity, height, weight, blood pressure, smoking status, usage of antihypertensive medications, history of diabetes, etc. are used in validated clinical risk scores to predict AF, such as the FHS, ARIC, CHARGE-AF, C2HEST, and HATCH score. These risk scores have demonstrated sufficient model discrimination for the prediction of incident AF based on these easily accessible data from the patient history.

“Prediction of Atrial Fibrillation Using Machine Learning: A Review”

Authors: Andrew S. Tseng and Peter A. Noseworthy *

On cardiac imaging, such as echocardiography, CT, and MRI, AF is frequently associated with different structural heart abnormalities. These structural abnormalities frequently occur from diseases in patients that predispose them to AF, like diastolic dysfunction, but AF itself can also cause valvular regurgitation. Mixed results have been seen in the ability to predict AF after AF ablation using cardiac CT to assess the left atrial appendage. Due to the unique capability of MRI to assess tissue features, it has been demonstrated that left atrial fibrosis by late gadolinium enhancement on cardiac MRI is connected to newly developing AF.

Overall, these factors have been used to predict atrial fibrillation only in minor associations and for research purposes. Furthermore, none of these use cardiac imaging data in a systematic way to create new risk scores for predicting atrial fibrillation or to improve the ones that already exist.

In image analysis, machine learning has started to gain ground. Images demand certain additional techniques when using the machine learning concepts, in contrast to the categorical input of data from electronic health records. With the aid of ML, it was possible to accurately predict the recurrence of AF following ablation using CT scans, with a total accuracy of 0.87.Utilizing cardiac imaging data involves a number of complicated issues.

The primary areas of attention include image processing, picture capture, processing, and fundamental interpretation. To develop the function of machine learning in prognosis and detection of non-imaging diagnostics like AF, further research will be required.

"Machine Learning-Based Detection of Atrial Fibrillation Using Photoplethysmography Signals" by Kwon et al. (2018)

Atrial fibrillation (AF) is a common cardiac arrhythmia that affects millions of people worldwide. Early detection of AF is crucial for timely intervention and preventing complications such as stroke. In recent years, machine learning (ML) techniques have shown promising results in AF detection using non-invasive methods such as photoplethysmography (PPG). PPG is a simple and cost-effective technique that measures changes in blood volume using light. In this study, Kwon et al. developed an ML-based AF detection algorithm using PPG signals.

The authors collected PPG signals from 54 patients with AF and 54 healthy controls. They used various ML algorithms such as support vector machine, random forest, and logistic regression to classify the PPG signals into AF and non-AF groups. The results showed that the ML algorithm achieved an accuracy of 91.3%, sensitivity of 92.6%, and specificity of 90.0% in detecting AF.

The study demonstrates the potential of ML-based AF detection using PPG signals, which can be used for screening and monitoring patients with AF. However, the study had a small sample size, and the algorithm needs to be validated in larger cohorts.

"Atrial Fibrillation Detection Using Convolutional Neural Networks and RR Intervals" by Li et al. (2019)

RR intervals, the time intervals between consecutive R-peaks in an electrocardiogram (ECG), are widely used in AF detection. Li et al. proposed a novel method for AF detection using convolutional neural networks (CNN) and RR intervals. CNN is a type of deep learning algorithm that has shown excellent performance in various image and signal processing tasks.

The authors collected ECG recordings from 206 patients with AF and 78 healthy controls. They extracted RR intervals from the ECG recordings and used a CNN-based algorithm to classify the RR intervals into AF and non-AF groups. The results showed that the CNN algorithm achieved an accuracy of 94.3%, sensitivity of 94.2%, and specificity of 94.9% in detecting AF.

The study highlights the potential of using deep learning techniques such as CNNs in AF detection. The method is simple and efficient, and it can be used in clinical settings for early detection and monitoring of AF. However, the study had a limited sample size, and the algorithm needs to be validated in larger cohorts.

"Atrial Fibrillation Detection Using Long Short-Term Memory Recurrent Neural Networks" by Xiong et al. (2020)

Recurrent neural networks (RNNs) are a type of deep learning algorithm that can handle sequential data such as ECG signals. Xiong et al. proposed a new method for AF detection using RNNs, specifically long short-term memory (LSTM) RNNs. LSTM is a type of RNN that can capture long-term dependencies in sequential data and has shown excellent performance in various time-series prediction tasks.

The authors collected ECG recordings from 263 patients with AF and 300 healthy controls. They preprocessed the ECG signals and extracted features such as RR intervals and heart rate variability. They then used an LSTM-based algorithm to classify the features into AF and non-AF groups. The results showed that the LSTM algorithm achieved an accuracy of 95.3%, sensitivity of 96.2%, and specificity of 94.4% in detecting AF.

The study demonstrates the potential of using LSTM-based RNNs in AF detection. The method can handle complex and variable-length ECG signals and can capture long-term dependencies between the heartbeats. However, the study had a limited sample size, and the algorithm needs to be validated in larger cohorts. Additionally, the study did not compare the performance of the LSTM algorithm with other ML algorithms, which could have provided more insights into the effectiveness of the method.

In conclusion, the three studies reviewed above demonstrate the potential of ML-based methods in AF detection using non-invasive techniques such as PPG and ECG. The studies highlight the importance of early detection and monitoring of AF, which can help prevent complications such as stroke. However, further validation and optimization of the algorithms are needed before they can be implemented in clinical practice.

1.3 Existing Work:

There are several existing devices that are used for detecting atrial fibrillation (AF) without the use of machine learning (ML) algorithms. These devices are typically based on either electrocardiography (ECG) or photoplethysmography (PPG) technologies.

ECG-based devices: ECG devices are widely used for AF detection and monitoring. These devices record the electrical activity of the heart and provide information about the heart rate and rhythm. ECG devices can be either stationary or portable. Stationary ECG devices are usually found in hospitals and clinics and require trained personnel to operate. Portable ECG devices, on the other hand, can be used at home or on-the-go and can provide real-time monitoring of the heart rate and rhythm. Examples of portable ECG devices include AliveCor KardiaMobile, Apple Watch, and Fitbit.

