Automated Identification of Heart Arrhythmias through HRV Analysis and Machine Learning

S. K. Lawal[1], I. O. Muniru[1], S. A. Yahaya[1], M. O Ibitoye[1]*

[1]Department of Biomedical Engineering, Faculty of Engineering and Technology, University of Ilorin, Ilorin, Nigeria.

ABSTRACT: Sudden cardiac death and arrhythmia are responsible for about 15-20% of cardiovascular disease incidences. Conventionally, the prediction and diagnosis of cardiovascular disorders (CVDs) have been mainly through the evaluation of ECG patterns by cardiologists. To improve the accuracy of and automate this process, and facilitate early detection, Heart Rate Variability (HRV) analysis has been promoted as a diagnostic and predictive tool for CVDs. In the present study, a machine learning model capable of detecting the presence of arrhythmia, using HRV indices obtained from ECG signals was built. Unlike similar works in the literature, this study deployed the developed model on Raspberry Pi with Streamlit software. Two ECG datasets from the Physionet database, one with arrhythmia patients (48 half-hour recordings) and another with healthy individuals (18 24-hour recordings), were employed. An ensemble of seven different machine learning models was used on the two sets of datasets to classify ECG recordings into Arrhythmia and Normal Sinus Rhythm (NSR). The best models were able to predict the presence of Arrhythmia in a 3-minute recording with an accuracy of 95.96%, and in a 10-minute recording with an accuracy of 96.20%. These performance measures were calculated using test dataset. The Random Forest models also had the highest precision, AUC, (Area under the Curve) recall, and F1 scores compared to the other models tested. The highest performing model (i.e., Random Forest Model) was then deployed onto a Raspberry Pi with Streamlit as the software interface for usability. This was done to facilitate a smooth user experience for faster and seamless diagnoses for cardiologists.

KEYWORDS: Cardiovascular disorders; Arrhythmia; ECG; Heart Rate Variability; Diagnosis

[Received Dec. 3, 2023; Revised Jan 11, 2024; Accepted Jan 27, 2024] Print ISSN: 0189-9546 | Online ISSN: 2437-2110

I. INTRODUCTION

The heart is arguably the most important organ in the body, and any problems with the heart or the cardiovascular system must be swiftly investigated and corrected to maintain a healthy life. Unhealthy lifestyles or genetic complications could result in heart defects and disorders which are usually characterized by reduced quality of life and often lead to death. About 15-20% of all deaths worldwide has been accounted for by sudden cardiac death (Srinivasan and Schilling, 2018). It has also been widely observed that ventricular arrhythmias are responsible for 80% of sudden cardiac deaths (Mehra, 2007, Harris and Lysitsas, 2015). Arrhythmias are characterized by irregular heart rhythms caused by disturbances in the electrical activity of the heart (Fletcher and Rea, 2008). Arrhythmias are classified into different types including extra beats, supraventricular tachycardias, ventricular arrhythmias, and bradyarrhythmias" (Lipshultz *et al.*, 2019). One of the most dangerous types of arrhythmias is ventricular fibrillation which is a rapid, life-threatening heart rhythm starting in the ventricles, the bottom chambers of the heart (Handa *et al.*, 2020, Pai, 2011). This may cause an erratic, disorganized firing of signals from the ventricles which severely limits the efficiency of the heart in pumping blood to the other organs and parts of the body (Jalife, 2000).

Prolonged ventricular fibrillation can lead to high blood pressure, loss of consciousness, or even death (Pai, 2011). Arrhythmias can be treated with a variety of methods. This includes lifestyle changes, medication for arrhythmia control and prevention, prescriptions to treat related conditions such as high blood pressure, coronary artery disease, and blood thinners to reduce or manage fast heartbeat. In order for cardiologists to be able to manage patients with arrhythmia effectively, an early and accurate diagnosis is a requirement.

Heart Rate Variability (HRV) is the physiological phenomenon of variation in the time interval between heartbeats (ChuDuc *et al.*, 2013). HRV is measured by the variation in the beat-to-beat interval (or RR interval, "the time elapsed between two successive R-waves of the QRS signal on the Electrocardiogram (ECG) recordings"). In addition, HRV is a measurement of the seemingly insignificant changes in the pace of the heart. It has been shown that reduced HRV is associated with poor cardiovascular outcomes (Tsuji *et al.*, 1996, Connell *et al.*, 2023), as well as the prediction of death after acute myocardial infarction (Kleiger *et al.*, 1987, Tang *et al.*, 2023). HRV indices can be extracted from ECG signals collected using ECG electrodes to monitor the electrical

*Corresponding author: ibitoye.mo@unilorin.edu.ng
doi: http://dx.doi.org/10.4314/njtd.v21i1.2208

activity of the heart. Generally, ECG is a non-invasive approach to detect heartbeats, and measurements are taken from the skin surface through electrode positioning on standard anatomical locations (Werner *et al*., 2016).

