IMPLEMENTATION OF AN IoT BASED HEALTHCARE KIT

PROJECT ID
S - 13

6 Credits

Team composition

Sinchitha H V	PES1UG19EC301	
Tanya Chanchalani	PES1UG19EC326	
Bhushan Kiran Munoli	PES1UG19EC359	
Gaurav R	PES1201801939	

Guide : Dr. Purushotham U (ECE)

Contents

- Introduction
- Problem Statement
- Literature Review
- Motivation
- Block Diagram
- Flow chart / Circuit Design
- Work Progress
- Hardware Implementation details / Software details
- Results / Summary/ Conclusion
- References
- Deliverables
- Project timeline Gantt Chart (Aug-Dec 2022)
- Q & A

Problem Statement / Objective

- To build and design a Healthcare kit using IoT which measures ECG in order to predict the type of heart disease using Machine Learning. Additionally, provide an application interface for IoT devices for easy access and analysis.

Motivation

Studies have shown that 30 percent of patients with a discharge diagnosis of heart failure are readmitted at least once within 90 days and such cases happen because, there is provision for continuous monitoring of patient's health in hospitals. However, there is no provision for checking health parameters when they return home, so there is a risk of disease recurrence. Hence this proposed system should measure patients' health parameters related to heart diseases and show results accordingly. And, also our research at a government hospital has shown that at least 1 in every 10 patients is positive for CVD.

The prime objective is to build up a dependable patient-checking framework with the goal that the patients, who are either hospitalized or executing their ordinary day-to-day life exercises can screen and self-examine their heart-related health parameters safely.

Literature Review

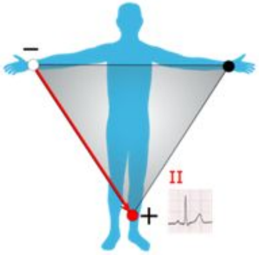
Reference	Key Takeaway
[1] Ahamed, Jameel & Koli, Abdul & Ahmad, Khaleel & Jamal, Mohammad & Gupta, B B. (2021). CDPS-IoT: "Cardiovascular Disease Prediction System Based on IoT using Machine Learning." International Journal of Interactive Multimedia and Artificial Intelligence. In Press. 1-9. 10.9781/ijimai.2021.09.002.	Random Forest search gave 90.08% accuracy which is higher than all other proposed classifiers. This method is hence the best for implementation and model building as it gives lesser error rates also. The dataset used here is taken from kaggle - heart disease uci.
[2]Shadman Nashif, Md. Rakib Raihan, Md. Rasedul Islam, Mohammad Hasan Imam, **"Heart Disease Detection by Using Machine Learning Algorithms and a Real-Time Cardiovascular Health Monitoring System"**, Department of Electrical & Electronic Engineering, American International University-Bangladesh (AIUB), Dhaka, Bangladesh, Vol.6 No.4, November 2018	A total of 13 features were taken into consideration. It was seen that SVM, Random Forest and Simple Logistic models showed higher accuracy rates of more than 95% which make them considerable models for this problem. Parameters like accuracy, sensitivity, specificity, precision, and F-score were determined for each model. SVM showed the best performance among other models with radial basis kernel function.

Literature Review

References	Key Takeaway
[3] Mr. Yathish DP & Dr. Dayanand Lal N, "**Early Detection of Cardiac Arrhythmia Disease using Machine Learning and IoT Technologies**", 2021 2nd International Conference on smart Electronics and Communication (ICOSEC) 2021 IEEE DOI:10.1109/ICOSEC51865.2021.9591884.	The heart rate variability (HRV) features were considered and the SVM model was used mainly with Linear Kernel and accuracy was about 76%. For ECG feature extraction Pan-tompkins QRS detection algorithm is used.
[4] Padmavathi Kora, A Rajini, M C Chinnaiah & K Meenakshi, "**IoT Based Wearable Monitoring Structure for Detecting Abnormal Heart**", 2021 International Conference on Sustainable Energy and Future Electric Transportation (SEFET), 21-23 January 2021 DOI : 10.1109/SeFet48154.2021.9375787	A wearable ECG is designed to detect abnormal heart conditions. It uses a three wireless electrodes, a specialist framework focused on Java and a web-enabled surveillance network. If any change of the health condition from their normal is observed, then it will be transmitted it to a health center for early analysis and preventative actions. This saves the life of the patients from Heart attacks.

Methodology

- The ECG electrodes which are extended from AD8232 sensor are placed on the body in the form of a triangle famously known as einthoven triangle and enables recording of electrical currents/biopotential signals.
- the red electrode on the right arm/shoulder, the yellow electrode on the left arm/ shoulder and the green electrode on the leg/thigh. The heart is in the center of the triangle.Each electrode gives its own voltage. Different combination of +ve and -ve electrodes give different ECG graphs.
- **Lead I=LA-RA.**
- **Lead II=LL-RA.**
- **Lead III=LL-LA.**



- Arduino Uno is connected to Rpi using a communication bus(UART).
- The Rpi provides specific ECG related information like Max heart Rate, ST slope and T inversion which is required in the prediction of CVD.
- Two Butterworth filters in the code are used for ECG signal processing.
- Libraries namely heartpy and neurokit2 are used for ECG peaks detection.

<u>The features considered for CVD :</u>

- Age, gender , chest pain type, resting blood pressure, cholesterol, resting ECG values, exercise induced angina, maximum heart rate, oldpeak, slope, number of major vessels involved and thalassemia.

Methodology

The features considered for Arrhythmia :

Age, Sex, Height, Weight, QRS duration, P-R interval, Q-T interval, T interval, P interval, Vector angles in degrees on front plane of QRS, T, P, QRST and J, Heart rate

- Training and prediction of data is done using UCI datasets for CVD and Cardiac arrhythmia by applying various machine learning algorithms.
- Support Vector Method is giving the highest accuracy in both machine learning models.

The features considered for Hospital Dataset:

Age, Gender, Height, Weight, Resting Blood Pressure, Cholesterol Blood Sugar, Alcohol, Cad History, Stroke History, Peripheral Heart Disease, Chest Pain Type, Exercise Induced Angina, Max Heart Rate, ST Peak, T Wave Inversion, Condition

- Oversampling, undersampling is done on the unbalanced hospital dataset.
- Then, SMOTETomek technique is applied which combines both undersampling (using Tomek Links) and oversampling (using SMOTE) in order to obtain better accuracy for the model.
- To test the model Random Forest Classifier is used as it can handle the data set containing categorical variables in the case of classification.