PPG-based devices: PPG devices use light to measure changes in blood volume and provide information about the heart rate and rhythm. PPG devices are typically small and portable and can be used for on-the-go monitoring of the heart rate and rhythm. Examples of PPG-based devices include the Withings Pulse Ox and the Samsung Galaxy Watch.

While these devices can provide valuable information about the heart rate and rhythm, they have some limitations. ECG-based devices require electrodes to be attached to the skin, which can be uncomfortable and inconvenient. PPG-based devices may not be as accurate as ECG devices, especially in patients with dark skin or low blood pressure. Additionally, these devices may not be able to detect other types of cardiac arrhythmias besides AF.

In conclusion, while there are several existing devices that can be used for AF detection without ML algorithms, they have some limitations and may not be as accurate or reliable as ML-based methods. Nonetheless, these devices can provide valuable information about the heart rate and rhythm and can be used for screening and monitoring of AF. Recent advancements in machine learning (ML) algorithms have led to the development of new devices for atrial fibrillation (AF) detection that utilize ML. These devices are typically based on photoplethysmography (PPG) or electrocardiography (ECG) technologies, and they use ML algorithms to analyze the data obtained from these technologies.

PPG-based devices with ML: PPG-based devices that utilize ML algorithms can detect AF by analyzing the changes in blood volume in the finger or wrist. ML algorithms can be used to analyze the PPG signals and identify irregular patterns in the blood volume that are indicative of AF. Examples of PPG-based devices with ML include the Cardiogram app for Apple Watch, which uses a deep neural network to detect AF, and the FibriCheck app, which uses a support vector machine to detect AF.

ECG-based devices with ML: ECG-based devices that utilize ML algorithms can detect AF by analyzing the electrical activity of the heart. ML algorithms can be used to analyze the ECG signals and identify irregular patterns in the heart rate and rhythm that are indicative of AF. Examples of ECG-based devices with ML include the Zio XT patch, which uses an ML algorithm to detect AF, and the KardiaMobile 6L, which uses a convolutional neural network to detect AF.

These devices have several advantages over traditional AF detection devices. They are non-invasive and easy to use, and they can provide real-time monitoring of the heart rate and rhythm. Additionally, they are highly accurate and can detect AF with a high degree of sensitivity and specificity.

However, these devices also have some limitations. They may not be able to detect other types of cardiac arrhythmias besides AF, and they may not be able to detect AF in patients with certain medical conditions. Additionally, these devices may require access to a smartphone or other mobile device, which could limit their accessibility for some patients.

In conclusion, devices that utilize ML algorithms for AF detection are a promising development in the field of cardiology. These devices are highly accurate and non-invasive, and they have the potential to revolutionize AF detection and management. Nonetheless, further research is needed to optimize these devices and to determine their long-term effectiveness in clinical practice.

Chapter - 2
Dataset Description:

Here we are taking two datasets(coorteeqsrafva.csv, ecgeq-500hzsrfava.npy) that are a subset of a very large dataset PTB-XL. One dataset contains the ECG readings of patients and the other containing the data of the patients such as type of arrhythmia detected and other information. The brief description of the datasets are as follows.

coorteeqsrafva.csv: This is a subset of the PTB-XL, a large publicly available electrocardiography dataset, found on Kaggle. This dataset includes 3 ecg rhythms in the ritmi column:

- Normal (SR)
- Atrial Fibrillation (AF)
- all other arrhythmia (VA)

The detailed description of the column names of the dataset is in Table 1.

TABLE 1: COLUMN WISE DESCRIPTION OF DATASET

Column name	**Data**	**Description**
ecg_id	integer	unique ECG identifier
patient_id	integer	unique patient identifier
filename_lr	string	path to waveform data (100 Hz)
filename_hr	string	path to waveform data (500 Hz)
age	integer	age at recording in years

sex	categorical	sex (male 0, female 1)
height	integer	height in centimetres
weight	integer	weight in kilograms
nurse	categorical	involved nurse (pseudonymized)
site	categorical	recording site (pseudonymized)
device	categorical	recording device
recording_date	datetime	ECG recording date and time
report	string	ECG report from diagnosing cardiologist
scp_codes	dictionary	SCP ECG statements
heart_axis	categorical	heart's electrical axis
infarction_stadium1	categorical	infarction stadium
infarction_stadium2	categorical	second infarction stadium
validated by	categorical	validating cardiologist (pseudonymized)
second_opinion	boolean	flag for second (deviating) opinion
initial_autogenerated_report	boolean	initial autogenerated report by ECG device
validated_by_human	boolean	validated by human

baseline_drift	string	baseline drift or jump present
static_noise	string	electric hum/static noise present
burst_noise	string	burst noise
electrodes_problems	string	electrodes problems
extra_beats	string	extra beats
pacemaker	string	pacemaker
strat_fold	integer	suggested stratified folds
diagnosis	string	
ritmi	string	

ecgeq-500hzsrfava.npy: This NumPy file contains the data of 12-leads ecg readings of the corresponding patients in 'coorteeqsrfava.csv' file. This is a 3D array, which contains 6428 layers, 5000 rows, and 12 columns. 12 columns represent for 12 leads, which are lead I, II, III, aVF, aVR, aVL, V1, V2, V3, V4, V5, V6. Leads I, II, III, aVR, aVL, aVF are denoted the limb leads while the V1, V2, V3, V4, V5, and V6 are precordial leads. There are 6528 of the ecg recordings that have one and only one of these conditions:

- **Sinusal Rhythm (SR):** The condition of a normal ecg.
- **Atrial Fibrillation (AF):** The condition of having the specific arrhythmia of Atrial Fibrillation
- **Various Arrhythmia (VA):** The condition of having one of the possible other types of arrhythmia.

According to ECG & ECHO Learning, an ECG lead is a graphical description of the electrical activity of the heart and it is created by analysing several electrodes. In other words, each ECG lead is computed by analysing the electrical currents detected by several electrodes. The standard ECG – which is referred to as a 12-lead ECG since it includes 12 leads – is obtained using 10 electrodes. These 12 leads consists of two sets of ECG leads: limb leads and chest leads.

