Software Documentation

The software “Arrythmia Classification System” is designed with Java thus is cross-platform and hence, runs on Windows, Linux, Mac OSX etc. the complied software product performs preprocessing on ECG comma separated value (CSV) file or simple text file (TXT) files, containing numerical values. Then applies the provided machine learning model (Stacked Random Forest and J48 model) to classify the CSV or TXT file to dragonize the heart condition ECG patient, gives recommendations and stores the data for retraining purposes in the database.

This section will answer the following frequently asked questions (FAQ) on system usage

1. How to launch the software tool
2. Entering inputs in the software
3. Preprocessing the ECG CSV or TXT file
4. Classifying the ECG file (diagnosis)
5. Viewing summary and printing the diagnosis result.
6. Saving the data
7. Why the machine learning model is not hard coded into the application
8. Displaying the machine learning model statistics
9. Denoising raw or noisy ECG data

1. Loading the Application

The software is launched by double clicking on the executable jar in windows or entering the Linux command in the Linux terminal “java -jar ECG Arrythmia Classification System.jar” taking note of the path where the software tool is located.

After launching the application, the login screen is displayed, the default user name and password is admin, admin. Figure 4.1 shows the login interface for the application.

On the login screen, before logging in, check DB Status, if it is connected or not; once logged in, the application interface is displayed. Figure 4.2 shows the application window.

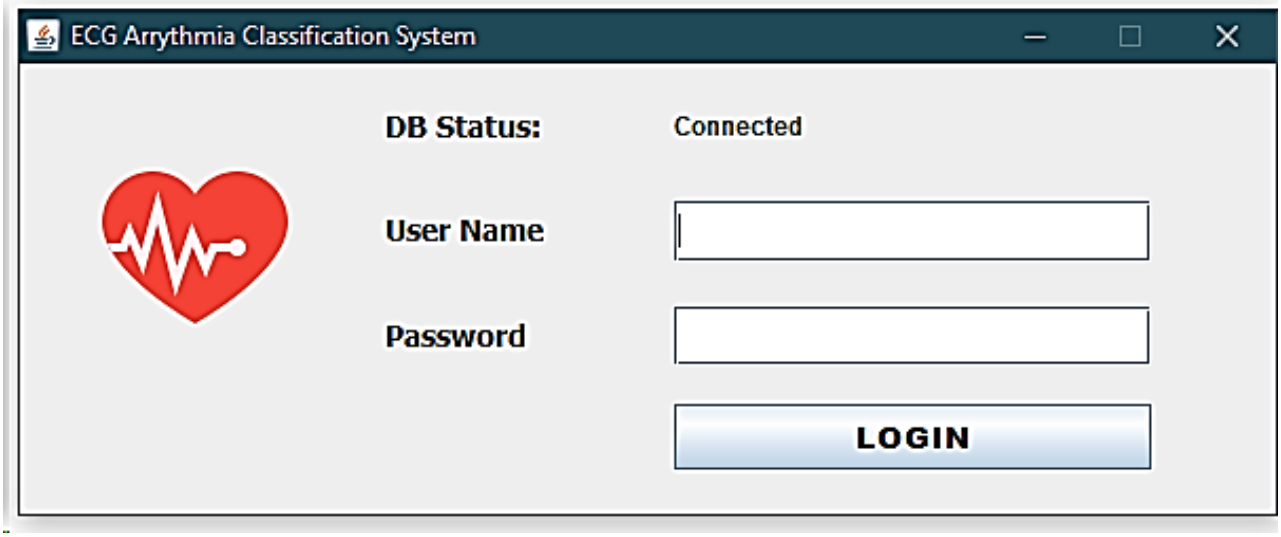

Figure 4.1: Login Page.