Methodology

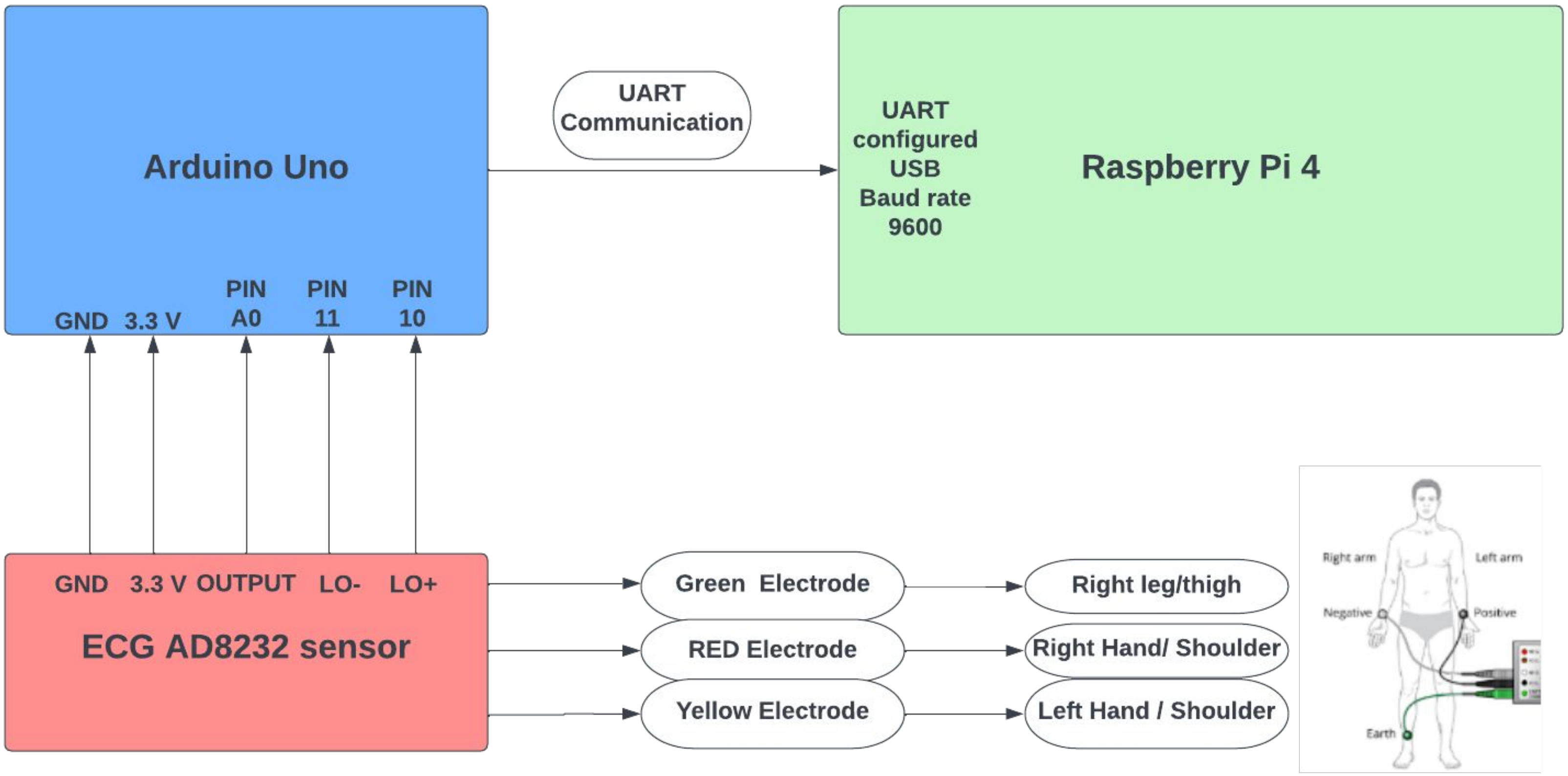
Interface

- A separate independent user interface is built using HTML, CSS, Jinja2, Gunicorn, Flask framework, including modules like MarkupSafe.
- For each disease where the user provides necessary input which is tested using the machine learning models and final results are displayed on the GUI.

Implementation details of the projects:

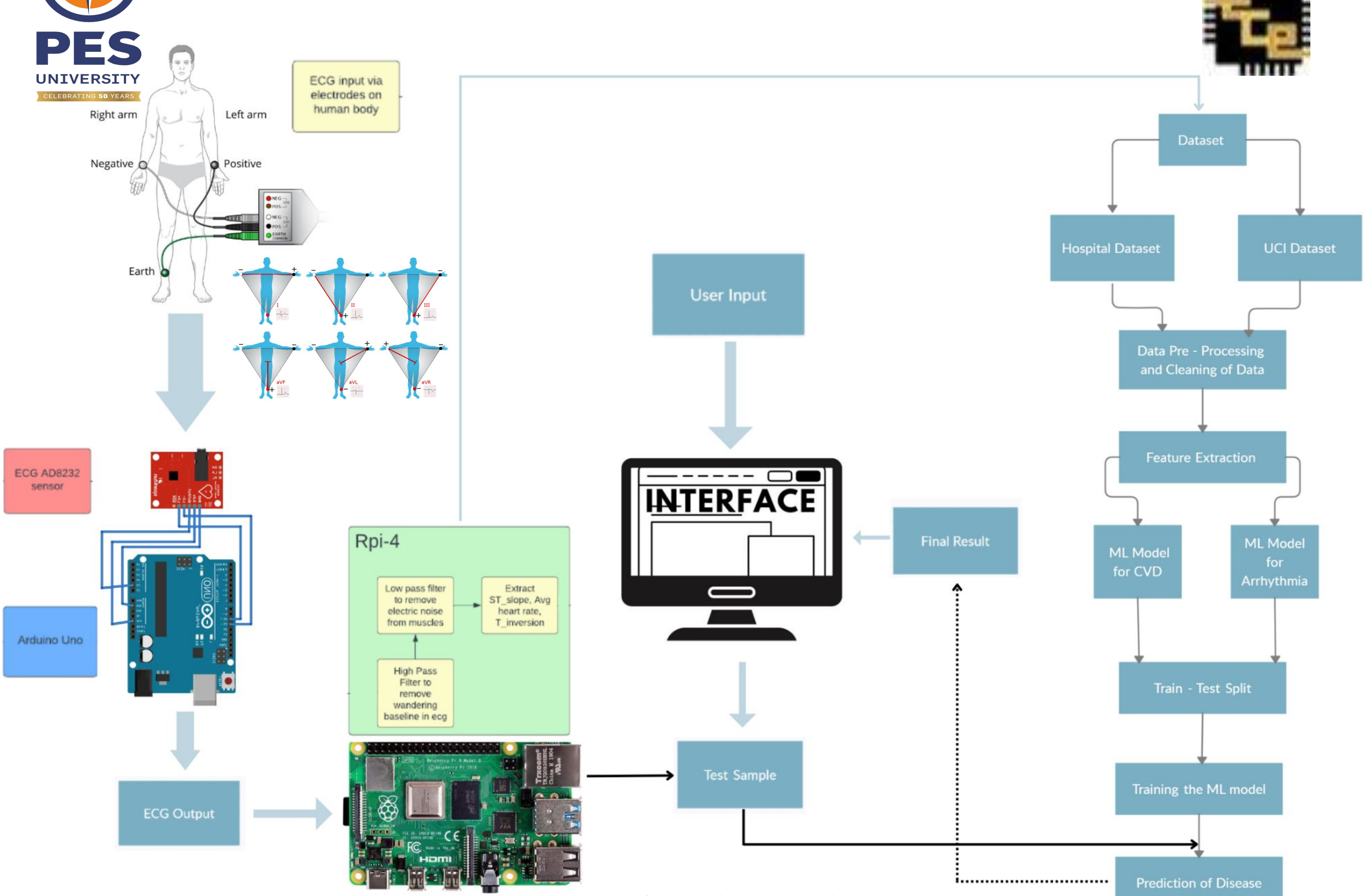
- Hardware - Raspberry pi-4, Arduino, ECG AD8232 Sensor
- Software/Programming Language - Python with Flask Framework for rendering the code onto the GUI which is built using HTML and CSS.
- IDE used - Jupyter Notebook, Arduino IDE
- Libraries - Numpy, Pandas, Keras, TensorFlow, Math, Matplotlib, Operator, Seaborn, SKLearn, Collections, Flask(framework) and Tkinter for GUI to include a few.
- Hospital Dataset - Dataset collected from the Hospital visited.
- Heart disease dataset from UCI Website and Arrhythmia dataset from the UCI Website.