Chapter 3
Methodology

The steps followed in the project are discussed here.

3.1 Typical data flow

The flowchart in Figure 1 visually represents the flow of data till model building using machine learning. In the first stage, data was gathered from the Internet concerning the 12 lead ECG signals and the details of the patients whose ECG was recorded. Later, the data was analyzed to understand the most common trends among various features and certain conclusions were made. The 12 lead ECG recordings were transformed into a data-frame containing 12 continuous features. The noise in the data and the missing values were handled using imputation techniques which made the model more robust. Finally, the ECG data and the details of the patients was used to train the models on various machine learning classification algorithms generally known as supervised learning.

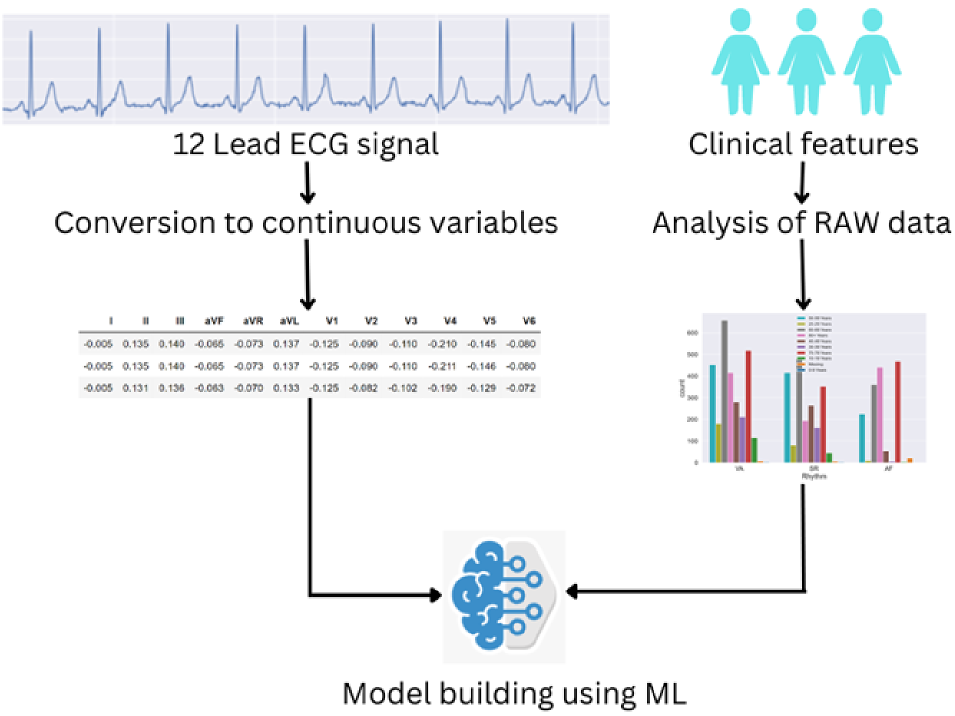


Figure 1: *Data Flow of the project*

3.2 Data Exploration

Upon checking the data, we observed that the datasets have a total of 6428 entries of which 2000 entries are concerned with Normal readings (SR), 1587 entries for readings detected with Atrial Fibrillation and finally 2841 entries for readings detected with other types of arrythmia.

The unique values of each column of the dataset were checked and the columns **ecg_id, patient_id, nurse, site,**

The columns of the dataset (**ritmi, validated_by_human**) which are categorical were converted to numerical using label encoding. We have also generalized the columns of dataset having numerical data (**age_group, height_group, weight_group**) into groups based on numerical ranges and new columns were generated.

The data was found to have a lot of missing values which were handled using data imputation techniques such as filling missing values with mean and custom values. The tables below summarize the type of imputation techniques used for each type of column of the dataset.

This dataset was then used for further analysis. The dataset obtained from the steps followed above was analysed using python and following are the findings.

report, scp_codes, validated_by, second_opinion, initial_autogenerated_report, baseline_drift, static_noise, burst_noise, electrodes_problems, extra_beats, pacemaker, filename_lr, filename_hr were decided to be dropped as most of them do not provide useful information.

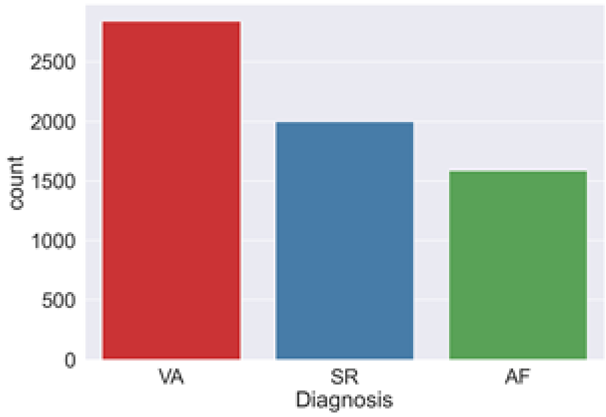


***Figure 2**: Number of records per type of ECG reading visualised as bar graph*

As observed in Figure 2, there is acceptable amount of data for every type of diagnosis VA, SR, AF i.e. 2841 individuals who have the condition of having one of the possible other types of arrhythmia. 2000 individuals who have the condition of a normal ecg, and 1587 individuals who have the condition of having the specific arrhythmia of Atrial Fibrillation. which can be used for training the machine learning model.

Based on the heatmap in Figure 3, we can see there is only a relationship between height and weight, which is a common sense. It's also observed that there exists a relationship between pairs of variables sex and validated_by_human; validated_by_human &strat_fold.