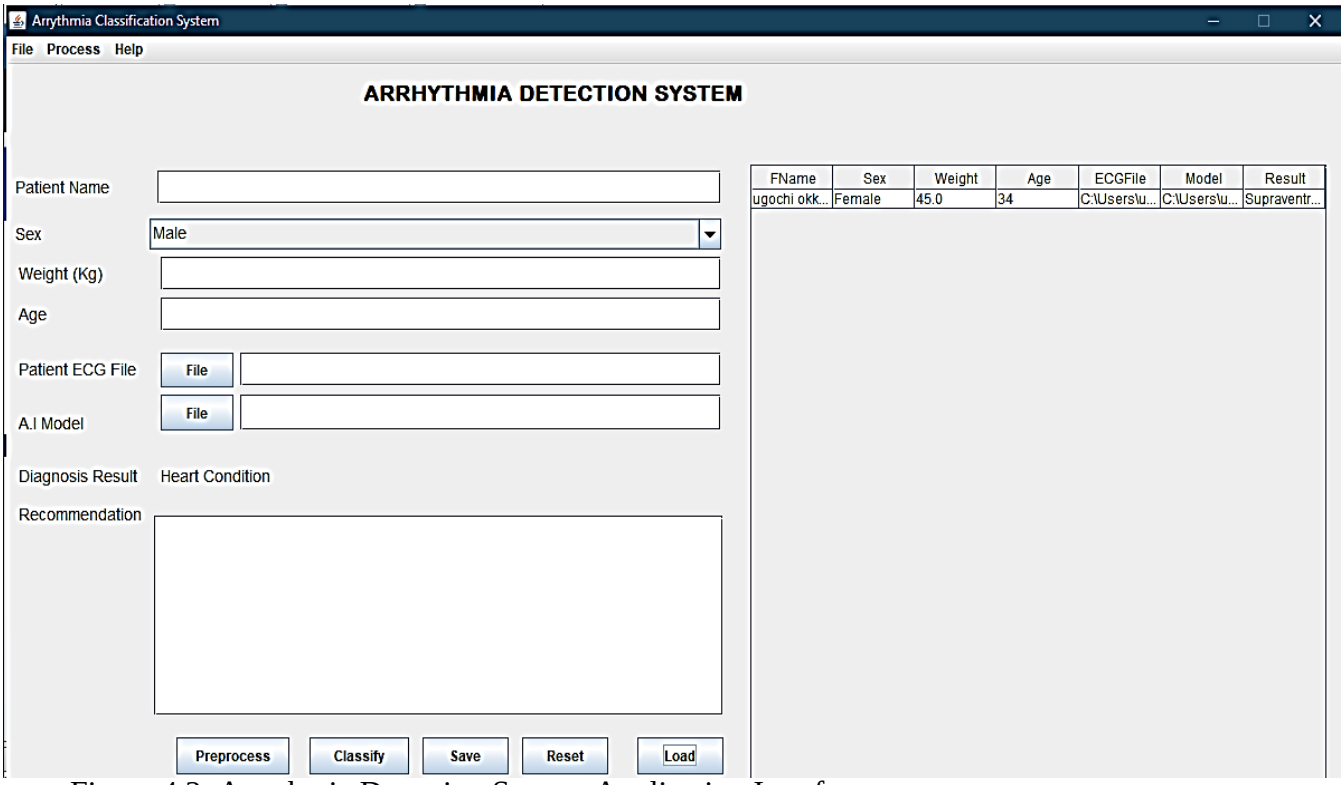

Figure 4.2: Arrythmia Detection System Application Interface.

The application interface shows the title bar, menu bar at the top, patient information and table showing previously stored patient data.

2. Entering Information in the Application

The application has the following fields

Patient Name: The patient full name is here example, James Hope.

Sex: the sex of the patient (male, female, transgender male or transgender female).

The weight and the age of the patient is also taken.

The Patients name, sex, age and gender is not used for the diagnosis, but strictly for record purposes and subsequent data analysis if need be.

The patient ECG file: This the CSV or TXT file, containing raw ECG record of the patient, made up of 188 data instances in a single row, last data instance should be a question mark "?" signifying the class attribute to be predicted. A sample CSV is provided in appendix D.

A.I Model: The trained machine learning model, this is also provided on the folder containing the application.

3. Preprocessing the ECG CSV or TXT File

Preprocessing the ECG CSV or TXT is done by selecting file from the application interface and selecting **Process menu** on the menu bar or the **Preprocess button** on the buttons below the program. Note, without doing this, you won't get any diagnosis result.