Block Diagram



Circuit Diagram For Hardware

Block Diagram



Working Principle / Details

- The ECG patches are placed in the form of a einthoven triangle i.e. the red electrode on the right arm/shoulder, the yellow electrode on the left arm/ shoulder and the green electrode on the leg/thigh. The heart is in the center of the triangle. Then ECG signal is extracted using AD8232 sensor.
- Then this signal is processed using butterworth filter, in order to obtain ST slope, Max Heart is acquired by extrapolating the number of R peaks in the ECG signal, and T inversion checking is done by estimating the T peak to be less than -0.1mv. This is done because the baseline of the ECG can have an increase or decrease of 0.1mV, so to avoid false positive results or vice versa.
- These three parameters obtained from the hardware are the inputs to the interface for CVD along with other important parameters.
- Further, two separate machine learning models are developed and trained using UCI datasets of CVD and Arrhythmia.
- Here, data is cleaned and processed to extract features according to the parameters required for the two heart diseases selected.
- The algorithm used for CVD are Logistic Regression, KNN, Support vector machine(SVM), Random Forest and Decision Tree Classifier.

Working Principle / Details

- And, the algorithms used for cardiac Arrhythmia are, initially PCA is applied for feature extraction followed by SVM with Random Forest Classifier using linear, RBF, and polynomial kernels with various regularization parameters to further enhance the accuracies.
- It was observed that SVM with linear kernel gave the highest accuracy for CVD and Arrhythmia respectively while using testing dataset.
- After training the models, in order to test them, data is manually entered onto the interface according to the parameters specified for each diseases.
- The inputs given are processed in the ML models connected to the two interfaces. And the final outputs are displayed through the GUI built for both diseases.
- For CVD the results displayed is either class 1 or 0 indicating the presence of disease or not.
- For cardiac arrhythmia the output includes 16 classes based on value of attributes.
- Class 1 - Normal case, Class 2 to 15 - Each class indicating various types of Arrhythmia. Class 16 - Unknowns

Working Principle / Details

Disease Considered	Cardiac Arrhythmia
Name of creators	H. Altay Guvenir, PhD, Bilkent University; Burak Acar, M.S., Bilkent University; Haldun Muderrisoglu, M.D., Ph.D., Baskent University; H. Altay Guvenir Bilkent University
Type of Dataset	Multivariate
Attribute Characteristics	Categorical, Integer, Real
Size (Instances x Attributes)	452 x 280
Train:Test split	70:30 (316 & 136 records)
Total Number of Attributes	279
Number of Attributes considered	15
Total Positive Cases	207
Total Negative Cases	245
Output Values	0 - Class 1, Normal 1 - Class 2-15; various types of Arrhythmia
Name of Attributes	Age, Sex, Height, Weight, QRS duration, P-R interval, Q-T interval, T interval, P interval, Vector angles in degrees on front plane of QRS, T, P, QRST and J, Heart rate
Number of Missing Values	376
SVM w/ RF (Linear, RBF, Polynomial)	73%, 66%, 68%
Linear SVM Accuracy	74%

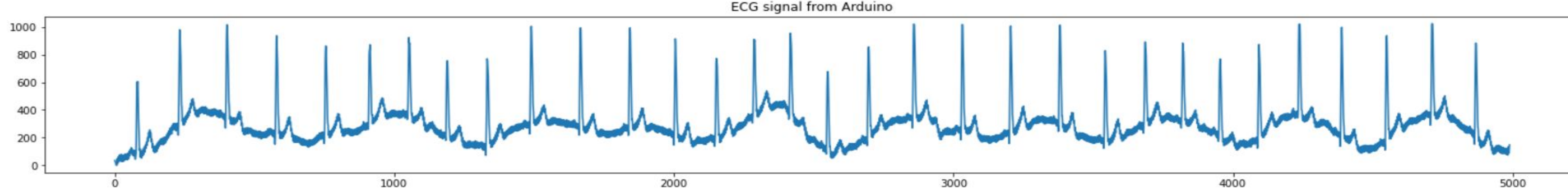
UCI Dataset for Prediction of Cardiovascular Disease