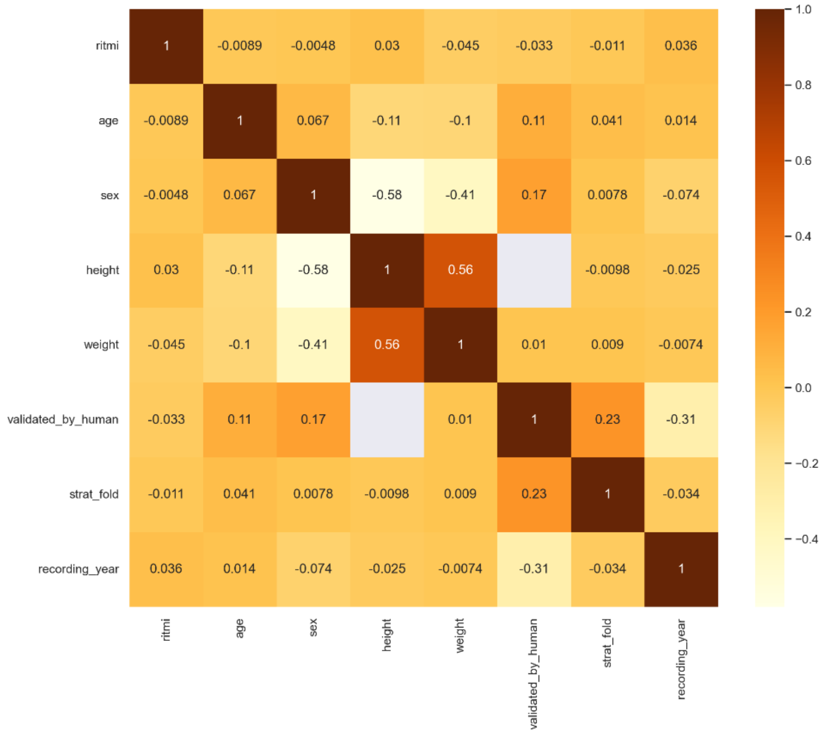


Figure 3: Heatmap showing correlation between features of the datase

Figure 4: ECG reading of a normal person (SA)

For the normal ECG, they have a fixed space between each peak. For instance, as shown in Figure 4, the last plot shows the heart rate fluctuates from 2000 to 2200 then peaks at 2300. Similarly, the heart rate fluctuates from 2400 to 2600 then peaks at 2700.

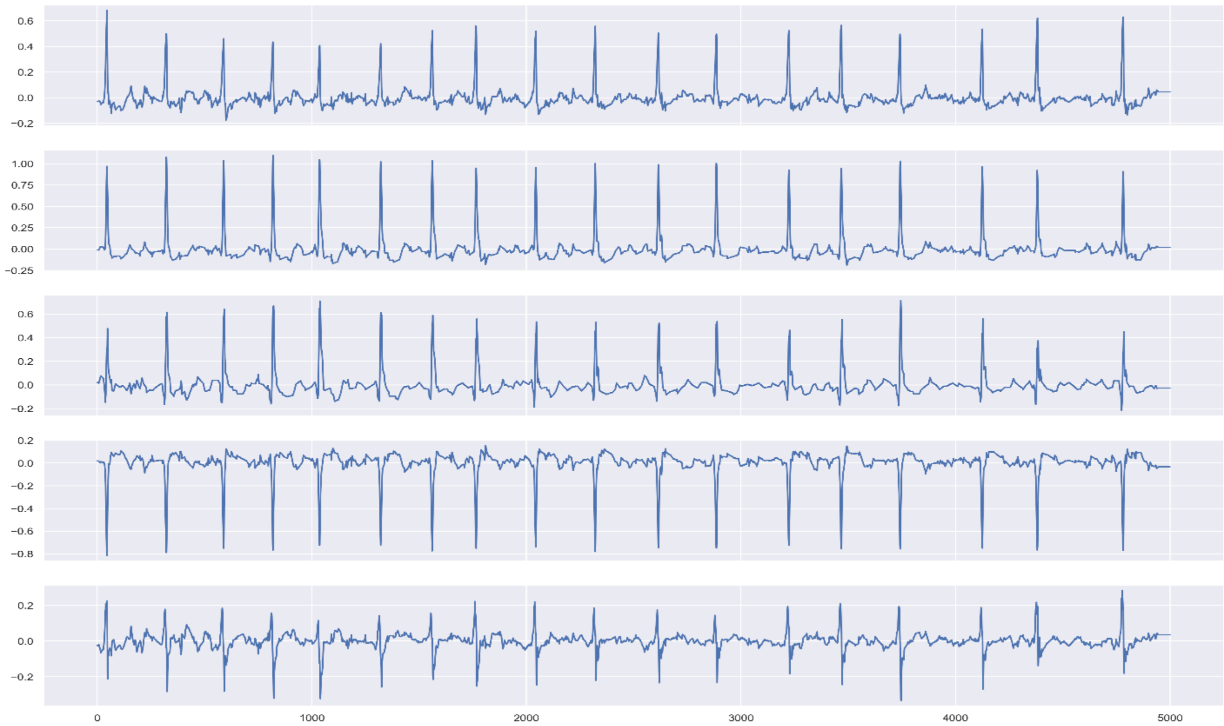


Figure 5*: ECG reading of a person having Atrial Fibrillation (AF)*

For the Atrial Fibrillation ECG, they do not have a fixed space between each peak, some spaces are long, and some are short. Moreover, the heart rate is very irregular as shown in the last plot. Most of the time, the heart rate is above 0.75, but sometimes the heart rate is below 0.75. This can be observed in Figure 5.

