TRIBHUVAN UNIVERSITY

INSTITUTE OF ENGINEERING

PULCHOWK CAMPUS

A THIRD YEAR PROJECT REPORT

ON

ANALYSIS OF ECG USING AI

[CT-455]

SUBMITTED TO:

DEPARTMENT OF ELECTRONICS AND COMPUTER ENGINEERING

SUBMITTED BY:

Rajat Parajuli 073BCT532

Sushant Thapa 073BCT547

Pujan Karki 073BCT549

Pramesh Regmi 073bct550

August, 2019

[Analysis of ECG Using AI]

A THIRD YEAR MINOR PROJECT REPORT

[CT-455]

"A Third Year Report Submitted for Partial Fulfillment of Degree of Bachelors' in Computer Engineering"

SUPERVISOR

Dr. Basanta Joshi

SUBMITTED TO:

DEPARTMENT OF ELECTRONICS AND COMPUTER ENGINEERING

PULCHOWK, LALITPUR

SUBMITTED BY:

Rajat Parajuli	**073BCT532**
Sushant Thapa	**073BCT547**
Pujan Karki	**073BCT549**
Pramesh Regmi	**073BCT550**

August, 2019

ACKNOWLEDGEMENTS

The proof that any student has learnt a particular discipline is through the demonstration of the application of learnt concepts. In particular, the proof that an engineering concept has been solidified is through the quality of the project that the students present.

We are honored to have been given this opportunity to showcase our knowledge and skills through the medium of this minor project. Not only will we be able to master our previously learnt skills but we will also get an amazing opportunity to learn many new things as we go. And on this process, we were never all on our own. We had great help in this project.

We are highly indebted to the Department of Electronics and Computer Engineering and would like to express our gratitude and acknowledgement by presenting a project of the highest quality. We would like to express our deepest thanks to our teachers and faculty members. We would like in particular to thank Dr. Basanta Joshi, our supervisor for his overall guidance. We also express our hearty gratitude to Dr. Arun K. Timalsina and Bibha Sthapit for being all along together with us in all of our difficulties and solving our queries.

Rajat Parajuli (073BCT532)
Sushant Thapa (073BCT547)
Pujan Karki (073BCT549)
Pramesh Regmi (073BCT550)

ABSTRACT

Electrocardiogram (ECG) is a P, QRS and T wave demonstrating the electrical activity of the heart. The cardiac problem where the rate or rhythm of the heartbeat changes i.e. it beats too quickly, too slowly, or with an irregular pattern is Arrhythmia. Since, machine learning has been quite accurate in predicting patterns which is not apparently visible, it can also be used in the classification of arrhythmia in a heart-beat. The main objective of this project, thus is being able to classify the arrhythmia, given the ECG wave, with the highest possible accuracy. The dataset of more than 200,000 heart-beats from MIT-BIH Arrhythmia Database is extracted from signals and converted into single heart-beat using annotations provided using parallel processing and filtering. This data is then fed into machine learning algorithms to classify arrhythmia, with accuracy of 96.61% accuracy. Further, we also present various machine learning algorithms and results used to classify the arrhythmia in a heart-beat.

TABLE OF CONTENTS

Table of Figures

List of Tables

1. INTRODUCTION

1.1 Background

From a simple mathematical calculator to a machine with self-learning ability, computers have been a fantasy of science. The current human civilization is fond of Artificial Intelligence and implementing them in day-to-day life. Not only does this technology share the burden of cumbersome mathematical and computational work, Artificial Intelligence in recent years has been our friend for different intelligent and high-risk tasks. Likewise, Machine Learning has started being used widely in the health sector too. A number of researches is being done regarding implementation of Machine Learning in health sector. One of the best uses of this implementation is analyzing ECG signal which is traditionally done by human medical personnel. Sometimes, the abnormality may go undetected by the human eye and may be associated with stroke, heart failure, and other serious fatalities. We aim to develop an inexpensive means to identify various types of arrhythmia assisting medical personnel with more accuracy and better precision. Our project doesn't intend to completely supersede doctors in this analysis rather it will be assisting medical personnel for more accurate and quicker predictions.

Random forest algorithm was implemented for classification purpose as it is one of the widely used algorithms for similar classification task. It is able to classify the beats with desired accuracy than other short-listed classifiers. The algorithm created the multitude trees with different classes which operate as an ensemble. Each individual tree in the random forest was able to spit out class prediction and the class with the majority votes becomes the final prediction of the algorithm. Except it, we considered Support Vector Machine (SVM) as our alternate classifier which was found to be less accurate than Random Forest classifier. Also, it was a bit impractical for this kind of data. Hence, we finalized Random Forest to be more appropriate in our case.

1.2 Objectives

The primary objective of our project is to develop the supervised batch learning system to classify various arrhythmias seen in human ECG signal. The objectives can be listed as: -

- To develop Machine Learning System implementing available algorithms for precise prediction and classifications.
- To provide easier identification of the abnormal beats using various annotations within the beat.
- To visually graph the predicted result in a graph
- To generate a summary report of the ECG.

1.3 Scope of The Project

ECG analyzer can be implemented to supervise the heart-beat of the patients and predict the risk of various heart beat disorders with the help of analysis. Also, the main scope is to provide easier reading of the ECG with the help of annotations displayed in each beat. The main aim is to improve the accuracy in prediction of arrhythmia saving time and, in the case, where better accuracy is the life-threatening matter.

The input beat is fed to the classifier and the beat is classified on the basis of presence of any disorders along with the notations displayed over the beats. The filter is used to remove the noise present due to AC signal from the Cardiogram reader which ensures the noise free signal and avoids the error due to noise. We emphasized on getting quicker prediction by our algorithms

2. LITERATURE REVIEW

The heart provides electric pulses that ensures the circulatory system of the human body, which is a critical and important one for the human survival. The preceding electrical activation of the heart provides mechanical force to pump the oxygenated blood. Disruption in this orderly pattern of the cardiac excitation can lead to arrhythmias. [1]

Irregularity in heart beat may be harmless or life threatening. Thus, making the accurate detection and classification of arrhythmia important. The electric activity of the heart can be plotted by an Electro Cardiogram (ECG) with extraction of different parameter values. However, finding of minute irregularities can be difficult for a doctor from a long duration of ECG recordings. Thus, machine learning can be effective for automating the detection of arrhythmia. [2]

Among many algorithms that are effective in machine learning, ensemble methods are said to have high accuracy for non-linear data of the ECG waves. Since, Random Forest creates trees drawn at random from a set of possible trees, the accuracy of classification is higher. [3]

Random Forest is one of the ensemble algorithms widely used in machine learning. It is based on the decision tree method. As the base constituents of the ensemble are tree-structured predictors, and each of these trees is constructed randomly, this algorithm is called “random forests.” The tree is grown using CART methodology to maximum size, without pruning. Formally, a random forest is a predictor consisting of a collection of randomized base regression trees. These random trees are combined to form the aggregated regression estimate. [4]

3. SYSTEM REQUIREMENTS AND DESIGN

The data extraction and splitting of a heartbeat series into its constituents are coded in Python programming language. The filtering of data and implementation of the digital notch filter was done in Octave language. All the frontend and Graphics User Interface (GUI) is done in JavaScript and HTML using React framework.

3.1 Software Development Model

Although initially we opted to go for waterfall model, we found it to be unsuitable because of the changing requirements. The change in requirements were due to programming constraints, limited time and availability of required data to train our system. Thus, we opted for spiral model of software development.