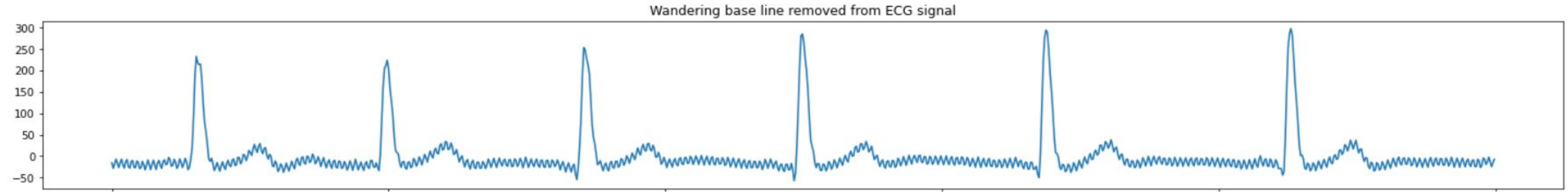
Name of creators	Andras janosi, M.D Hungarian Institute of Cardiology, William Steinbrunn, M.D, University Hospital,Zurich, Switzerland Matthias Pfisterer, M.D, University Hospital,Basel,Switzerland Robert Detrano,M.D.,Ph.D, V.A Medical Center,Long Beach and Cleveland Clinic Foundation
Disease Considered	CVD (CardioVascular Disease)
Type of dataset	Multivariate
Attribute Characteristics	Categorical, Integer, Real
Size (Rows x Cols)	297x14
Train:Test split	75:25 (222 & 75 records)
Number of Attributes Considered	14
Name of Attributes	age, sex, cp, trestbps, chol, fbs, restecg, thalach, exang, oldpeak, slope, ca, thal, num
Total Positive Cases	137
Total Negative Cases	160
Number of Missing Values	0
Output Values	0 - no disease ; 1 - disease
Logistic Regression Accuracy	76.0 %
KNN Accuracy	74.67 %
Support Vector Classifier Accuracy	78.67 %
Random Forest Classifier Accuracy	77.33 %
Decision Tree Classifier Accuracy	70.67 %

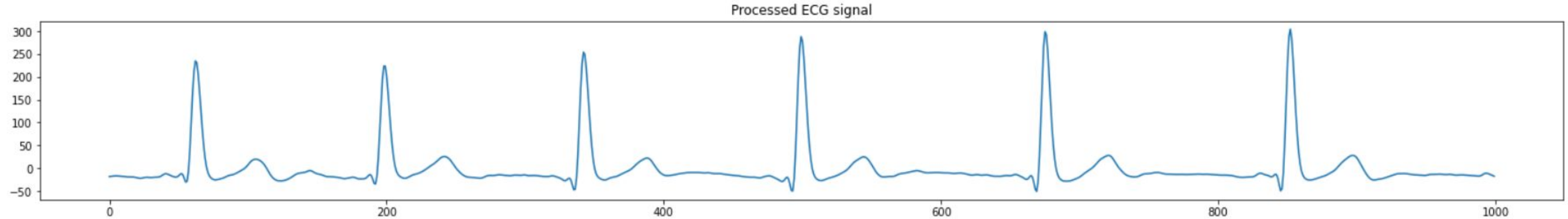
Working Principle / Details

Hospital Dataset for Prediction of Cardiovascular Disease

Disease Considered	CVD (Cardiovascular Disease)
Attribute Characteristics	Integer
Size of Dataset (Rows x Cols)	143 x 17
Train:Test split	70 : 30 (100 & 43 records)
Number of Attributes	16
Name of Attributes	age, gender, height, weight, resting_blood_pressure, cholesterol, blood_sugar, alcohol, cad_history, stroke_history, peripheral_heart_disease, chest_pain_type, exercise_induced_angina, max_heart_rate, st_peak, t_wave_inversion
Total Positive Cases	63
Total Negative Cases	85
Number of Missing Values	0
Output Values	0 - No disease ; 1 - Disease
Logistic Regression Accuracy	93.35 %
Support Vector Classifier Accuracy	90.7 %
Random Forest Classifier Accuracy	95.5 %







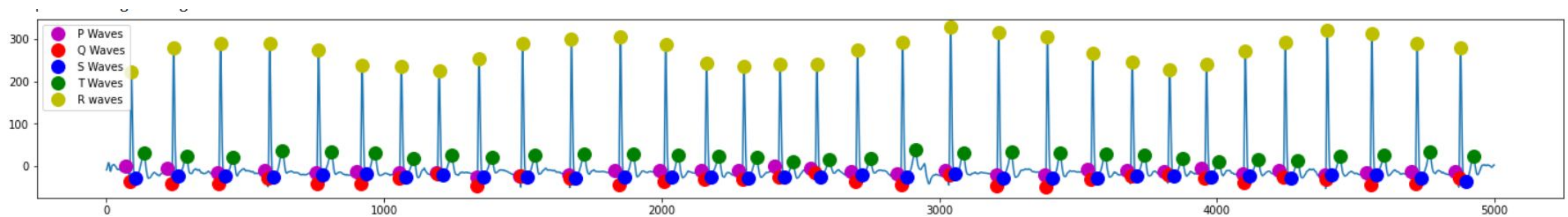


Fig1 : Original signal from sensor
Fig2 : Signal with muscle noise.
Fig3 : Completely Processed signal
Fig4 : Signal with P, Q, R , S and T waves detected
y-axis ECG signal in mV
x-axis count of the ECG signal

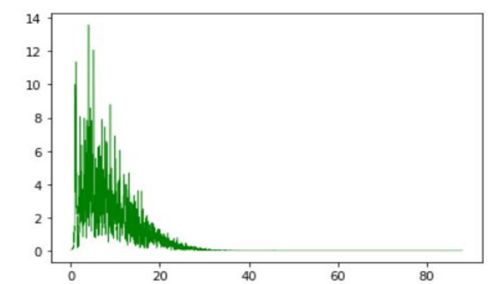
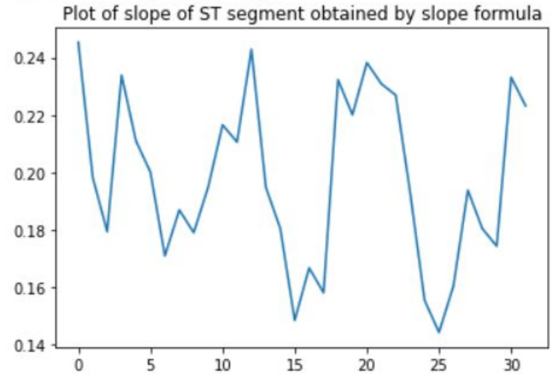


Fig 5: Frequency response of filtered ECG
y axis - Mag of Frequency
x axis - Frequencies