When this button is clicked, it prepares a copy of the file by applying the Pan Thompson algorithm, adds labels and converting the copy of the file to ARFF file format which is readable

by the Weka API. Note, If the desired response is not achieved, please click the **Rest button**, then try again.

4. Classifying the ECG File (Diagnosis)

When the necessary information has been inputted into the application, the classify button is selected to perform the machine learning diagnosis. The diagnostic information is provided on the diagnosis result field and recommendation field on the application. As show on Figure 4.3.

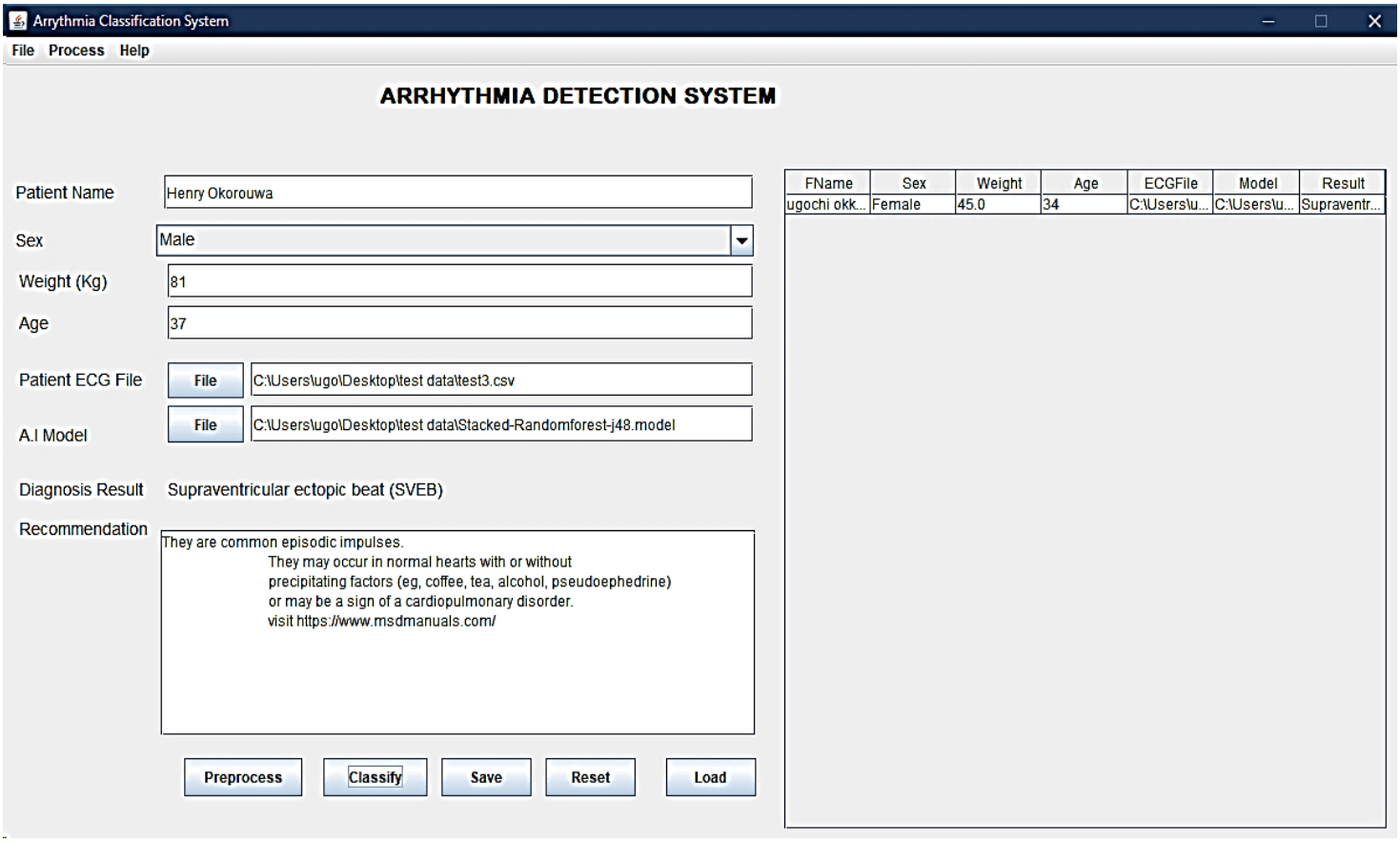

Figure 4.3: Classification Results

5 Viewing Summary and Printing the Diagnosis Information.

After the classification results has been provided, to view the summary and print the diagnosis report, select the **File menu**, on the menu bar on the drop-down menu select **Print Report**. The summary of the report is shown. Then select print to print the information for onward medical examination by a cardiologist. Sample report is shown on figure 4.10.