Spiral model of software development combines some key aspect of the waterfall model and rapid prototyping methodologies, in an effort to combine advantages of top-down and bottom-up concepts. Due to limited time and resources, we had to start building the software along with analyzing the requirements and available data. The principle of spiral model is such that focus is on risk assessment and on minimizing project risk by breaking a project into smaller segments and providing more ease-of-change during the development process, as well as providing the opportunity to evaluate risks and weigh consideration of project continuation throughout the life cycle.

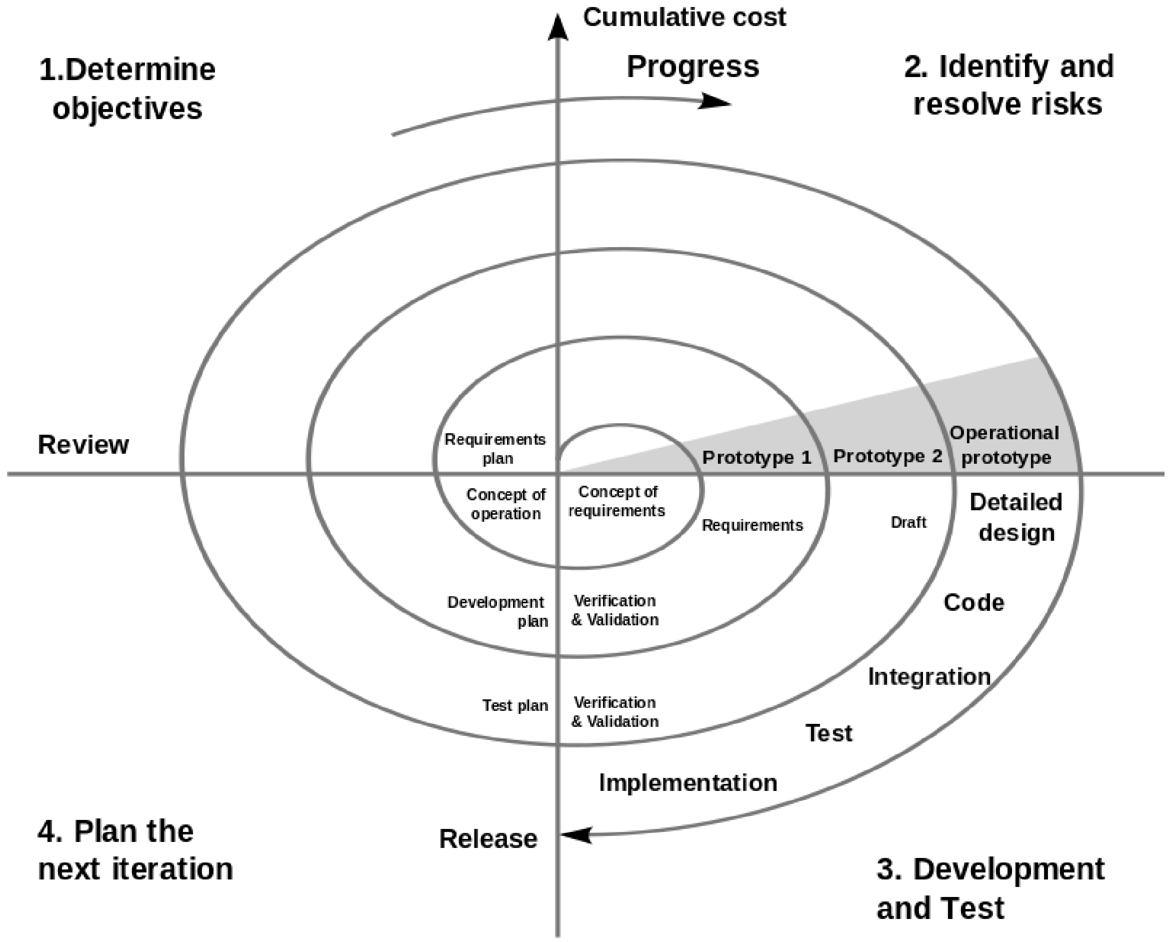

Figure 1 Spiral Model

3.2 Use Case Diagram

Figure 2 Use Case Diagram

3.3 Activity Diagram

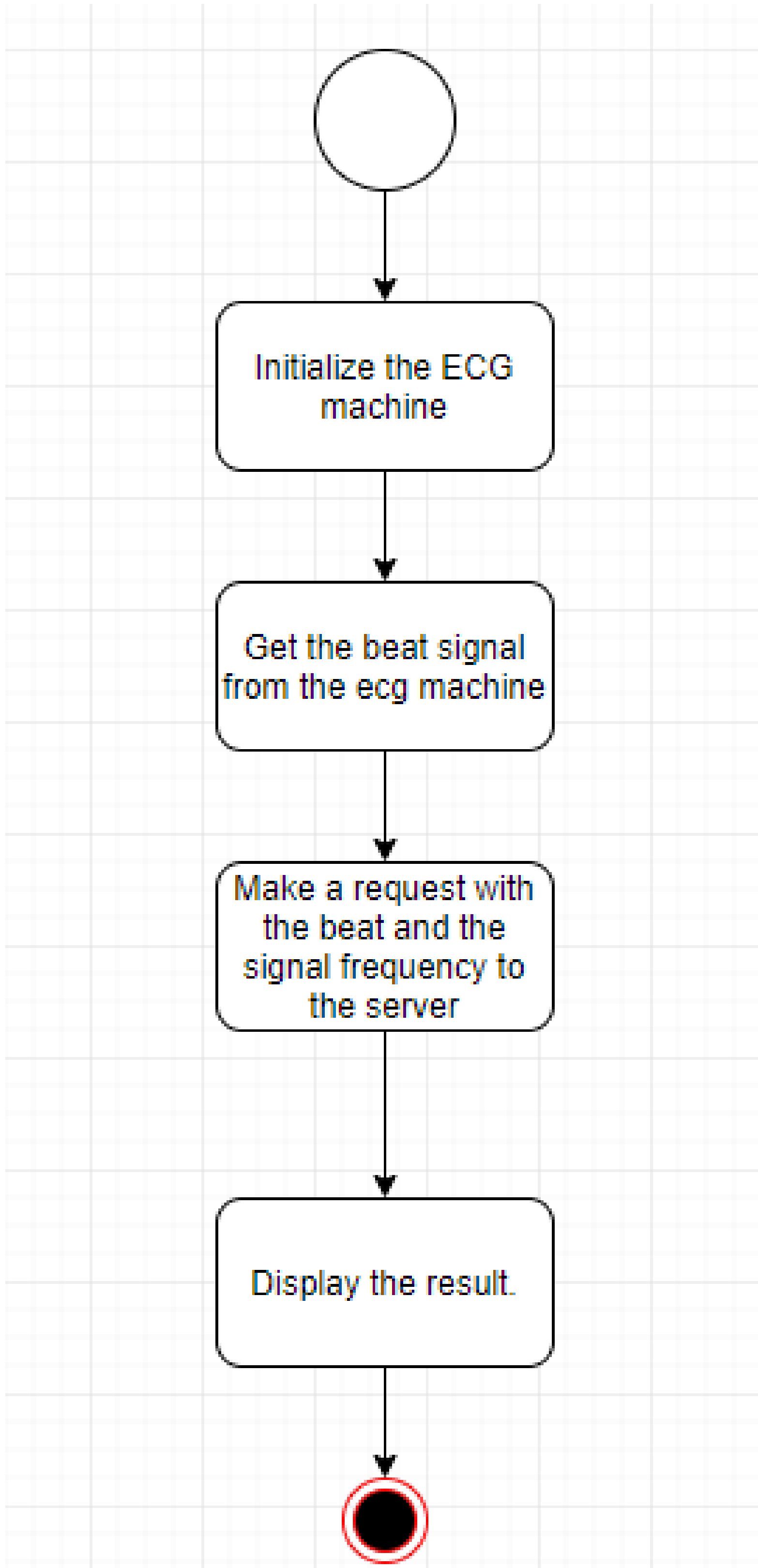

Figure 3 Activity Diagram of the System

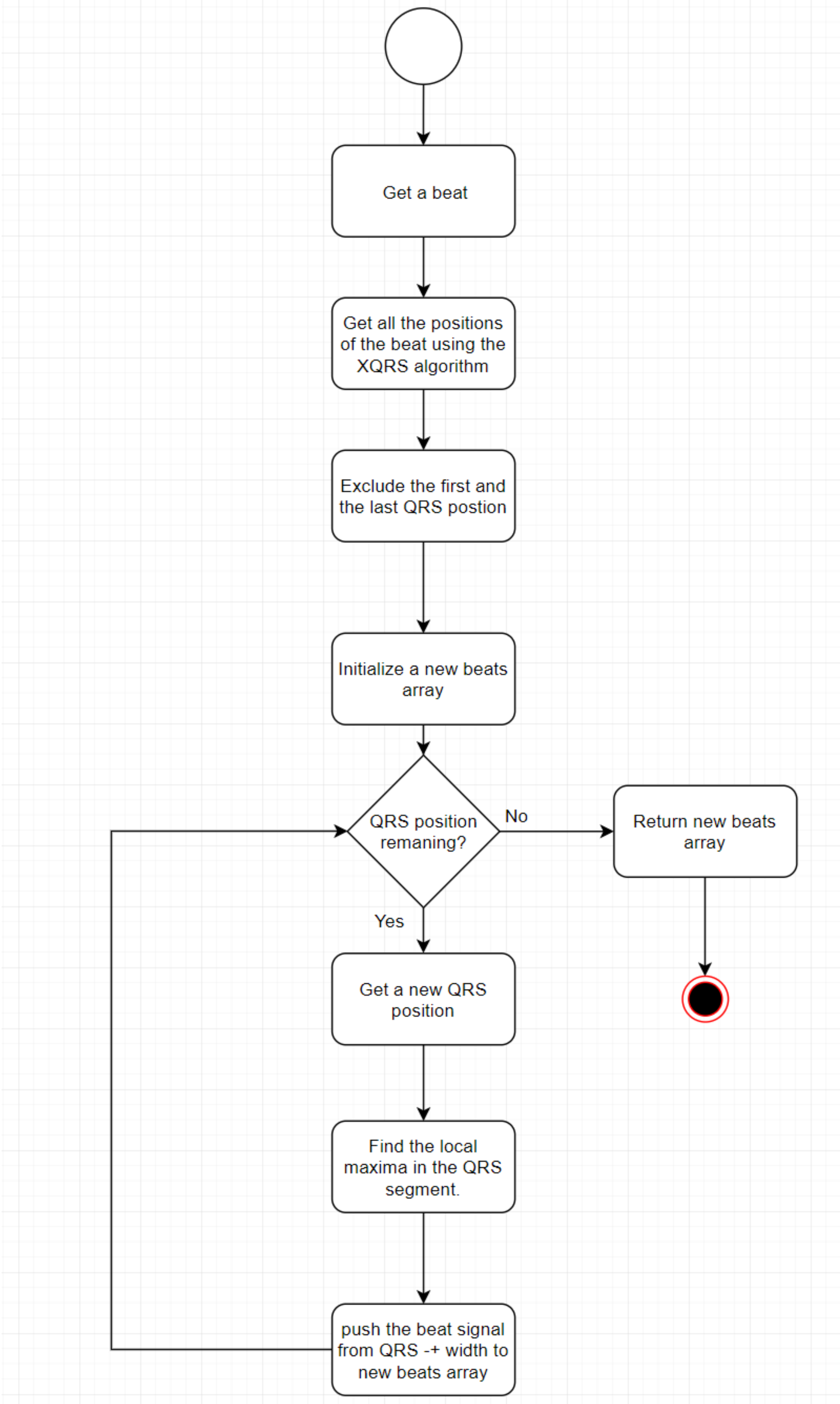

Figure 4 Activity Diagram of System of Splitting Series of Heartbeat into Constituent Beats

4. METHODOLOGY

4.1 Data Source and Extraction of Data

The data source of our project is the MIT-BIH Arrhythmia Database. The source of the ECGs included in the MIT-BIH Arrhythmia Database is a set of over 4000 long-term Holter recordings that were obtained by the Beth Israel Hospital Arrhythmia Laboratory between 1975 and 1979. Approximately 60% of these recordings were obtained from inpatients. The database contains 23 records (numbered from 100 to 124 inclusive with some numbers missing) chosen at random from this set, and 25 records (numbered from 200 to 234 inclusive, again with some numbers missing) selected from the same set to include a variety of rare but clinically important phenomena that would not be well-represented by a small random sample of Holter recordings. Each of the 48 records is slightly over 30 minutes long. The subjects were 25 men aged 32 to 89 years, and 22 women aged 23 to 89 years.

4.2 Conversion of Data