The HRV analysis can be done through the analysis of three major domains following ECG signal processing. These domains are a means of measuring and recording the different HRV indices that can be extracted from an ECG signal. Time-domain indices record "the amount of variability in the period between successive heartbeats", also known as the interbeat interval (IBI) (Shaffer and Ginsberg 2017). This consists of parameters such as Standard Deviation of Normal-to-Normal intervals (SDNN), Standard Deviation of R-to-R peak intervals (SDRR), Root mean square of successive RR interval differences (RMSSD), and various other parameters (Shaffer and Ginsberg 2017). The frequency-domain analysis reflects the distribution of power across different frequency bands. The task force of the "European Society of Cardiology and the North American Society of Pacing and Electrophysiology (Haines et al., 2014, Variability., 1996) divided heart rate (HR) oscillations into ultra-low-frequency (ULF), very-low-frequency (VLF), low-frequency (LF), and high-frequency (HF) bands" (Haines *et al*., 2014, Variability., 1996). Non-linear measurements were developed to quantify the dynamical properties of heart rate that the previously mentioned methods were unable to accurately record (Haines *et al*., 2014, Variability, 1996).

Due to their versatility and wide range of applications, machine learning models and techniques have been used extensively in various fields from pattern recognition, computer vision, spacecraft engineering, and computational biology to biomedical and medical applications (El Naqa and Murphy, 2015). In clinical practice, diagnosis is the first step in the treatment of a patient. This is because it is crucial to determine the clinical condition of a patient to guide the next steps to be taken by the clinicians. Machine learning is able to facilitate the process of diagnosis through data collection and analysis. By analyzing the data given to a model, a model can accurately return a prediction or diagnosis results as fast as possible. Thus, making the work of clinicians seamless and faster. Based on this background, as machine learning techniques can help analyze large amounts of data and recognize patterns much faster than humans, it was used in this study. This study aimed to build a machine learning model capable of detecting the presence of arrhythmia using HRV indices obtained from ECG signal recordings. The study also reports the deployment of the model on Raspberry Pi on the Streamlit software platform.

II. MATERIALS AND METHODS

A. Materials

In any machine learning project, deployment is vital to ensuring model usability in real-world applications. Deployment is done after various performance evaluations and testing methods have been carried out to measure functionality. For this study, deployment was achieved in two ways. These were hardware components and a software interface. The hardware component used for this project was a Raspberry Pi (Raspberry Pi Ltd, Cambridge CB4 0DS, UK) which is a line of single-board computers. The Raspberry Pi is used around the world to build hardware projects, for home automation, industrial applications, weather monitoring, robotics, and in other areas of human endeavours. This device consists of several USB ports that can be used to connect the board to a monitor to display various projects it contains. The specific device model used in this study was Raspberry Pi® 3 B+ with ARM Cortex-A53 processor and 1GB RAM capacity (Papakyriakou and Barbounakis, 2023). The Raspberry Pi can be connected to a WiFi network wirelessly or using a LAN cable. It can also be scanned and connected to external devices using Bluetooth. The operating system primarily runs Linux but can be used to run a host of other open-source software (Loyse, 2017). Table 1 presents the cost analysis of the device and other materials used in this study.

Table 1: Cost analysis of the materials used in the study

S/N	Components/ Material	Qty	Rate/(₦)	Amount/ (₦)
1	Raspberry Pi Kit	1	19805	19805
2	Computer Monitor	1	70000	70000
3	Computer Keyboard	1	4900	4900
4	Computer Mouse	1	2500	2500
5	HDMI Cable	1	4000	4000
TOTAL			**₦101,205/(Approx. USD 100)**	

B. System design and description

This research study was divided into three main stages: (i) data collection and preparation, (ii) data pre-processing and HRV extraction, and (iii) model building and validation. The datasets used for this project were obtained from the Physionet database: the MIT-BIH Arrhythmia Database (Moody et al., 2000, Moody and Mark, 2001, Goldberger *et al*., 2000) for ECG recordings of patients with Arrhythmia, and the MIT-BIH Normal Sinus Rhythm Database (Goldberger *et al*., 2000) for ECG recordings of healthy individuals. Snapshots of an arrhythmia recording and a normal sinus rhythm (NSR) recording at one-minute intervals are shown in Figures 1 and 2, respectively.