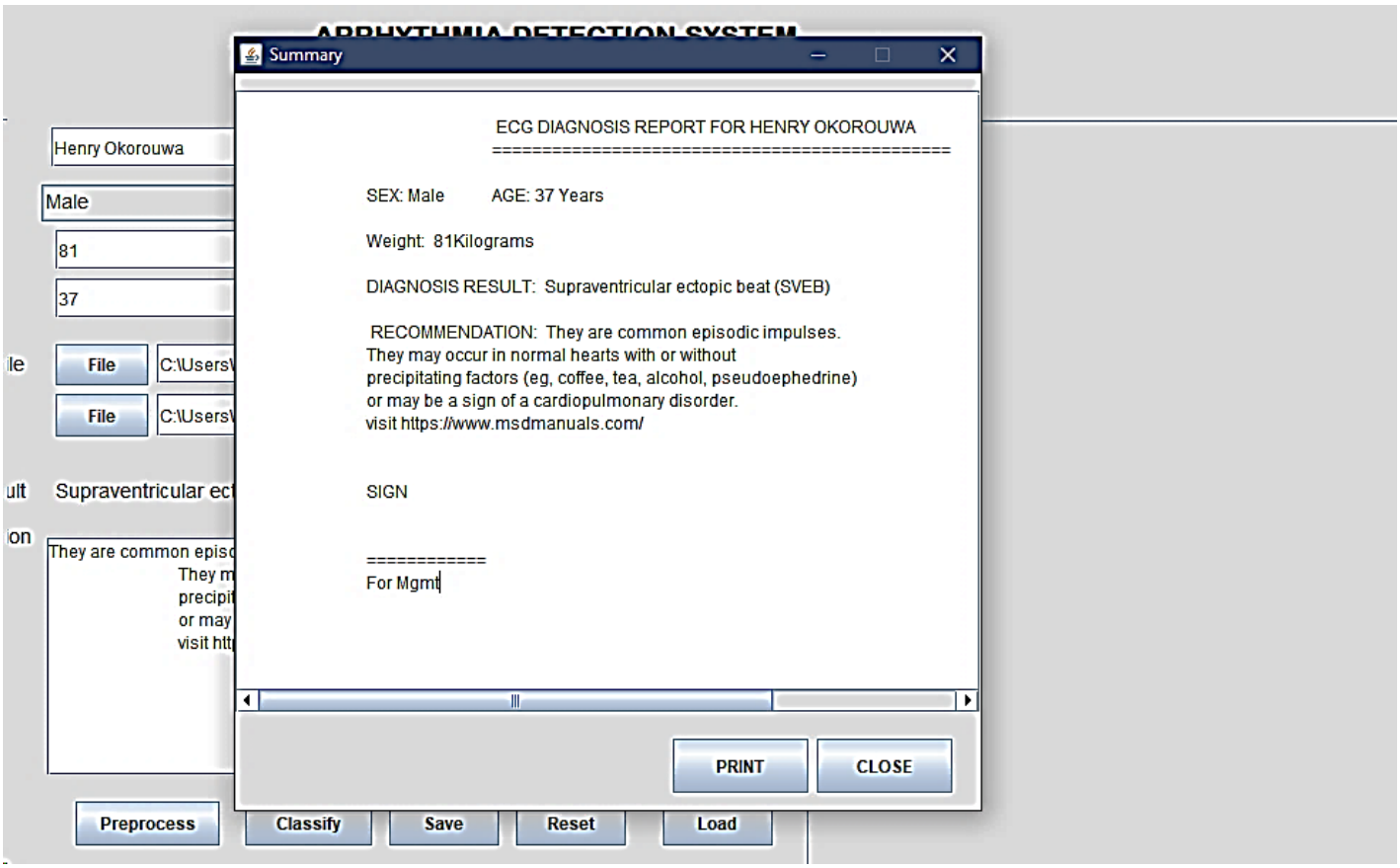

Figure 4.4: Sample Diagnosis Report

6 Saving the Data

After the classification results has been provided, it is also necessary to save the data for reference purposes, hence the save option. To the save the data, click on the **save button**. These saves the information to the SQLite database, click on the **Load button** to view the data saved.