The data of ECG recordings from the MIT-BIT Database are encoded in a waveform database which have the presence of noise and cannot be directly fed into the training system. Thus, the data was converted to a comma-separated-value (.csv) format which could then be filtered as well as trained.

Since there was need for individual beats and the database had many beats of the same person, the data was also split into the constituent beats algorithmically. It was accomplished by locating the R peak of the QRS complex and then recording the data points on either side of the peak. Thus, a series of heartbeats were separated into individual beats, with the information of their arrhythmia embedded in the last row of the comma-separated-value file. The following classes of arrhythmia were provided by the database and which our system is trained to classify.

Symbol	**Meaning**
· *or* N	Normal beat
L	Left bundle branch block beat
R	Right bundle branch block beat
A	Atrial premature beat
a	Aberrated atrial premature beat
J	Nodal (junctional) premature beat
S	Supraventricular premature beat
V	Premature ventricular contraction
F	Fusion of ventricular and normal beat
[	Start of ventricular flutter/fibrillation
!	Ventricular flutter wave
]	End of ventricular flutter/fibrillation
e	Atrial escape beat
j	Nodal (junctional) escape beat
E	Ventricular escape beat
/	Paced beat

f	Fusion of paced and normal beat
x	Non-conducted P-wave (blocked APB)
Q	Unclassifiable beat
|	Isolated QRS-like artifact

Table 1 Labels of Arrhythmia

4.3 Filtering Database to Remove Noise

Whenever there is a digital signal involved, there are chances of noise. When filtering any signal, care should be taken not to alter the desired information in any way.

In the MIT-BIH Database Directory, during digitization of the database, analog outputs of the playback unit were filtered using a passband from 0.1 to 100 Hz relative to real time. The bandpass-filtered signals were digitized at 360 Hz per signal relative to real time using hardware constructed at the MIT Biomedical Engineering Center and at the BIH Biomedical Engineering Laboratory. The database explicitly mentioned that there is noise of 60 Hz in the ECG recordings.