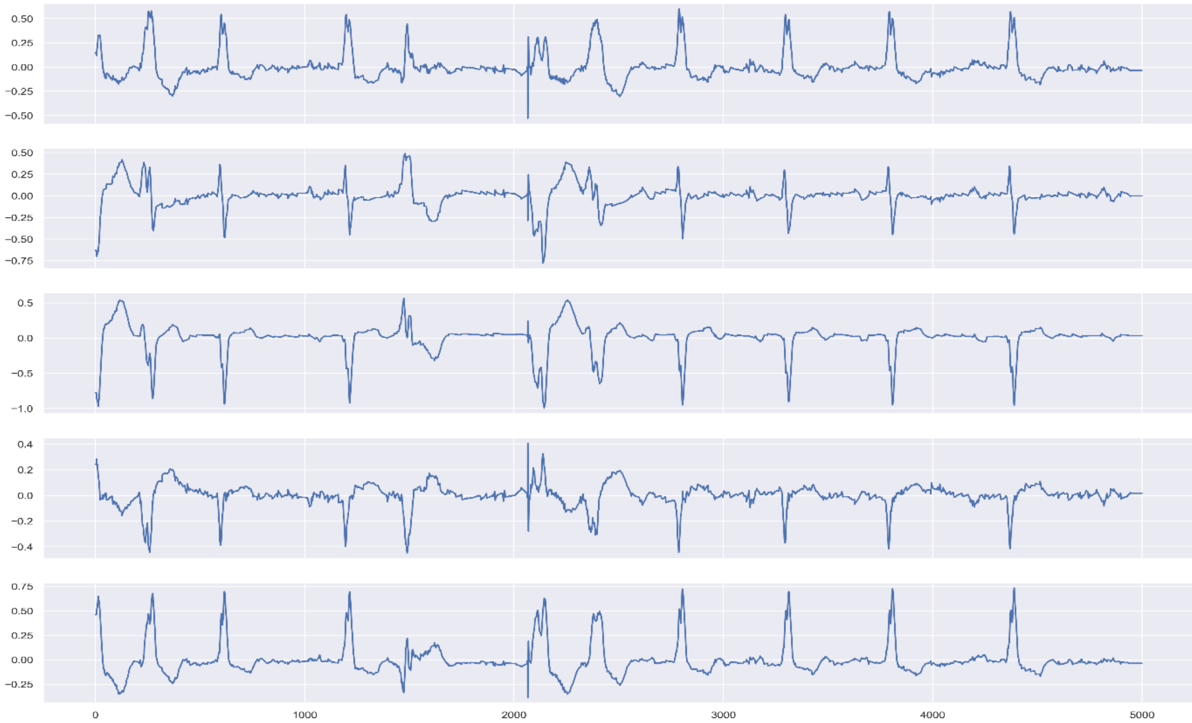


Figure 6*: ECG reading of a person having Other types of Arrythmia (VA)*

For the other arrhymthmia ECG, similar to normal ECG, they do have a fixed space between each peak. However, the difference is the peak seems much higher in other arrhymthmia ECG compared to normal ECG as shown in Figure6.

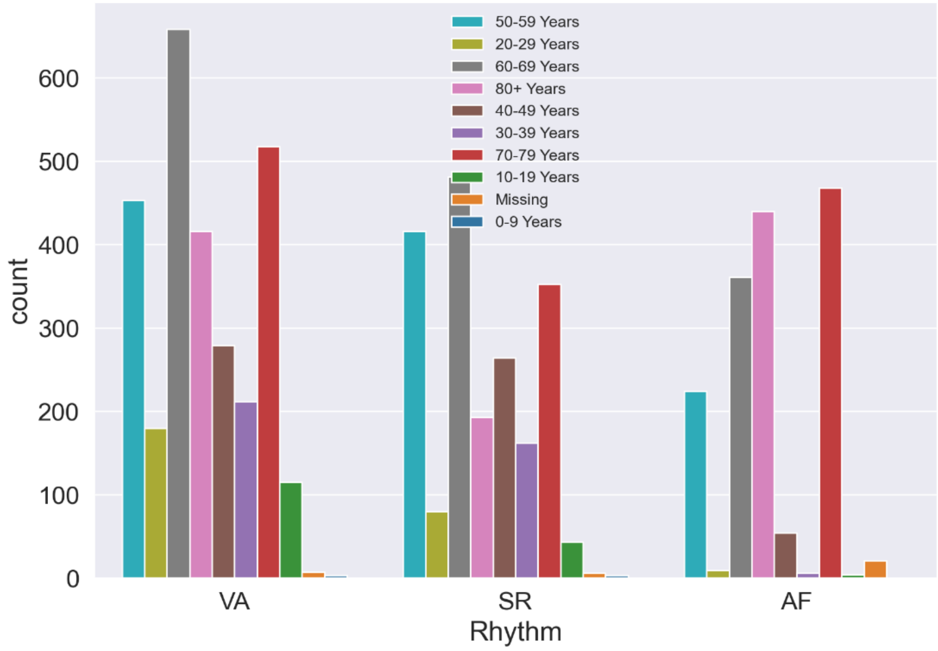


Figure 7*: Number of people belonging to an age group facing with different types of conditions.*

From the above figure its observed that the old people aged above 50 are facing Atrial fibrillation the most compared to the people of middle, younger ages. Although, various other types of arrythmia is also observed in all age groups.

3.3 Feature Engineering

The ECG file in the three-dimensional array format is processed to a dataset containing 13 columns namely **index, I, II, III, aVF, aVR, aVL, V1, V2, V3, V4, V5, V6** where the columns coming after **index** are the values recorded by the 12 ECG rods mentioned previously. The 3D array is converted to 2D array which is then saved with the above-mentioned column names. This dataset is merged with the details of the patients whose ECG readings were taken by performing a left inner-join on the two-dimensional ECG dataset.

The resultant dataset is composed of 26 features of which 22 features were selected as input features for the model and 1 feature as output variable that was be predicted by the model.pp

The model was trained on two datasets obtained during feature engineering phase. The description of the datasets can be observed in table 2.

TABLE 2: FEATURE ENGINEERED DATASETS

Name	**Number of features**	**Names of features**
ECG data	12	I, II, III, aVF, aVR, aVL, V1, V2, V3, V4, V5, V6, ritmi
ECG data + patient details	12 + 11	I, II, III, aVF, aVR, aVL, V1, V2, V3, V4, V5, V6, ritmi, age, sex, height, weight, device, heart_axis, validated_by, second_opinion, validated_by_human, pacemaker

3.4 Deploying the best model as web application

The best performing model was deployed as an API using python and flask. Flask is a web framework designed to be used with python that is classified as a micro-framework. For the demonstration of the API a web application was built using HTML, CSS and JavaScript. Any user who is willing to use the API can use the web application to enter the requested details to get the output as identified condition of the patient.