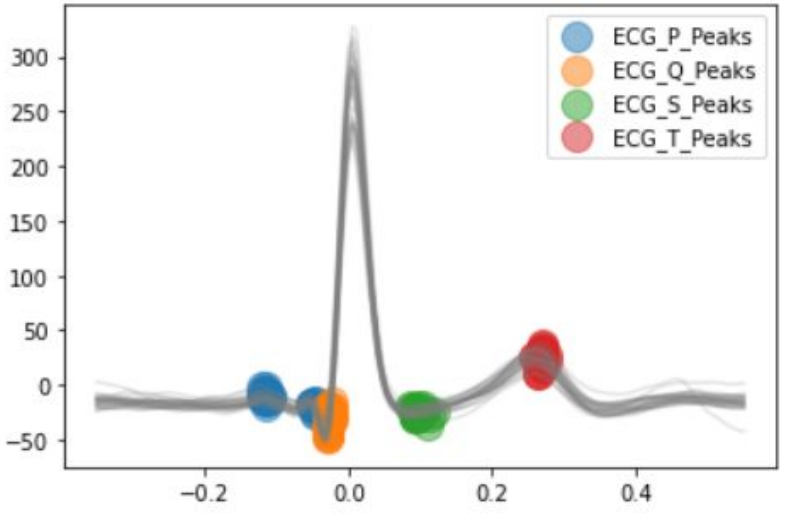
Results

Outputs from ECG

```
Slopes obtained using formula =  [0.2452477  0.19830768 0.17944603 0.23381315 0.21091395 0.20007977
 0.17100005 0.18696656 0.17907661 0.19486838 0.21654041 0.21057328
 0.24281082 0.19480789 0.18050972 0.14858213 0.16680103 0.15818027
 0.23225104 0.22005703 0.2381933  0.23083638 0.22688075 0.19304215
 0.15553912 0.14437222 0.16054441 0.19383145 0.18048518 0.17438015
 0.23307587 0.22317471]
number of slopes in ecg =  32
```




```
mean slope using formula =  0.19766216230330091
mean slope using differentiation =  0.20175104504029476
```




```
number_of_ecg_r_peaks    32
THE SAMPLE WAS COLLECTED FOR 30 SECONDS. TO FIND THE HEART RATE NUMBER OF R PEAKSx2 =  64
No T wave inversion detected
Enter Patient ID : 123123
Enter age : 21
average_heart_rate =  64
```

Fig1 : Graph representing the plot of slopes extracted from ECG.
Fig2 : Delineated ECG Graph and P,Q,S and T waves detected.

Results

Logistic Regression	76.0 %
K - Nearest Neighbour	74.67 %
Support Vector Classifier	78.67 %
Random Forest Classifier	77.33 %
Decision Tree Classifier	70.67 %

Table 4 : Accuracy of different algorithms applied on UCI dataset of CVD

Logistic Regression	93.35 %
Support Vector Classifier	90.7 %
Random Forest Classifier	95.5 %

Table 5 : Accuracy of different algorithms applied on Hospital dataset of CVD

SVM w/ RF (Linear, RBF, Polynomial)	73%, 66%, 68%
Linear SVM Accuracy	74%

Table 6 : Accuracy of different algorithms applied on UCI dataset of Arrhythmia

Results

Cardio Vascular Disease Predictor

At least three-quarters of the world's deaths from CVDs occur in low- and middle-income countries.
Early detection of cardiac diseases and continuous supervision of clinicians can reduce the mortality rate.

This model is running on Support Vector Machine Algorithm and is 78.67% accurate!

Age	Your age..
Sex	----select option----
Chest Pain Type	----select option----
Resting Blood Pressure	A number in range [94-200] mmHg
Serum Cholesterol	A number in range [126-564] mg/dl
Fasting Blood Sugar	----select option----
Resting ECG Results	----select option----
Max Heart Rate	A number in range [71-202] bpm
Exercise-induced Angina	----select option----
ST depression	ST depression, typically in [0-6.2]
slope of the peak exercise ST segment	----select option----
Number of Major vessels	Typically in [0-4]
Thalassemia	----select option----

Predict

Results

Cardio Vascular Disease Predictor

At least three-quarters of the world's deaths from CVDs occur in low- and middle-income countries.
Early detection of cardiac diseases and continuous supervision of clinicians can reduce the mortality rate.

This model is running on Support Vector Machine Algorithm and is 78.67% accurate!

Age	60
Sex	Female
Chest Pain Type	Typical Angina
Resting Blood Pressure	150
Serum Cholesterol	240
Fasting Blood Sugar	Less than 120 mg/dl
Resting ECG Results	Normal
Max Heart Rate	171
Exercise-induced Angina	No
ST depression	0.9
slope of the peak exercise ST segment	Upsloping
Number of Major vessels	0
Thalassemia	Normal

Predict

Cardio Vascular Disease Predictor

This is a Cardio Vascular Disease Predictor based on our Machine Learning Model!
Your results are displayed below! Thank you!

RESULT = You do not have chances of Cardio Vascular Disease!

<u>Output obtained for normal case</u>

Results

Cardio Vascular Disease Predictor

At least three-quarters of the world's deaths from CVDs occur in low- and middle-income countries.
Early detection of cardiac diseases and continuous supervision of clinicians can reduce the mortality rate.

This model is running on Support Vector Machine Algorithm and is 78.67% accurate!

Age	65
Sex	Male
Chest Pain Type	Typical Angina
Resting Blood Pressure	138
Serum Cholesterol	282
Fasting Blood Sugar	Greater than 120 mg/dl
Resting ECG Results	Probable or definite left ventricular hypertrophy
Max Heart Rate	174
Exercise-induced Angina	No
ST depression	1.4
slope of the peak exercise ST segment	Flat
Number of Major vessels	1
Thalassemia	Normal