Therefore, to remove the 60Hz noise, we needed to remove the 60Hz noise by digitally implementing a band-stop filter. A notch filter is a band-stop filter with a narrow stopband (high Q factor), which was suitable for our needs, so we implemented the notch filter.

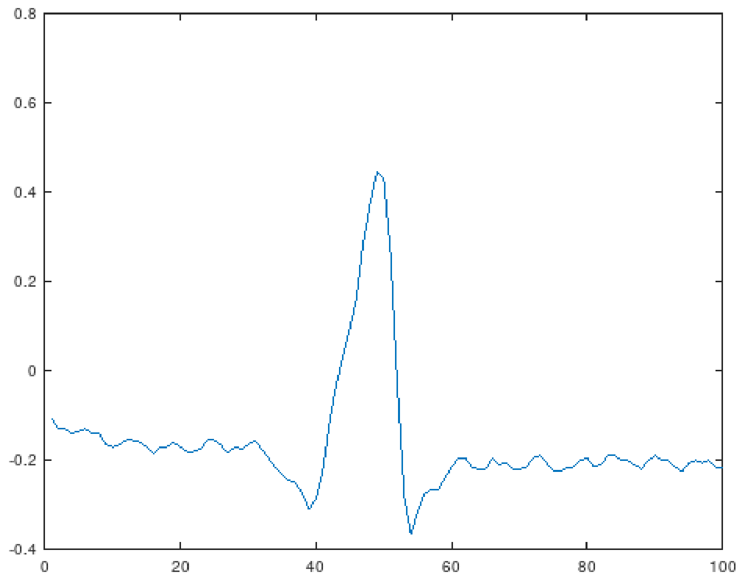

Figure 5 Sample of unfiltered data

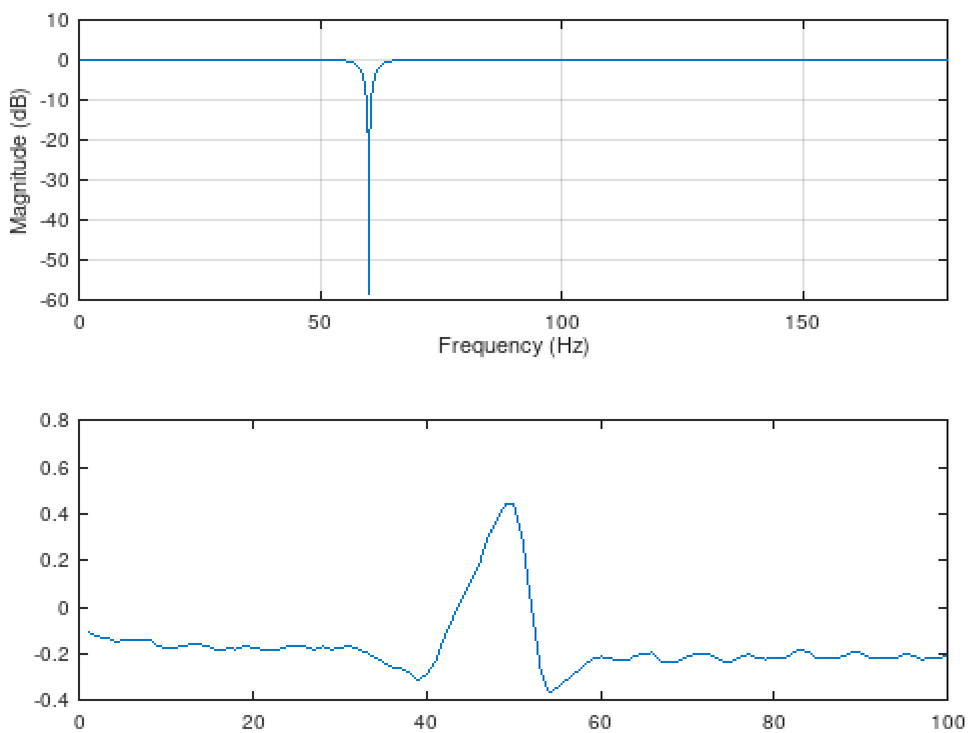

Figure 6 Sample of Filtered Data Using Notch Filter

4.4 Algorithm Used

Random forests, also known as random decision forests, are a popular ensemble method that can be used to build predictive models for both classification and regression problem. Ensemble methods use multiple learning models to gain better predictive results and in the case of a random forest, the model creates an entire forest of random uncorrelated decision trees to arrive at the best possible answer.

The fundamental concept behind random forest is a simple but powerful one — the wisdom of crowds. In data science speak, the reason that the random forest model works so well is:

A large number of relatively uncorrelated models (trees) operating as a committee will outperform any of the individual constituent models.

The reason for this high accuracy from random forest is because the trees protect each other from their individual errors (as long as they don't constantly all err in the same direction).

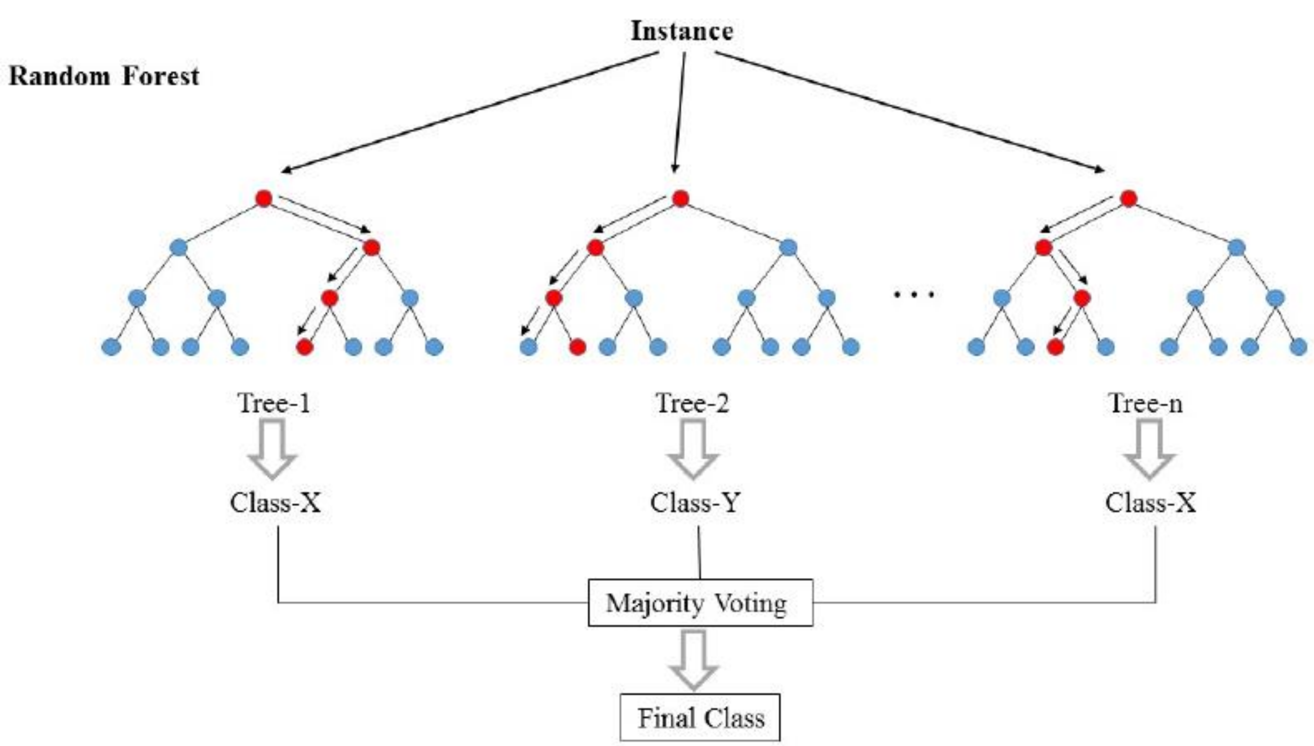

Figure 7 Random Forest

4.5 Handling Client Request

After the training of the system, clients can request their ECG data to be classified. Our system is deployed on a server which can be accessed through an API that we provide so that the front end can be designed on any platform. The way it works is by implementing a queue data structure on the server side, which forwards the client request in a First in First out (FIFO) manner.