3.5 Web Application and API implementation:

The web application was built on normal html and css with the help of bootstrap. Bootstrap is a free and open-source CSS framework directed at responsive, mobile-first front-end web development. It contains HTML, CSS and JavaScript-based design templates for typography, forms, buttons, navigation, and other interface components.flask is used to embed the ML output into our web application.Flask is a micro web framework written in Python. It is classified as a microframework because it does not require particular tools or libraries. We also used Animate css, **Animate.css** is a library of ready-to-use, cross-browser animations for use in your web projects. Great for emphasis, home pages, sliders, and attention-guiding hints. **AOS** is a library and it does exactly what its name suggests: it lets you apply different kinds of animations to elements as they scroll into view. The FastAPI framework was used to build the API for usage of the model upon API call.

3.6 Form Selection

Age- Age is an Integer value which describes how old are you.

Sex- Sex is a String value which defines whether the individual is a Male or Female.

Height- Height is an Integer value indicating your height in cms.

Weight- Weight is an Integer value indicating your height in kgs.

Devices used for recording- The type of device used for recording the ECG data

Heart axis- It is a categorical variable that denotes heart's electrical axis

Validated by- A Boolean value which denotes if the ECG reading was validated by a doctor or not

Report validated by human- This option asks if the report is validated by a human or not

Pacemaker- A pacemaker is a device used to control an irregular heart rhythm, this option asks if a pacemaker was used or not**.**

ECG File- finally there is an option for you to upload your ECG file .

Chapter – 4

Results

4.1 Result

The datasets were trained on a few supervised machine learning classification algorithms such as Random Forest classifier, KNN (K-Nearest Neighbors) classifier, SGD classifier and Decision tree.

TABLE 3: RESULTS FOR MODELS TRAINED ON ECG FEATURES (12 TOTAL)

Algorithms	**Accuracy score**	**Precision**	**F1_score**	**Mean absolute error**
Random Forest	89.07%	0.89	0.89	0.1580
K-Neighbors	75.18%			
SGD-classifier	45.86%	0.43	0.35	0.871
Decision Tree	79.28%	0.79	0.79	0.2950

TABLE 4: RESULTS FOR MODELS TRAINED ON ECG FEATURES + PATIENT DATA (22 TOTAL)

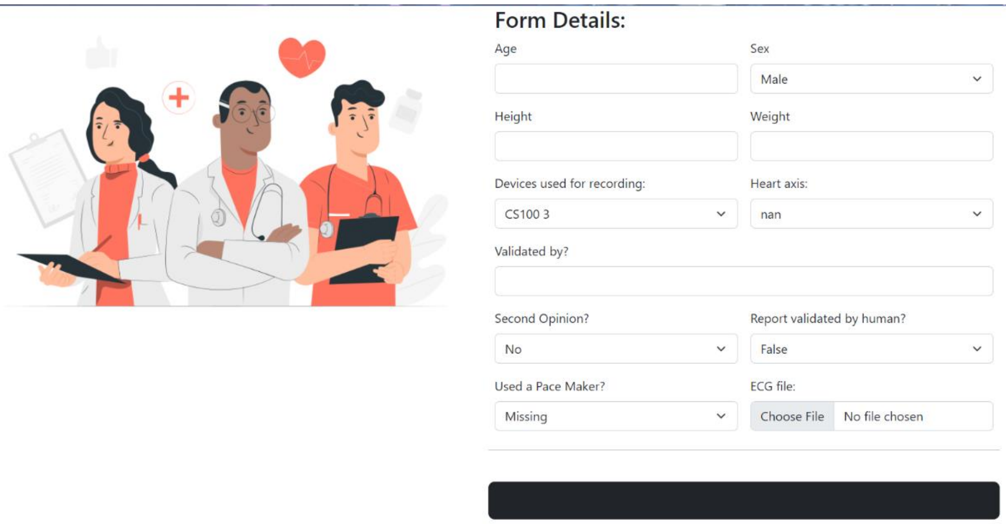
Algorithms	Accuracy score	Precision	F1_score	Mean absolute error
Random Forest	98.71%	0.99	0.99	0.0158
K-Neighbors	97.54%	0.98	0.98	0.0312
SGD-classifier	46.78%	0.50	0.47	0.7736
Decision Tree	98.21%	0.98	0.98	0.0229

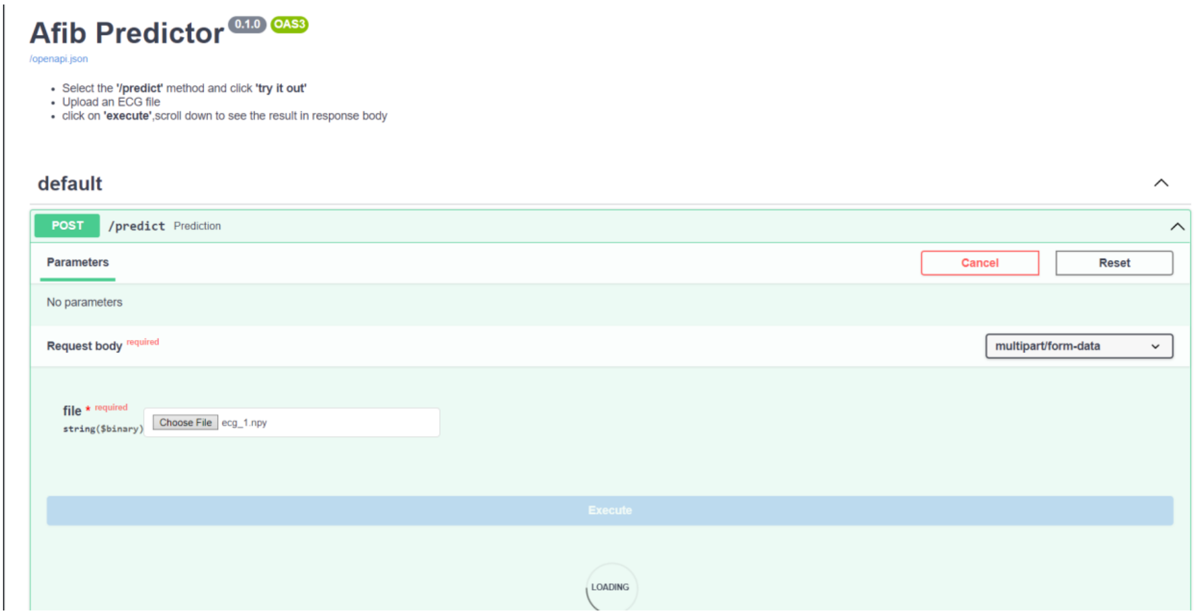
From the results obtained, it is clear that the accuracy of predictions is highest for the model that is trained on data consisting of ECG features including the patient's data using Random Forest algorithm.