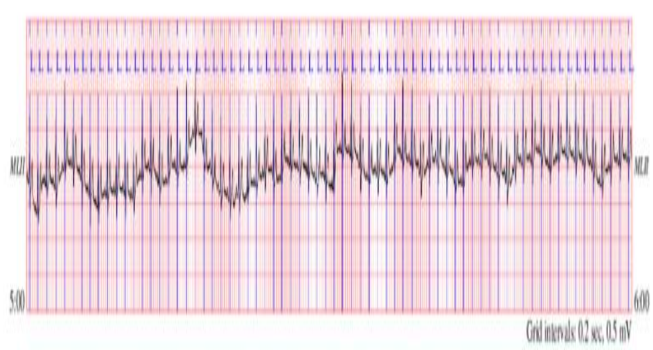

Figure 1: A one-minute interval of an arrhythmia recording from the Physiobank tool.

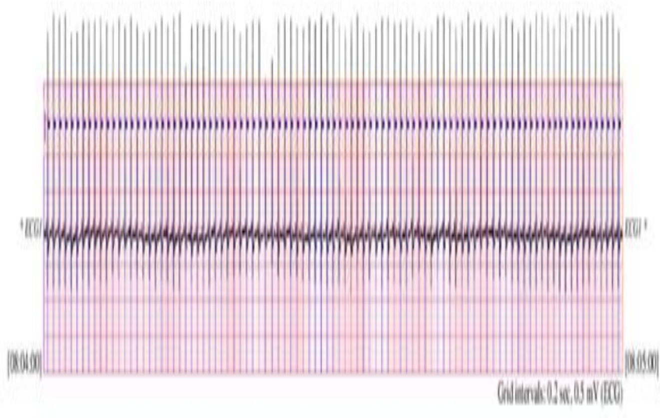

Figure 2: A one-minute interval of a normal sinus rhythm recording from the Physiobank tool.

Access to the recordings in the Physionet Database can be accessed through a toolkit called Physiobank. Physiobank can display recordings as waveforms, show samples as text, show RR intervals, save samples as a CSV file, and others. It also has some controls that can display the recordings in 10-second intervals, 1-minute intervals, 1-hour intervals, etc. Physiobank is a great way of previewing the ECG data to be sure of its contents and to have a feel of what the data looks like.

Physionet data can also be accessed by installing the WFDB (Waveform Database) Software Package, a specialised software designed by Physionet to enable the effective use of Physiobank data.

Using the *rdsamp*, *sampfrom* and *sampto* methods, we collected the digital ECG the recordings from both the Arrhythmia dataset and the NSR dataset and segment into 10-minute portions initially and then 3-minute portions.. .. The segmentation was meant to test the difference in results between 10-minute-long recordings and 3-minute-long recordings, as well as to increase the sample size by reducing the length of each recording. This process was iterated over each record in the two datasets.

The Physionet arrhythmia database contains 48 half-hour segments of two-channel ECG signal recordings. In order to test the results of HRV indices from different recording lengths, the 48 half-hour segments were split into 120 10-minute recordings chosen at random for model training while some recordings were left out to be used in the final testing phase. The Physionet NSR database consists of 24-hour recordings of 18 healthy individuals. After splitting the recordings into 10-minute segments, 104 recordings were selected at random for testing and HRV extraction, resulting in a total of 224 recordings as shown in Figure 3. A similar process was performed in the second part with the recordings now split into 3-minute segments. 450 recordings were then selected at random from the arrhythmia portion while 900 recordings were selected from the NSR portion for testing and HRV extraction, resulting in a total of 1350 recordings as shown in Figure 4.

NeuroKit2 is a user-friendly package that provides easy access to advanced biosignal processing routines (Makowski *et al*., 2021). It is widely known for its simplicity in use allowing researchers and clinicians to analyse physiological data without a lot of programming or biomedical signal processing experience. NeuroKit2 was used in this project for data pre-processing (i.e., signal resampling and signal cleaning) and HRV extraction. Table 2 summarises the features extracted using NeuroKit2.