The Client data is expected to be in a .csv format and it is analyzed by our system to predict the class of arrhythmia, which can be graphed as well as represented in a summary generated by the system.

5. UML DIAGRAMS

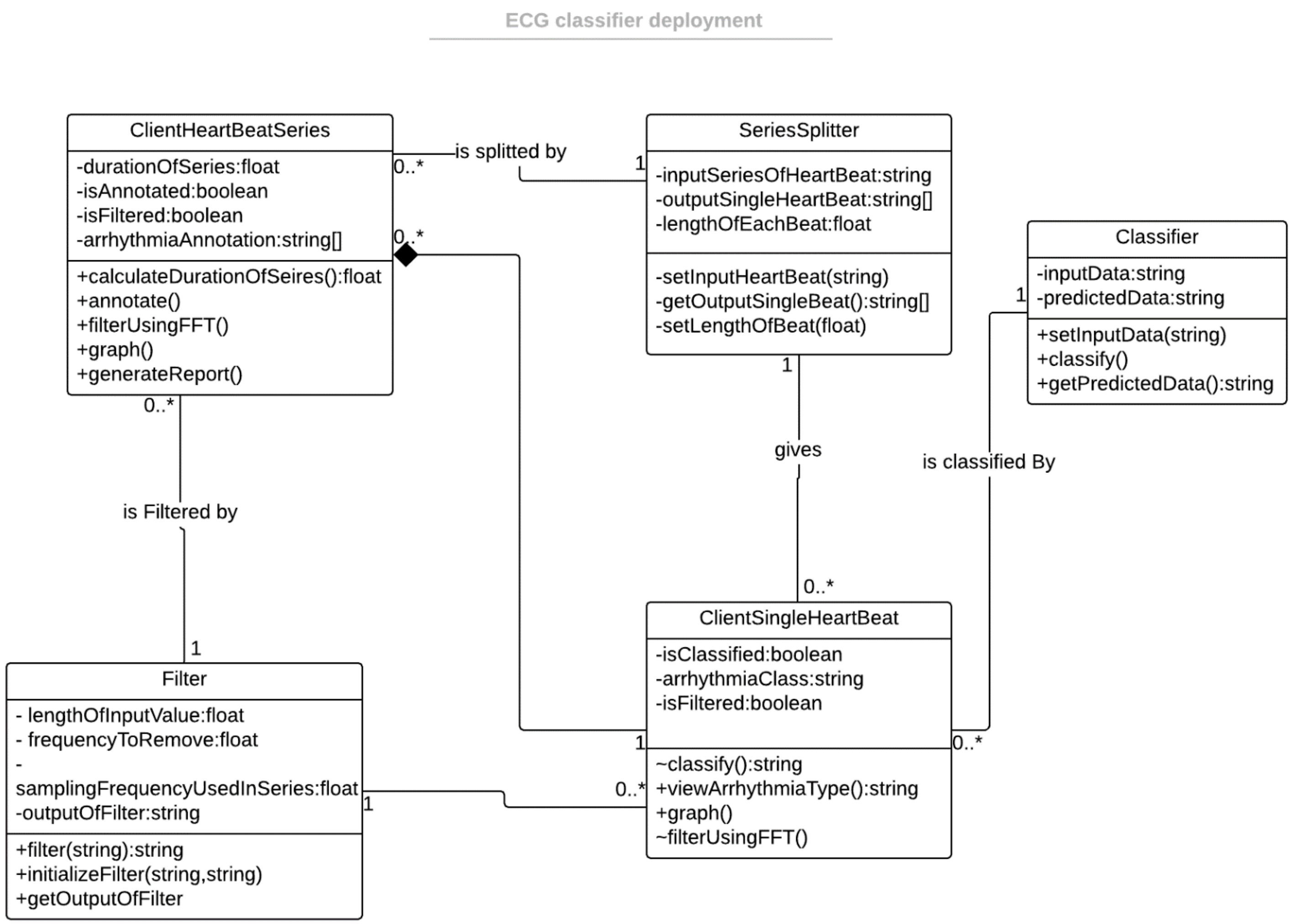

Figure 8 Class Diagram of Deployment System

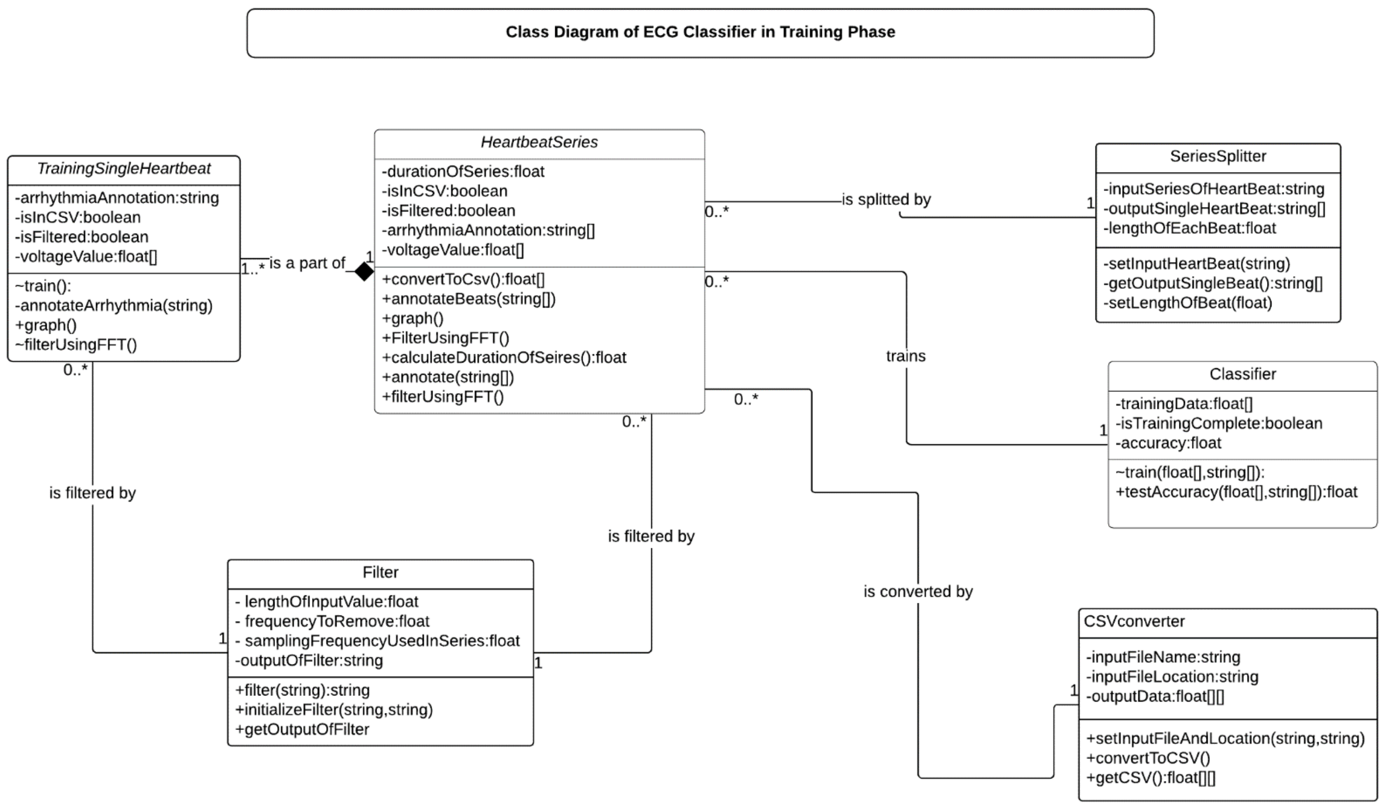