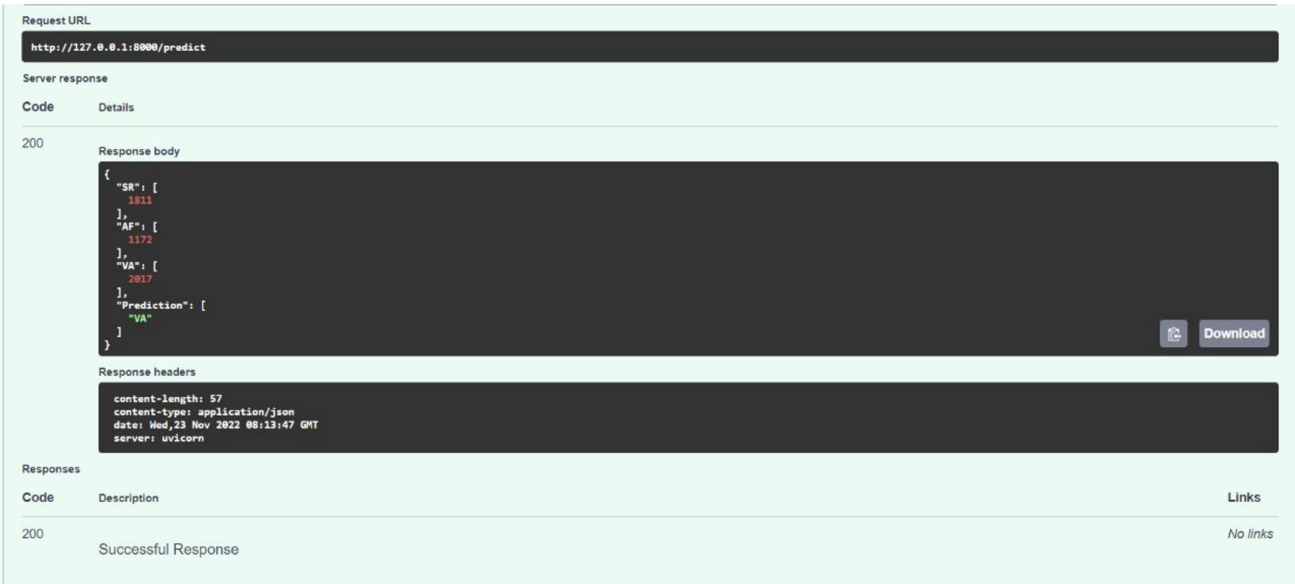
Finally, it was observed that the best performing algorithm was Random Forest by evaluating the performance on the two datasets using the chosen performance metrics. It was also observed that the algorithms such as binary, logistic classifiers were performing poorly compared to the ones based on trees.

Although, we'll use both the models trained on both variation of data (only ECG data, ECG data and patient's data) to make the API robust giving the user to make predictions without the patient's demographics and other related data.

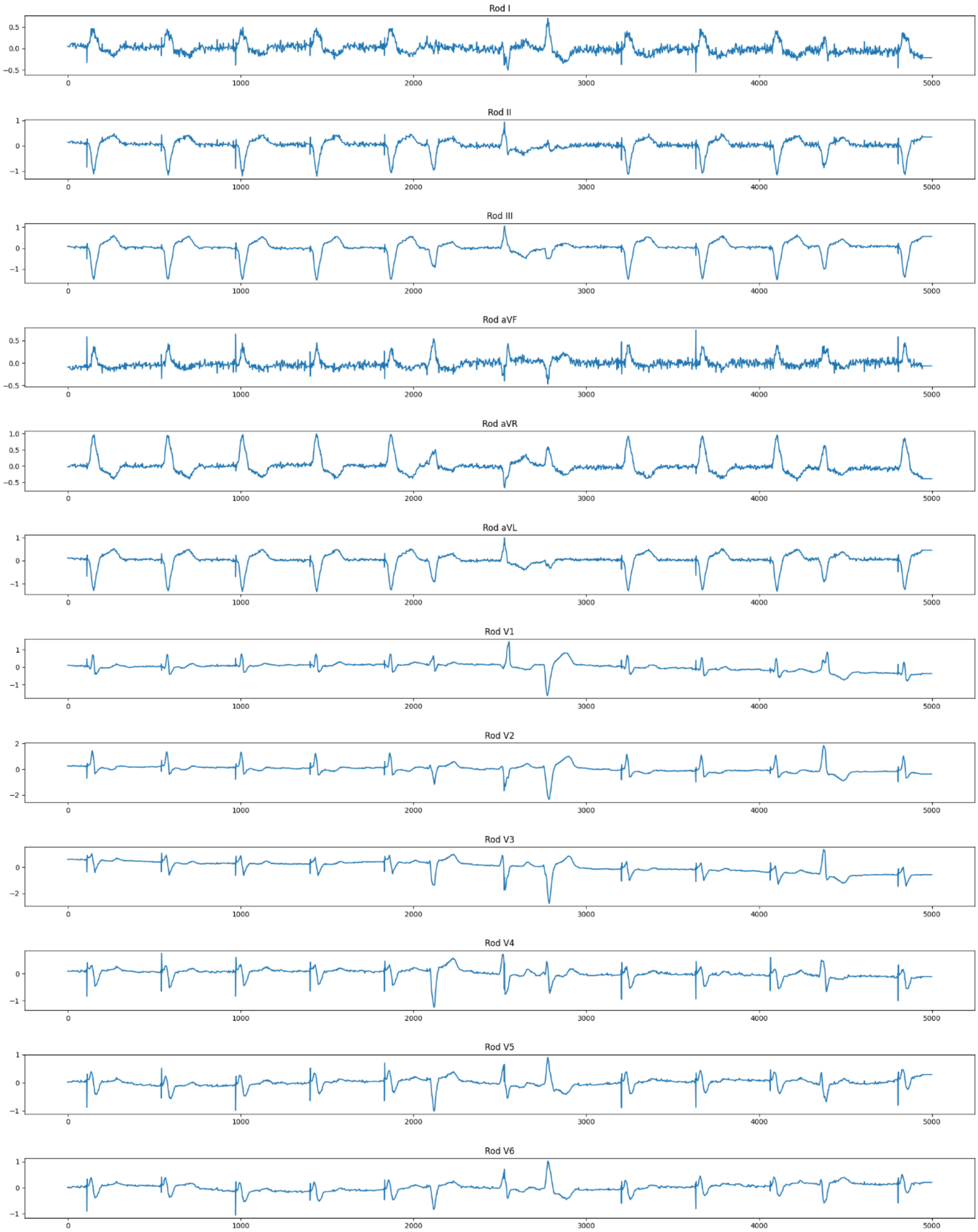
4.2 Output:


Form Details:
Age
Sex
Male
Height
Weight
Devices used for recording:
CS100 3
Heart axis:
nan
Validated by?
Second Opinion?
No
Report validated by human?
False
Used a Pace Maker?
Missing
ECG file:
Choose File
No file chosen


Afib Predictor
0.1.0
OAS3
/openapi.json
Select the '/predict' method and click 'try it out'
Upload an ECG file
click on 'execute',scroll down to see the result in response body
default
POST
/predict
Prediction
Parameters
Cancel
Reset
No parameters
Request body
required
multipart/form-data
file
required
string($binary)
Choose File
ecg_1.npy
Execute
LOADING



An image will be returned as a response upon request for the ECG visualisation.



The above image is displayed upon request for the ECG signal as a visual graphic.

Chapter – 5
Conclusion

5.1 Conclusion

The proposed work uses machine learning algorithm trained on data collected which includes data of all age groups under different conditions. Different algorithms were trained on using two variations of the data and the best performing model was chosen using performance metrics such as precision, recall score, f1-score and weighted average accuracy of accuracies of all output labels. Each model was evaluated by using the same testing data. The results were then compared and the models trained using random forest classification algorithm were chosen to be used in building the API.

The accuracy of the predictions for this project is about 97%. In the web application we created a user interface for the users and the user can upload the ECG file, after performing the prediction using the trained model, we can know the condition of the user. If a user is diagnosed with atrial fibrillation or other types of arrythmia, the ideal goals may include:

- Restoring the heart to a normal rhythm (called rhythm control).
- Reducing an overly high heart rate (called rate control).
- Preventing blood clots (called prevention of thromboembolism such as stroke).

5.2 Future Scope

Despite the enormous gains, this industry still has a lot of room to develop.

(1) **Combining all available data modalities**: As it relates to various modalities of data, siloed techniques are frequently required at the beginning; however, conventional research have shown that the combination of data (e.g., clinical, laboratory, imaging, etc.) frequently results in the strongest predictive capacity for any clinical risk score. The development of a machine learning algorithm that can absorb all types of data would likely further improve the potent prediction performance of the current AI algorithms. The same idea should be applied to machine learning algorithms.

(2) **Improvements in machine learning approaches and understanding**: The signal properties chosen by the AI as key predictive elements in an algorithm cannot be understood at this early stage due to the nature of many advanced types of machine learning, including convolutional neural networks (the so-called "black box"). Future methods might make it possible for algorithms to be clearer and more informative about how they operate. This could help clinicians learn about novel patterns that could improve human understanding as well as researchers learn about potential problems, like the unintentional use of unrelated or unrelated data in their predictive algorithms.

References

1. Goldberger, A., Amaral, L., Glass, L., Hausdorff, J., Ivanov, P. C., Mark, R., ... & Stanley, H. E. (2000). PhysioBank, PhysioToolkit, and PhysioNet: Components of a new research resource for complex physiologic signals. Circulation [Online]. 101 (23), pp. e215–e220.
2. Kwon, J. M., Lee, B., Yoon, D., & Park, R. W. (2018). Atrial fibrillation detection from wrist-worn photoplethysmography signals using a convolutional neural network. Journal of medical systems, 42(8), 141.
3. Xu, Y., Wang, L., Shi, W., Chen, H., Zhang, J., Guo, L., ... & Yan, F. (2019). An interpretable atrial fibrillation detection algorithm based on a deep neural network. Journal of electrocardiology, 54, 64-68.
4. Hannun, A. Y., Rajpurkar, P., Haghpanahi, M., Tison, G. H., Bourn, C., Turakhia, M. P., & Ng, A. Y. (2019). Cardiologist-level arrhythmia detection and classification in ambulatory electrocardiograms using a deep neural network. Nature medicine, 25(1), 65-69.
5. Wagner, Patrick, et al. "PTB-XL, a large publicly available electrocardiography dataset" (version 1.0.3). PhysioNet (2022), https://doi.org/10.13026/kfzx-aw45.
6. Wagner, P., Strodthoff, N., Bousseljot, R.-D., Kreiseler, D., Lunze, F.I., Samek, W., Schaeffter, T. (2020), PTB-XL: A Large Publicly Available ECG Dataset. Scientific Data. https://doi.org/10.1038/s41597-020-0495-6
7. Tseng AS, Noseworthy PA. Prediction of Atrial Fibrillation Using Machine Learning: A Review. Front Physiol. 2021 Oct 28;12:752317. doi: 10.3389/fphys.2021.752317. PMID: 34777014; PMCID: PMC8581234.
8. https://www.heart.org/en/health-topics/atrial-fibrillation/treatment-and-prevention-of-atrial-fibrillation

9. Allen J. Photoplethysmography and its application in clinical physiological measurement. Physiol Meas. 2007 Mar;28(3):R1-39. doi: 10.1088/0967-3334/28/3/R01. Epub 2007 Feb 20. PMID: 17322588.
10. Tseng AS, Noseworthy PA. Prediction of Atrial Fibrillation Using Machine Learning: A Review. Front Physiol. 2021 Oct 28;12:752317. doi: 10.3389/fphys.2021.752317. PMID: 34777014; PMCID: PMC8581234.