Table 2. HRV indices extracted from Neurokit

S/N	Domains	Indices
1	Time Domain	'MeanNN', 'SDNN', 'SDANN1', 'SDNNI1', 'SDANN2', 'SDNNI2', 'SDANN5', 'SDNNI5', 'RMSSD', 'SDSD', 'CVNN', 'CVSD', 'MedianNN', 'MadNN', 'MCVNN', 'IQRNN', 'pNN50', 'pNN20', 'HTI', 'TINN'
2	Frequency Domain	'ULF', 'VLF', 'LF', 'HF', 'VHF', 'LFHF', 'LFn', 'HFn', 'LnHF'
3	Non-Linear Domain	'SD1', 'SD2', 'SD1SD2', 'S', 'CSI', 'CVI', 'CSI_Modified', 'PIP', 'IALS', 'PSS', 'PAS', 'GI', 'SI', 'AI', 'PI', 'C1d', 'C1a', 'SD1d', 'SD1a', 'C2d', 'C2a', 'SD2d', 'SD2a', 'Cd', 'Ca', 'SDNNd', 'SDNNa', 'DFA_alpha1', 'DFA_alpha1_ExpRange', 'DFA_alpha1_ExpMean', 'DFA_alpha1_DimRange', 'DFA_alpha1_DimMean', 'DFA_alpha2', 'DFA_alpha2_ExpRange', 'DFA_alpha2_ExpMean', 'DFA_alpha2_DimRange', 'DFA_alpha2_DimMean', 'ApEn', 'SampEn', 'ShanEn', 'FuzzyEn', 'MSE', 'CMSE', 'RCMSE', 'CD', 'HFD', 'KFD', 'LZC'

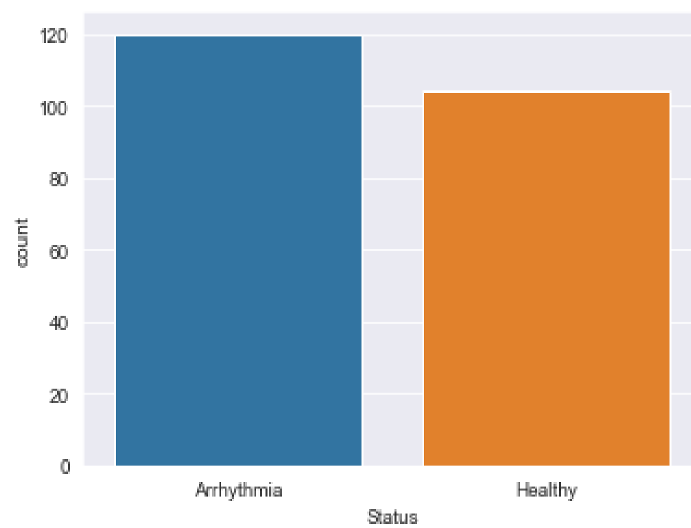

Figure 3: Sample count of the 10-minute dataset. 120 counts for Arrhythmia and 104 for NSR making 224 recordings.

In order to ensure a normal distribution and to ensure data organisation, all features were normalised using the "*StandardScaler*" class from Sci-kit Learn. In "*StandardScaler*", the standard score of a sample x is calculated as:

$$z = \frac{x - u}{s} \tag{1}$$

where u is the mean of the complete dataset and s is the standard deviation.

Data normalisation is conventionally done to ensure cohesion in the data across all fields and records. It helps to balance the impact of all features on a model.

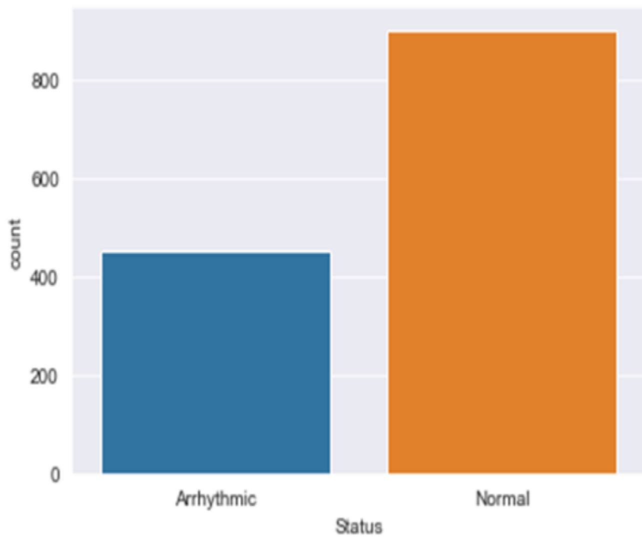

Figure 4: Sample count of the 3-minute dataset. 450 counts for Arrhythmia and 900 for NSR making 1350 recordings.