Figure 8 Class Diagram of Training Phase

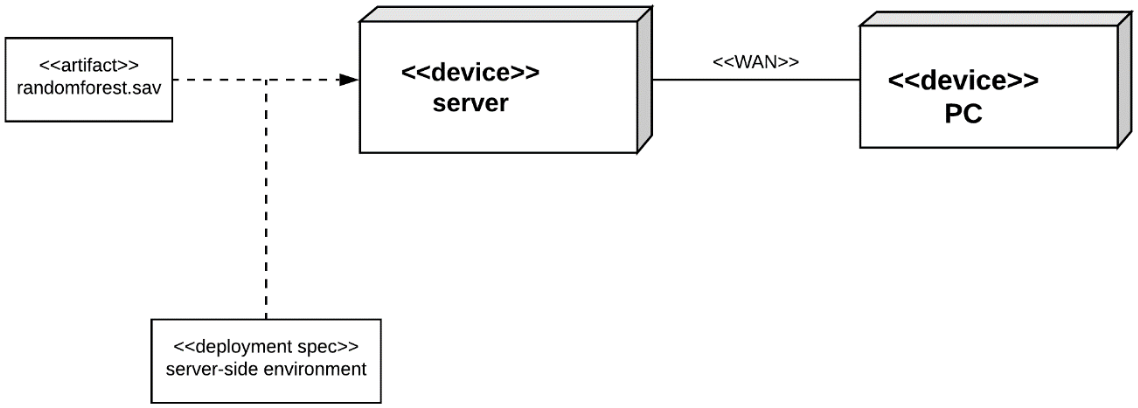

Figure 9 Deployment Diagram of Online System

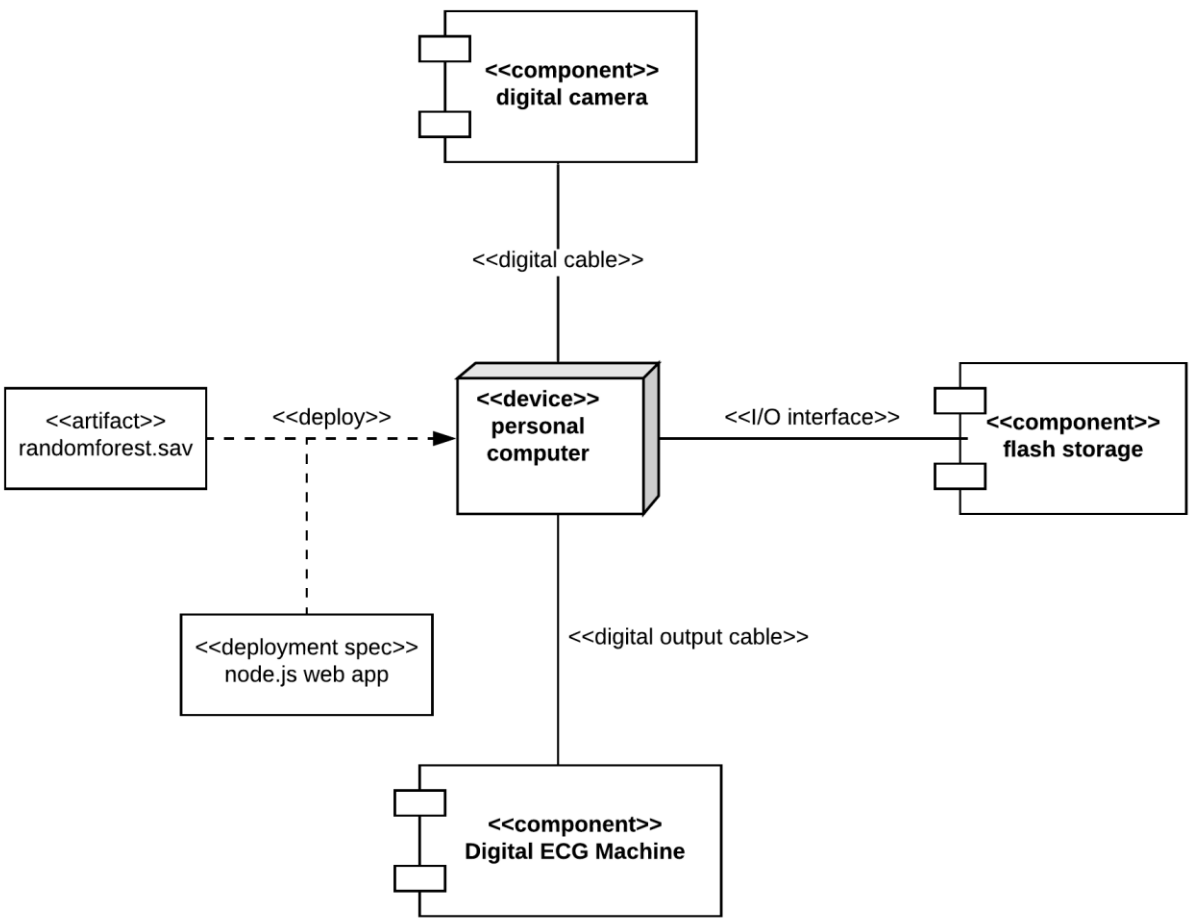

Figure 10 Deployment Diagram of Offline System

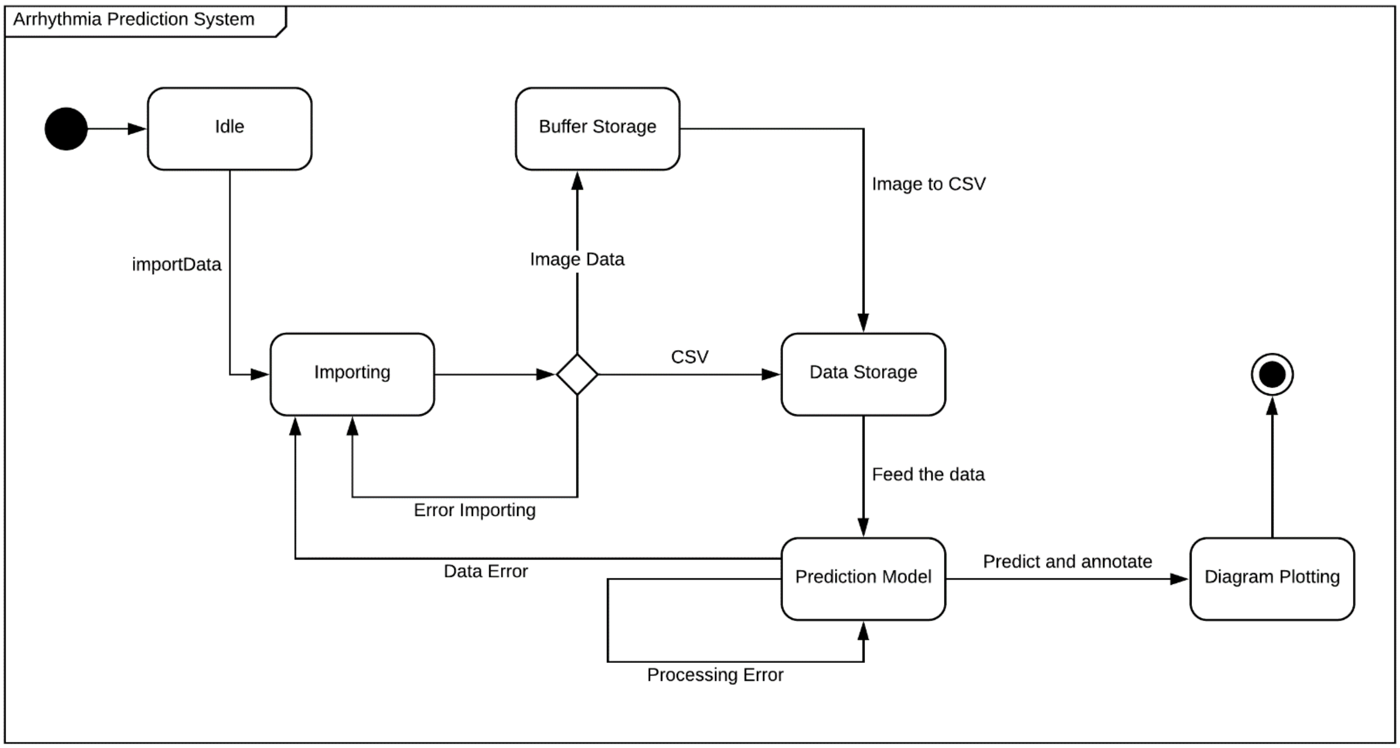

Figure 11 State Diagram

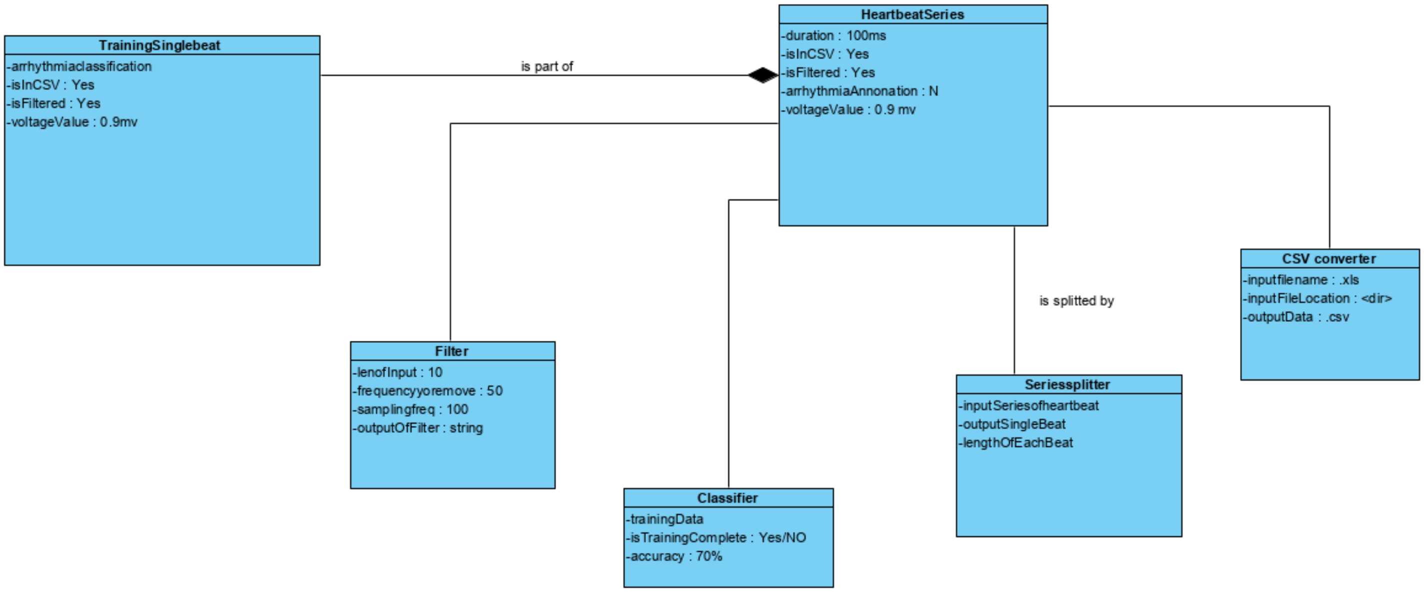

Figure 12 Object Diagram of Training System