Predict

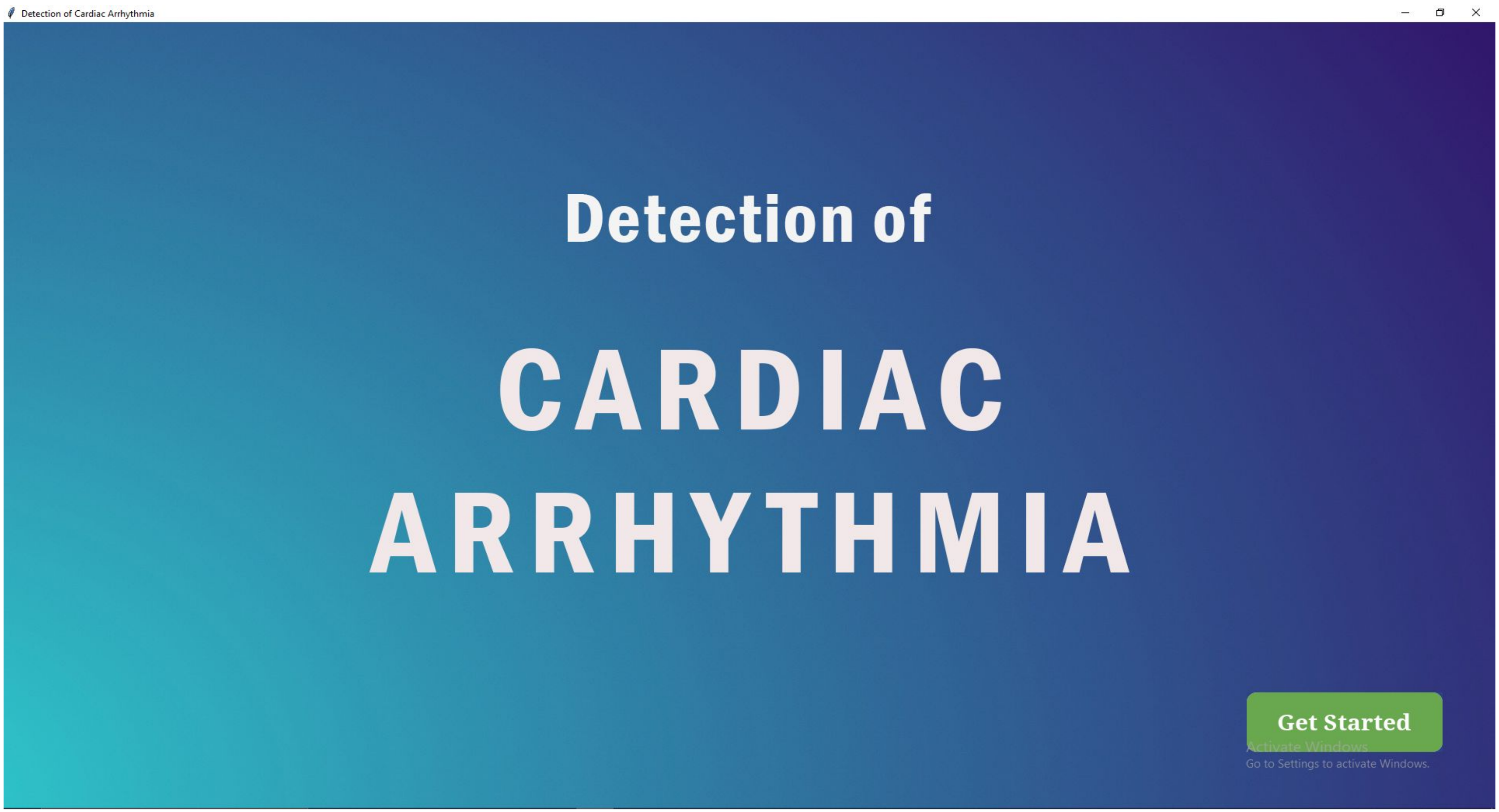
Cardio Vascular Disease Predictor

This is a Cardio Vascular Disease Predictor based on our Machine Learning Model!
Your results are displayed below! Thank you!

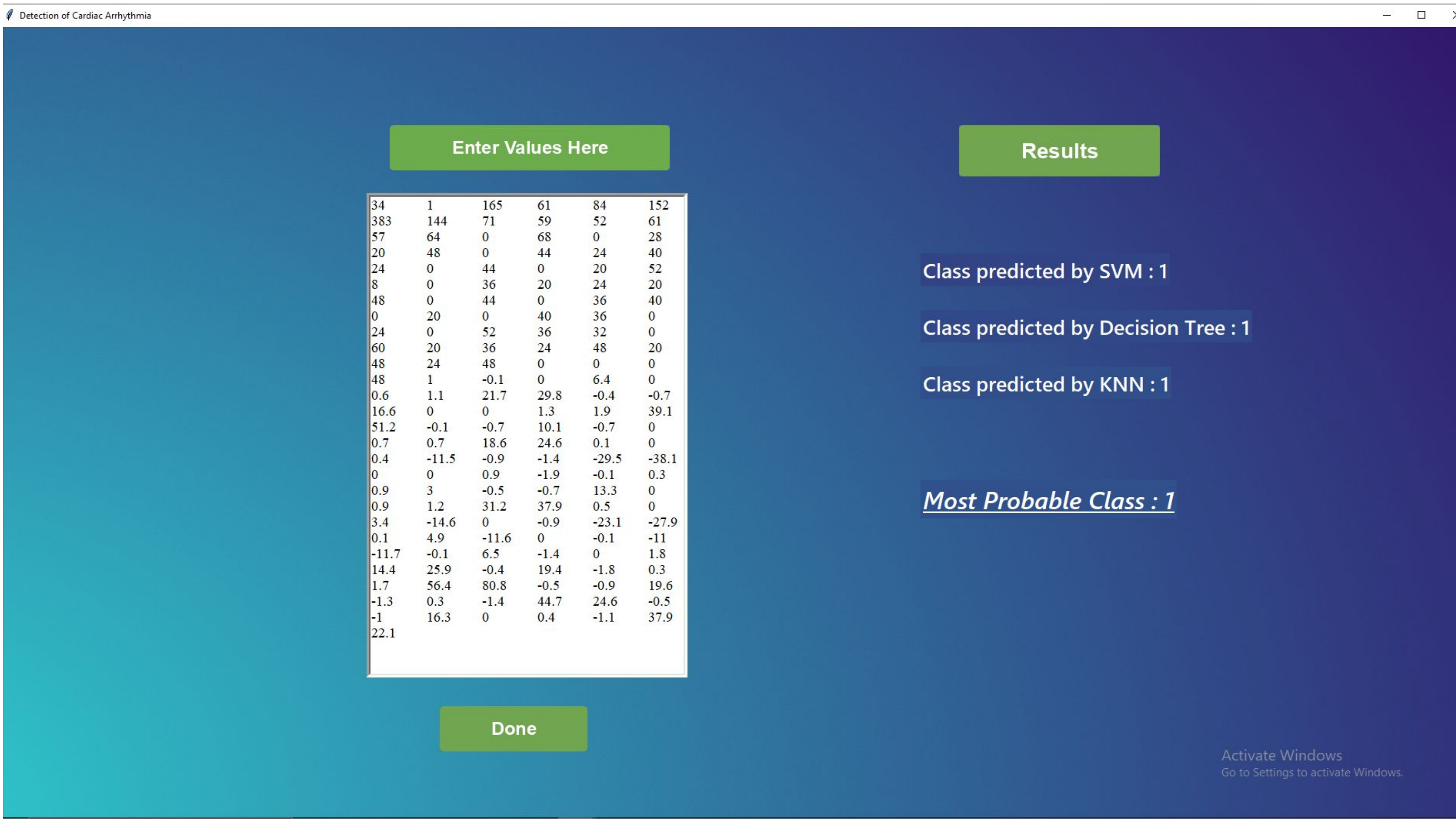
RESULT = Unfortunately, You have chances of Cardio Vascular Disease!