In total, seven models were trained with the data and a Voting Classifier was used to combine all the seven models into one ensemble model. These models were selected because of their robustness, freedom of ambiguity, ease of use, and effectiveness in working with relatively small datasets. The models used are as follows:

Random Forest Classifier: Random forests are relatively robust, easy to use and can be used effectively with small datasets. A random forest model is made up of a large number of small decision trees called estimators (Krittanawong *et al*., 2017). 200 estimators were used in the random forest model for this study.

K Nearest Neighbours (KNN): K-nearest neighbour (KNN) categorizes data points based on their proximity to and association with other data (Ali *et al*., 2019). KNN is relatively easy to use and performs calculations very fast. However, as the dataset increases in size, the time taken to process the data increases, making the KNN model less efficient for classification problems. The number of neighbours checked to determine classification is typically chosen as the square root of N, the total number of samples. As there were two different sample sizes, this number varied for the two datasets.

Support Vector Machine (SVM): This is a popular supervised learning model. It is most commonly used for classification but can be used for classification and regression (Noble, 2006, Ibitoye *et al*., 2020). The option of whether to enable probability estimates was set to “True” in this case.

Logistic Regression: While linear regression is preferred when dependent variables are continuous, logistical regression is chosen when the dependent variable is categorical, i.e. when they have binary outputs, such as "True" and "False" or "Yes" and "No." Even as both regression models seek to understand relationships between data inputs, logistic regression is mainly used to solve binary classification problems (Education, 2020, Zabor *et al.*, 2022).

Decision Tree Classifier: Decision trees have been applied in many fields that require data mining. It has been classified as one of the most effective methods due to its ease of use, freedom of ambiguity, and robustness even when the dataset has some missing values (Song and Lu, 2015). A dataset can be split into training and validation sets to train the model, and also measure its accuracy. Decision tree models use a validation dataset to decide on the appropriate tree size needed to optimize the model.

Naïve Bayes Classifier: A family of probabilistic classifiers in statistics known as "naive Bayes classifiers" applies the Bayes theorem with conditional independence between each pair of features based on the value of the class variable. This works quite well in many complex real-world situations despite having naive design and oversimplified assumptions. They have been effective in several areas including medical diagnosis (Ramanathan *et al*., 2022). Using Naive Bayes is a simple technique for constructing classifiers: models that assign class labels to problem instances, represented as vectors of feature values, where the class labels are drawn from some finite set.

Stochastic Gradient Descent Classifier (SGDC): SGD is an optimization method, SGD Classifier implements regularized linear models with Stochastic Gradient Descent. The gradient of the loss is computed for each sample at a time and the model is modified along the way with a diminishing severity schedule using this estimator, which employs regularized linear models with stochastic gradient descent (SGD) learning (Tsuruoka *et al*., 2009).

Voting Classifier: This is a model that combines different models in an ensemble and predicts an output based on the input of the other models (Kumar *et al.*, 2017). The estimators in this case are the other models that it considers. Voting can be set to either “soft” or “hard” but soft voting was used in this study as it does not just go with the majority rule but takes into account the confidence (probability) of each classifier’s prediction.

As stated earlier the model was deployed on a Raspberry Pi (Figure 5). The final model was incorporated into Streamlit software (San Francisco, CA, USA), a freely available web application framework that helps in the creation of web apps for data science and machine learning (Parker *et al*., 2021).

C. Deployment and mode of operation

Figure 6 presents the software interface on Streamlit wherein clinicians or users will be able to perform Arrhythmia diagnosis. Figure 7 is a screenshot showing a “Normal” classification while Figure 8 presents a screenshot showing an “Arrhythmic” classification, respectively.

The user enters the filename and sampling frequency of the patient’s ECG recording into the respective fields. The program processes the data, performs resampling and HRV extraction, and outputs a classification.