7 Why the Machine Learning Model Is Not Hard Coded into the Application

The machine learning model is not hard coded into the application for four (4) reasons;

a. To increase load time of the application
b. Manage the size of the application since we are considering the edge devices
c. To easily retrain the machine learning model as more data is collected.
d. To try other stacked machine learning models on the system.

4.2.8 Displaying the Machine Learning Model Statistics

To display a summarized model statistic of the A.I Model file inputted, selected the **File Menu**, on the drop-down menu, click on **Model Statistics,** a dialogue box is showing detailing the accuracy, weighted precision, weighted recall, F-Measure etc. of the ML algorithm model**.** An illustration of the output is given in figure 4.5.

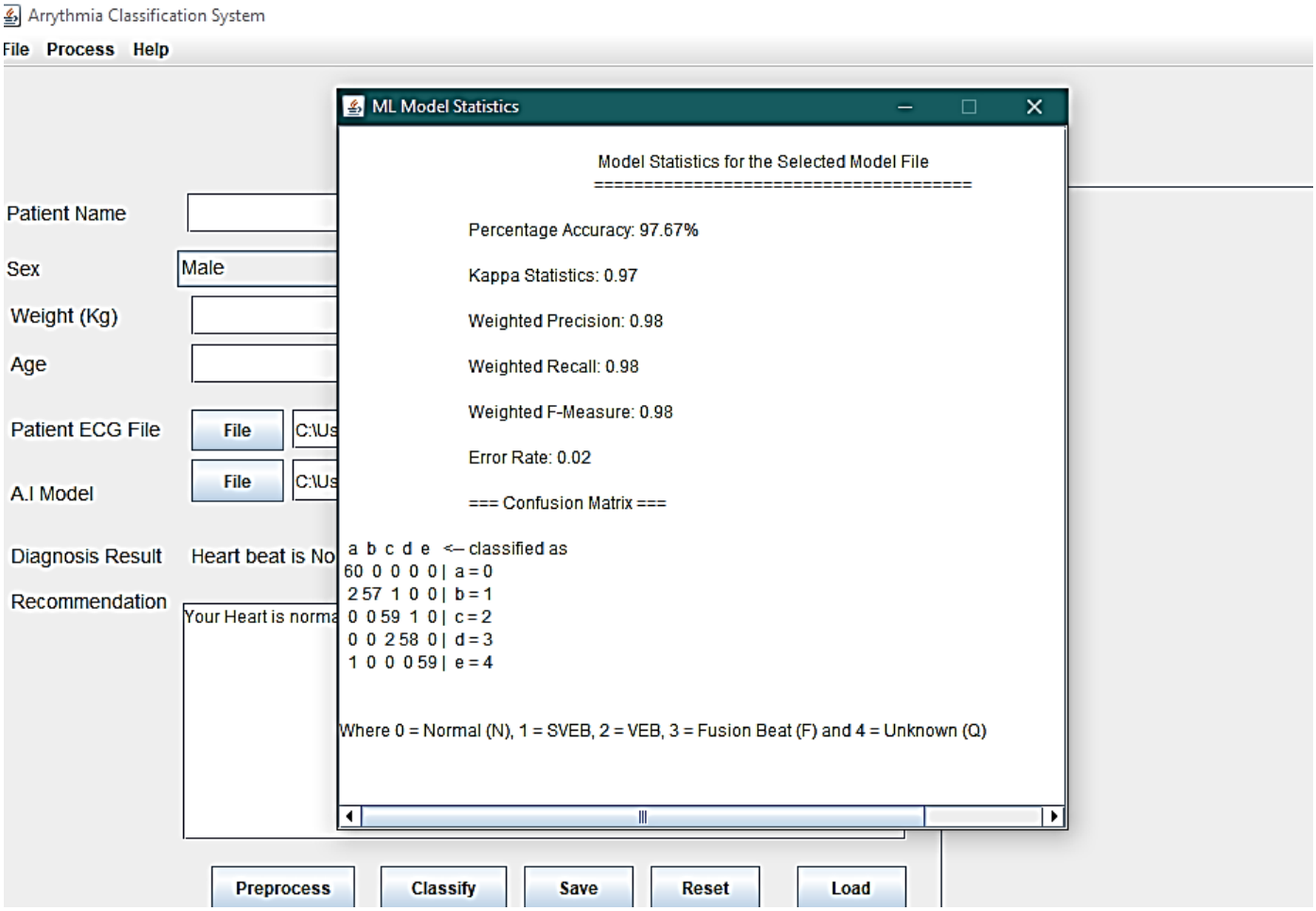

Figure 4.5 Machine Learning Model Statistics

9 Sample Input ECG Data

A sample ECG file in CSV or TXT file uploaded into the system for prediction is given below:

0.732,0.621,0.478,0.339,0.195,0.115,0.033,0.0191,0.0365,0.0852,0.212,0.322,0.33,0.393,0.449,0.513,0.555,0.581,0.597,0.597,0.593,0.617,0.637,0.654,0.652,0.677,0.677,0.697,0.706,0.727,0.744,0.769,0.783,0.809,0.817,0.835,0.847,0.87,0.863,0.868,0.842,0.816,0.776,0.75,0.69,0.656,0.617,0.584,0.557,0.548,0.53,0.536,0.52,0.525,0.518,0.52,0.522,0.523,0.51,0.525,0.529,0.53,0.52,0.515,0.506,0.513,0.513,0.518,0.511,0.506,0.504,0.52,0.515,0.517,0.508,0.522,0.517,0.52,0.513,0.511,0.508,0.517,0.513,0.506,0.511,0.52,0.513,0.522,0.522,0.532,0.525,0.539,0.588,1,0.988,0.87,0.765,0.67,0.522,0.374,0.216,0.104,0.0191,0,0.0209,0.08,0.188,0.315,0.357,0.428,0.477,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,?

10 Denoising Raw or Noisy ECG Data