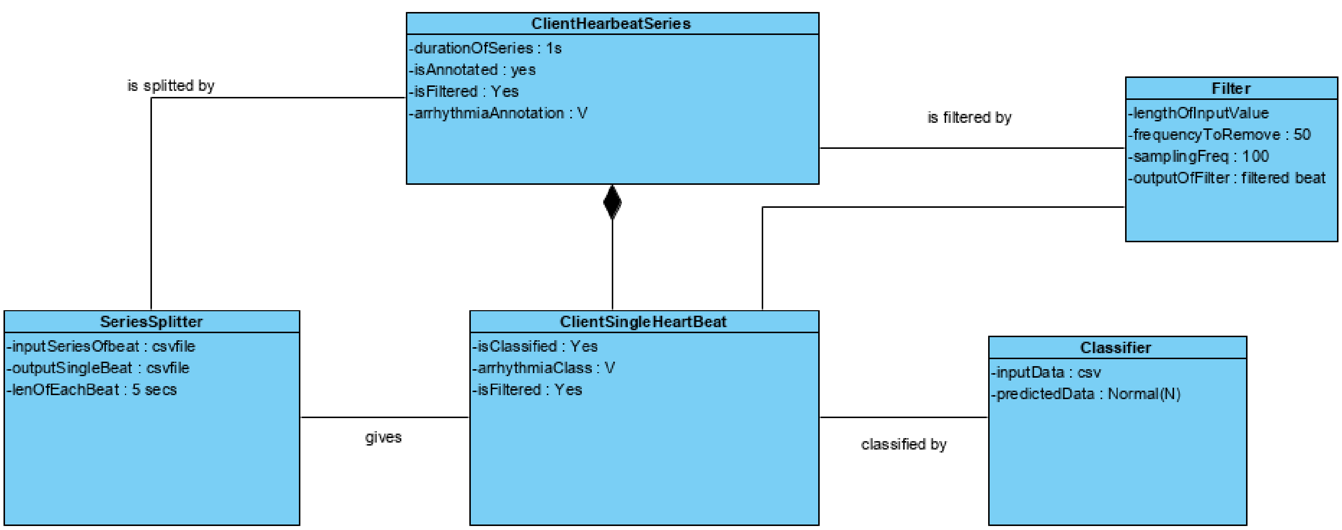

Figure 13 Object Diagram of Deployed System

6. DISCUSSION

Here we designed and implemented a system that classified the various kinds of arrhythmia by training the system on MIT-BIH Arrhythmia Database. We used the random forest algorithm for classification and the ECG data of the client.