Figure 5: Raspberry Pi

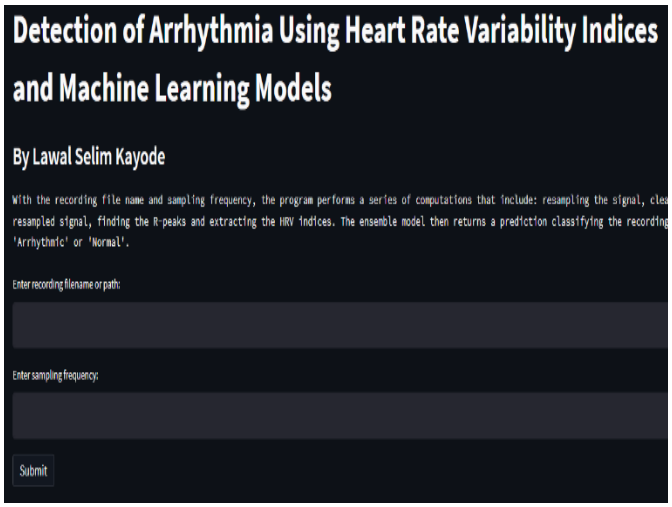

Figure 6: Software interface on Streamlit

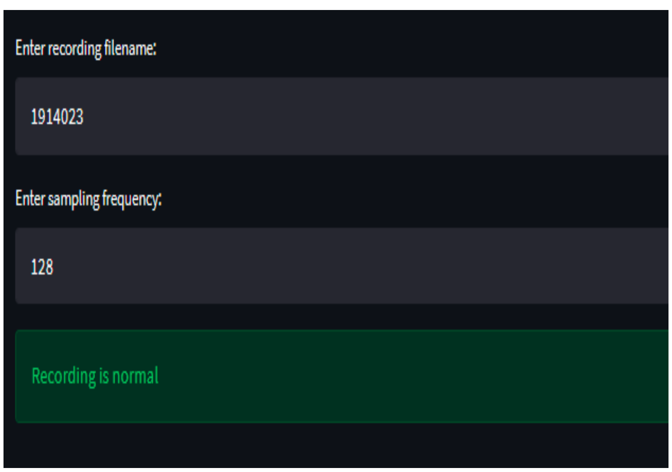

Figure 7: Screenshot showing a "Normal" classification on Streamlit

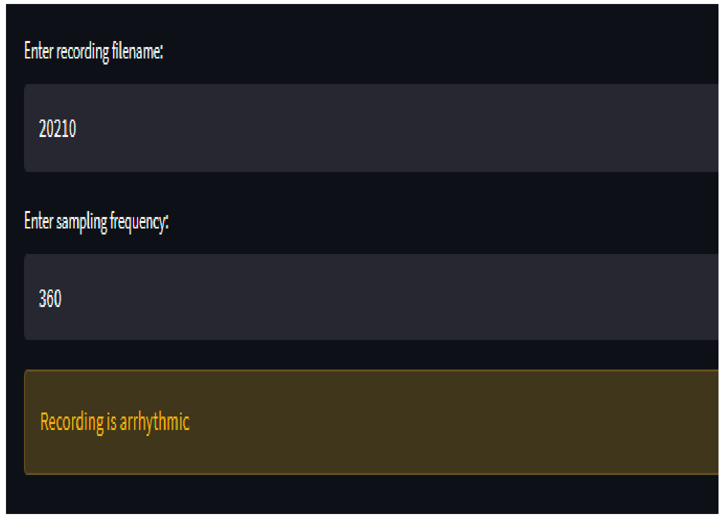

Figure 8: Screenshot showing an "Arrhythmic" classification

III. RESULTS

The performance validation and evaluation of the model involve predicting targets from the features on the test data and comparing the predicted labels to the actual labels. This tells us how good or bad the model's performance was. Some of the performance metrics used in this study are as follows:

i. True Positive (TP): The total number of cases that were correctly classified as "positive".

ii. True Negative (TN): The total number of cases that were correctly classified as "negative".

iii. False Positive (FP): The total number of cases that were incorrectly classified as "positive".

iv. False Negative (FN): The total number of cases that were incorrectly classified as "negative".