To denoise a raw ECG data, from the **Process** Menu item on the menu bar, select **denoise** on the drop-down menu, wait for a while for the python Jupyter notebook to launch in a browser, Click Run and enter the file-path to the raw ECG CSV file on the file input field that comes up (this may take a while), please note the format of the file provided as sample, as wrong format will display error while the script runs. After successfully denoising of the ECG file, an output is displayed and the denoised file is saved as data1.csv in the root folder where this application is stored. Edit the file by replacing the last zero in the file with "?" and save, then follow the instructions on item 4.2.4 to carry out diagnosis using the denoised ECG file.

Machine learning Model Results

A machine learning model is a file obtained as an output from a machine learning activity, fundamentally, this file encapsulates a pattern or a relationship that maps the input to the output; models are trained over a given set of data, using a tool and an algorithm, to recognize patterns and reason over those data. The results on table 4.1 and 4.2 were gotten during training and testing of the hybrid machine learning model using Weka machine learning tool. The dataset used consist of 109,446 instances split by 70-30 for training and testing.

Table 1: Summary of Machine Learning Model Results

Correctly Classified Instances	25643	97.63%
Incorrectly Classified Instances	623	2.37%
Kappa statistic	0.92	
Mean absolute error	0.015	
Total Number of Instances	26266	

Table 2: Detailed Accuracy by Class

	TP_Rate	FP_Rate	Precision	Recall	F-Measure	Class
Normal (N)	0.994	0.096	0.98	0.994	0.987	0
SVEB	0.724	0.003	0.887	0.724	0.797	1
VEB	0.905	0.003	0.954	0.905	0.929	2
F	0.689	0.001	0.865	0.689	0.767	3
Q	0.962	0.001	0.987	0.962	0.974	
Average	0.976	0.08	0.976	0.976	0.976	

The Model results showed good results, with an overall accuracy of 97.63%, average truth positive rate, precision and recall value of 97.6% indicates the model is reliable.