<u>Output obtained for abnormal case</u>

Results

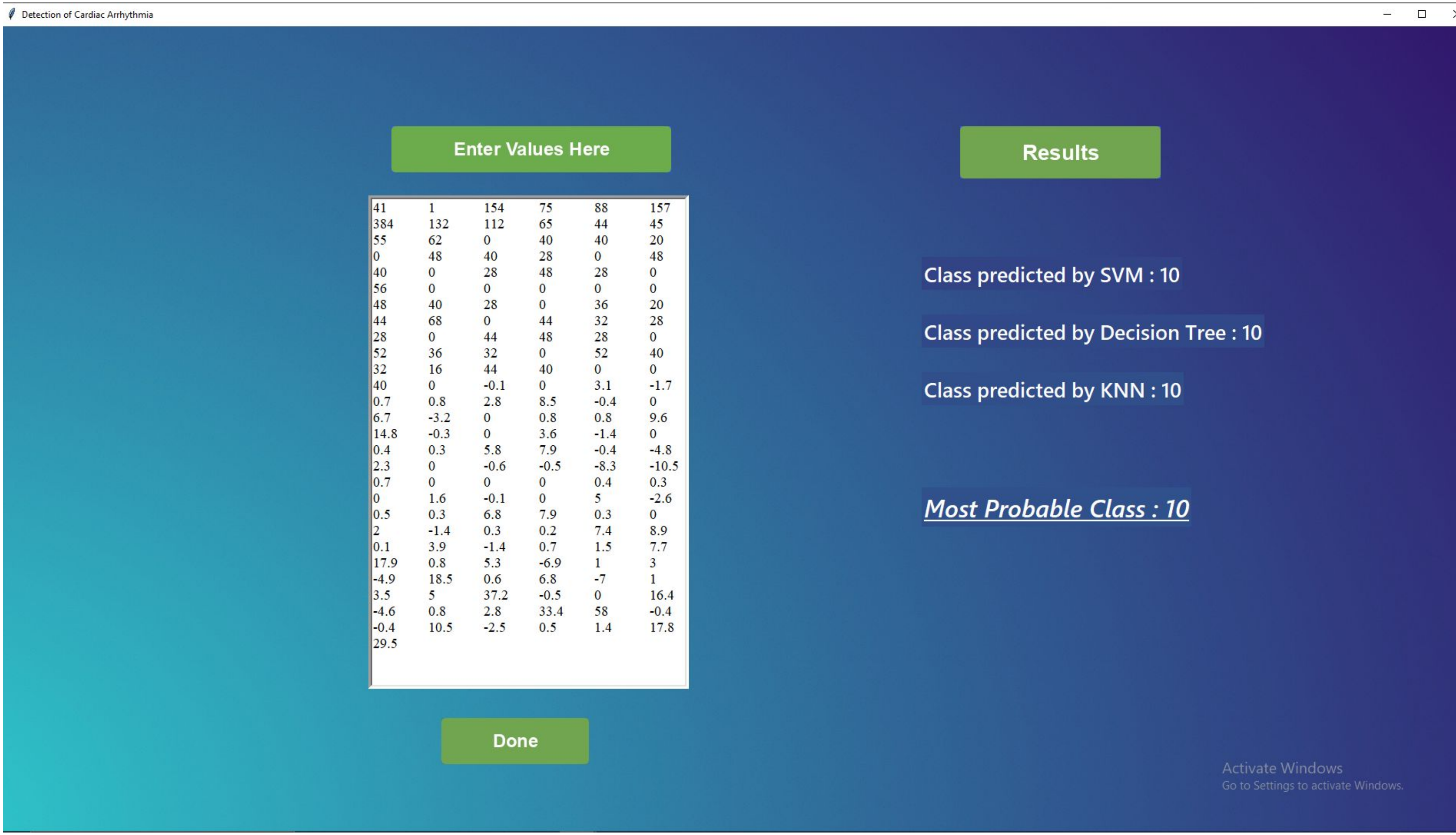

Detection of Cardiac Arrhythmia
Detection of
CARDIAC
ARRHYTHMIA
Get Started
Activate Windows
Go to Settings to activate Windows.

Results



Output obtained for normal case

Results



41	1	154	75	88	157
384	132	112	65	44	45
55	62	0	40	40	20
0	48	40	28	0	48
40	0	28	48	28	0
56	0	0	0	0	0
48	40	28	0	36	20
44	68	0	44	32	28
28	0	44	48	28	0
52	36	32	0	52	40
32	16	44	40	0	0
40	0	-0.1	0	3.1	-1.7
0.7	0.8	2.8	8.5	-0.4	0
6.7	-3.2	0	0.8	0.8	9.6
14.8	-0.3	0	3.6	-1.4	0
0.4	0.3	5.8	7.9	-0.4	-4.8
2.3	0	-0.6	-0.5	-8.3	-10.5
0.7	0	0	0	0.4	0.3
0	1.6	-0.1	0	5	-2.6
0.5	0.3	6.8	7.9	0.3	0
2	-1.4	0.3	0.2	7.4	8.9
0.1	3.9	-1.4	0.7	1.5	7.7
17.9	0.8	5.3	-6.9	1	3
-4.9	18.5	0.6	6.8	-7	1
3.5	5	37.2	-0.5	0	16.4
-4.6	0.8	2.8	33.4	58	-0.4
-0.4	10.5	-2.5	0.5	1.4	17.8
29.5					

Output obtained for abnormal case

Results

Link for UCI Dataset for Prediction of CVD :
https://docs.google.com/spreadsheets/d/1cvw03SwRSCOBlPPAyYez5ZHjwFYAeFkjrwxx9fUTevY/edit#gid=994735256

Link for UCI Dataset for Prediction of Arrhythmia :
https://docs.google.com/spreadsheets/d/1gtaoetSbni4JlRbqDfsOE0TQAmnExDCRbza3hSaIbHI/edit#gid=579262347

Machine Learning Code on Google Colab(both diseases) :

https://colab.research.google.com/drive/1ISF0F27hLLO_a2xy7jO6pGemIvnoEP0j?usp=sharing

Link for Hospital Dataset for Prediction of CVD :
https://docs.google.com/spreadsheets/d/1BM2HUWWK1CdBZih3k-GtHwmJFZU053-1PBb8NHS1FgE/edit#gid=1010088474