The algorithm provided around 96.61% accuracy with the following hyperparameters tuned as given. The grid search was implemented to find out the best hyperparameters value.

n_estimators: The n_estimators parameter specifies the number of trees in the forest of the model. After hyperparameter tuning, the n_estimators was selected as 100.

max_depth: The max_depth parameter specifies the maximum depth of each tree. The default value for max_depth is None, which means that each tree will expand until every leaf is pure. A pure leaf is one where all of the data on the leaf comes from the same class. Max_depth is kept as default.

max_features: The number of features to consider when looking for the best split. It is 8 in this peoject.

bootstrap: Whether bootstrap samples are used when building trees. If false, the whole dataset is used to build each tree. This hyperparameter is kept “False” in our algorithm.

7. OUTPUT

As described earlier, the output of the system can be viewed in the form of a predicted result graph and a summary report generated by the system is shown below:

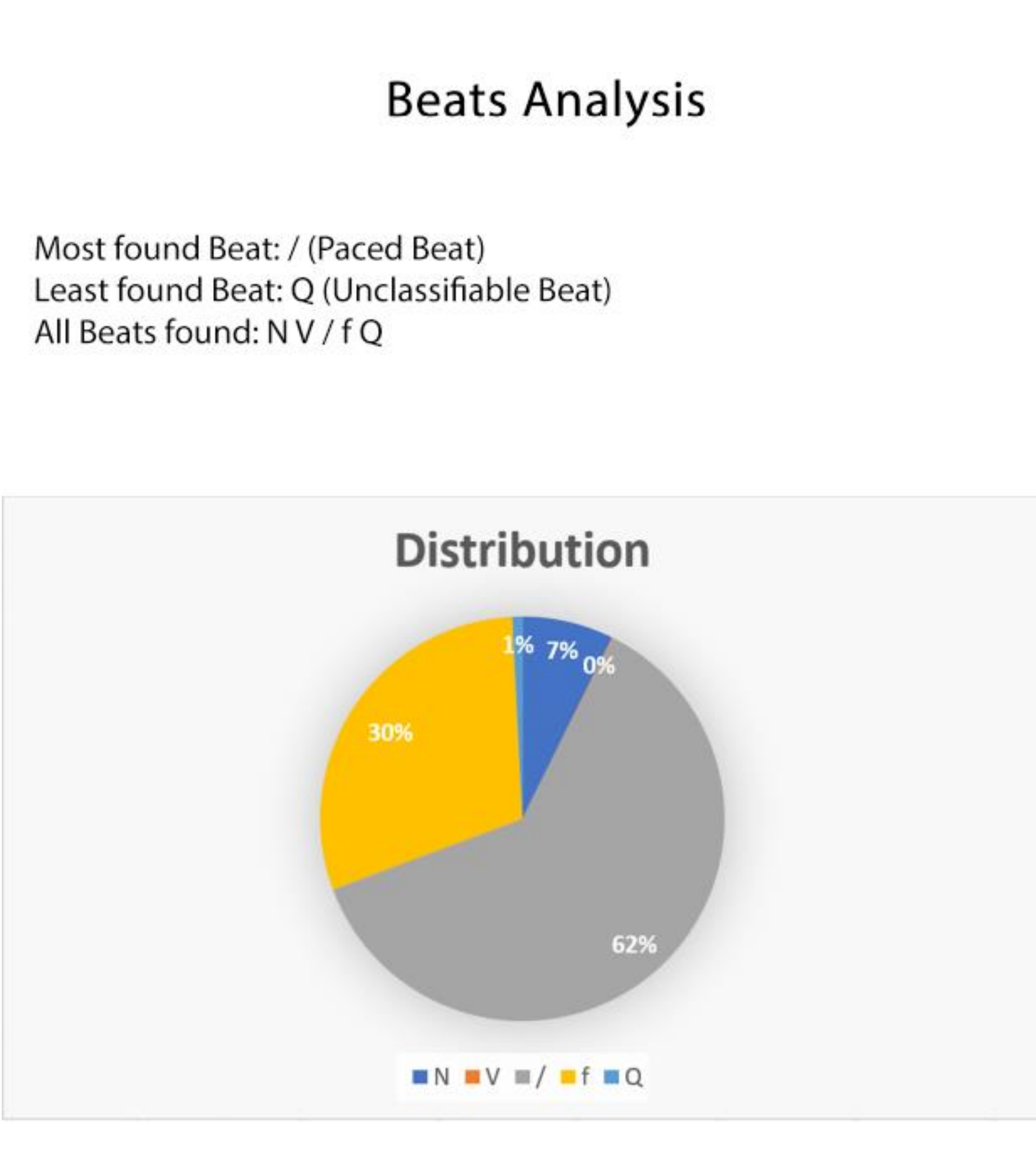

Figure 14 Arrhythmia Report

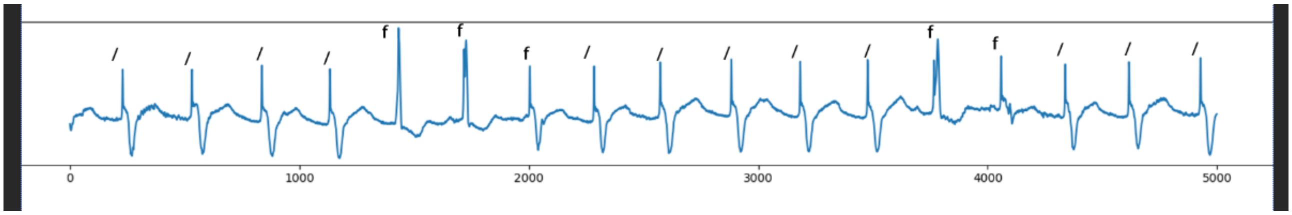

Figure 15 Annotation of Arrhythmia in a Graph

8. CONCLUSION

Machine Learning is a very important tool for pattern detection and can be used to accurately classify a lot of patterns. In our project, we have used Random Forest, an algorithm of Machine Learning to predict the patterns in Electrocardiograms so that we can accurately classify the various kinds of arrhythmia providing great assistance to doctors and other medical personnel in diagnosis of various heart conditions. Our system can classify with an accuracy of 96.61%. Although our system can accurately classify the arrhythmia accurately with graph and report, it is well beyond the scope of our project to interpret the medical consequences of the resultant data.

REFERENCES

[1] G. Tse, "Mechanisms of Cardiac Arrhythmias," Journal of Arryhthmia, 2015.

[2] V. Gupta, S. Srinivasan and S. S. Kudli, "Prediction and Classification of Cardiac Arrhythmia," Stanford University, California, 2014.

[3] R. G. Kumar and D. Y. S. Kumaraswamy, "Investigating Cardiac Arrhythmia in ECG using," International Journal of Computer Applications, 2012.

[4] G. Biau, "Analysis of a Random Forests Model," Journal of Machine Learning Research, Paris, 2012.