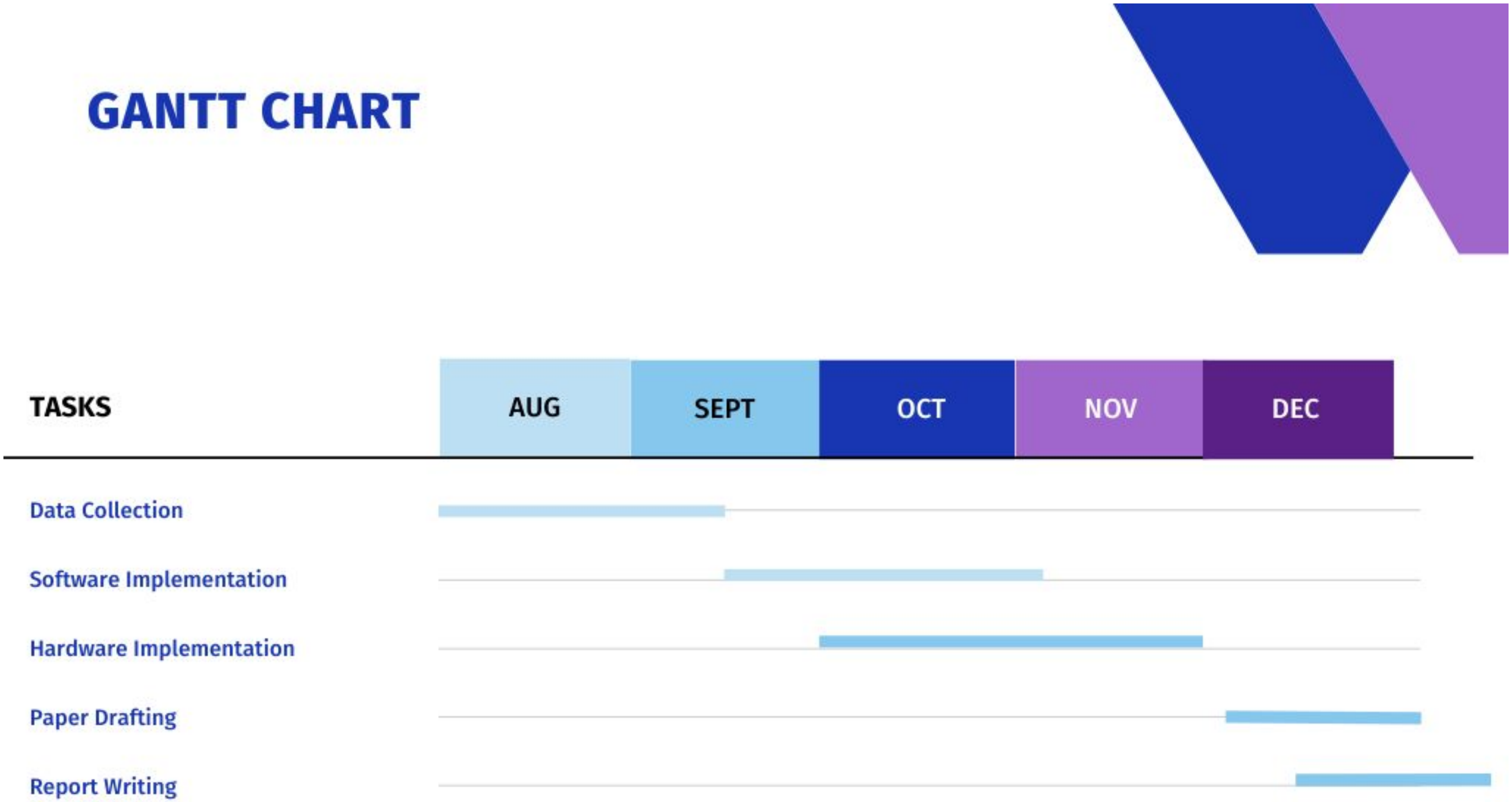
Code for Balancing Hospital Dataset :
https://colab.research.google.com/drive/1pMP4ZYEG2RTq1c90JHivbnW4b_O_3t0a#scrollTo=hhhH1mogHnlE

Code for ECG signal Processing and Feature extraction:
https://colab.research.google.com/drive/1GaSvajv0AaAX5fb0SdN5rp00UhZ8hwYc?usp=sharing

Conclusion

- In this project we have implemented various machine learning algorithm to classify two types of heart disease:- Cardiovascular disease(CVD) and Cardiac Arrhythmia.
- Data acquisition for heart parameters is being implemented using the hardware, signal processing is used to detect peaks in ECG signals/ waveforms, the final output is fed as inputs to the interface along with other important CVD parameters.
- The models built process these inputs received to send back the results to be displayed as the output on the GUI made for it.
- Using Hospital data and UCI datasets we have trained and tested the accuracy of the algorithms used.
- It was observed that SVM algorithm gave the best results for both CVD and Arrhythmia amongst the other algorithms with the accuracy of 74% and 78.67% respectively.
- And, when tested on the hospital dataset, random forest classifier gave the highest accuracy of 95.5%.

Project timeline



References

- Akhila Naz K A, Jeena R S, and Niyas P, "Deep Neural Network based Real Time Multi-class Arrhythmia classification in IOT cloud platform",2021,DOI: https://doi.org/10.21203/rs.3.rs-603991/v1
- Padmavathi Kora, A Rajini, M C Chinnaiah & K Meenakshi, "IoT Based Wearable Monitoring Structure for Detecting Abnormal Heart", 2021 International Conference on Sustainable Energy and Future Electric Transportation (SEFET), 21-23 January 2021 DOI : 10.1109/SeFet48154.2021.9375787
- Y. D P and D. L. N, "Early Detection of Cardiac Arrhythmia Disease using Machine Learning and IoT Technologies," 2021 2nd International Conference on Smart Electronics and Communication (ICOSEC) IEEE, 2021, pp. 1658-1661, doi: 10.1109/ICOSEC51865.2021.9591884.
- Shadman Nashif, Md. Rakib Raihan, Md. Rasedul Islam, Mohammad Hasan Imam, "Heart Disease Detection by Using Machine Learning Algorithms and a Real-Time Cardiovascular Health Monitoring System", Department of Electrical & Electronic Engineering, American International University-Bangladesh (AIUB), Dhaka, Bangladesh, Vol.6 No.4, November 2018

Project ID : S - 13		
Sl No.	**Submission**	**Additional details**
1	Project Report	As per format with Plagiarism report, Project Poster, Draft IEEE format Journal / Conference Paper signed by the guide
2	Project Poster	A4 size colour hard copy
3	Project Video	Shared with Guide
4	Individual Contribution Form	Submitted and signed by each student
5	Project Diary	Submitted to Panel

Q & A