v. Precision: This is the ratio of cases correctly classified as positive to the total number of cases classified as positive. It is denoted mathematically as (Eqn. 2):

$$Precision = \frac{TP}{(TP+FP)} \tag{2}$$

vi. Recall refers to the ratio of cases correctly classified as positive to the total number of true positives. It is denoted mathematically as (Eqn. 3):

$$Recall = \frac{TP}{(TP+FN)} \tag{3}$$

vii. Accuracy: Accuracy is the most widely used performance metric for evaluating the performance of a machine learning model. It describes the ratio of correctly classified instances to the total classified instances. It is denoted mathematically as (Eqn. 4):

$$Accuracy = \frac{(TP+TN)}{(TN+FN+TP+FP)} \tag{4}$$

viii. F1 Score: This is the harmonic mean of the precision and the recall. It is denoted mathematically as (Eqn. 5):

$$F1\ Score = \frac{2\times Precision\times Recall}{Precision+Recall} \tag{5}$$

It is important to note that the weighted average was chosen for all performance metrics. The results of each model in the 10- and 3-minute datasets are shown in Tables 5 and 6, respectively.

Figure 9 presents the procedure for the training, validation and testing processes. The random forest classifier for the 3-

minute dataset was selected as the model to be used for deployment because it was the highest performing model. The average time taken for prediction varied slightly with respect to computing and processing power. On an average 4 GB RAM computer system, the time taken for prediction was approximately 3 to 4 seconds, while on a 12/16 GB RAM system, the time taken for prediction was approximately 2 seconds. On the deployment platform (Raspberry Pi), the time taken for prediction was approximately 3 seconds.

v. LFHF (Ratio of Low Frequency to High Frequency)
vi. VHF (Very High Frequency)
vii. CSI_Modified (Modified Cardiac Sympathetic Index)
viii. HFn (Normalized High Frequency)
ix. LnHF (Log transformed High Frequency)
x. HFD (Higuchi Fractal Dimension)
xi. HF (High Frequency)

Table 3. Confusion matrix values for 10-minute data

S/N	MODEL	TRUE POSITIVE	FALSE POSITIVE	TRUE NEGATIVE	FALSE NEGATIVE
1	Random Forest Classifier	32	1	44	2
2	K Nearest Neighbours (KNN)	32	2	43	2
3	Support Vector Machine (SVM)	28	1	44	6
4	Logistic Regression	30	1	44	4
5	Decision Tree Classifier	34	3	42	0
6	Naïve Bayes Classifier	28	3	42	6
7	Stochastic Gradient Descent Classifier (SGDC)	30	0	45	4
8	Voting Classifier	30	0	45	4

Table 4. Confusion matrix values for 3-minute data

S/N	MODEL	TRUE POSITIVE	FALSE POSITIVE	TRUE NEGATIVE	FALSE NEGATIVE
1	Random Forest Classifier	299	14	152	8
2	K Nearest Neighbours (KNN)	297	12	154	10
3	Support Vector Machine (SVM)	304	18	148	3
4	Logistic Regression	305	22	144	2
5	Decision Tree Classifier	288	15	151	19
6	Naïve Bayes Classifier	285	28	138	22
7	Stochastic Gradient Descent Classifier (SGDC)	306	26	140	1
8	Voting Classifier	307	17	149	0

Using the "*feature_importances_*" property in the random forest classifier, the features (indices) that had the most relevance in training and evaluating the best 3-minute model were selected and used to retrain the final model for deployment. This resulted in the final accuracy of 95.96% as shown. The features selected had a score of above 1.70, which was the mean score, as shown in Figure 12. The most important indices for the 3-minute models were found to be:

i. SD1SD2 (Ratio of SD1 to SD2)
ii. CSI (Cardiac Sympathetic Index)
iii. DFA_alpha1 (Detrended Fluctuation Analysis corresponding to short-term correlations)
iv. LFn (Normalized Low Frequency)
xii. SD1d (Short-term variance of contributions of decelerations)
xiii. CVSD (Continuously Variable Slope Delta)
xiv. SD1a (Short-term variance of contributions of accelerations)

The SD1SD2 index (Ratio of SD1 to SD2) was found to be the most important feature in training and prediction. SD1SD2 had the highest score in the 3-minute dataset as shown in Figure 10.

Using the same property, the most important features of the 10-minute dataset were computed. The most important features for the 10-minute dataset were selected based on the

Table 5: Prediction results for the 10-minute recordings

S/N	Model	Precision	Recall	F1-score	Accuracy	AUC
1	Random Forest Classifier	96%	96%	0.96	96.20%	1.00
2	K Nearest Neighbours (KNN)	95%	95%	0.95	94.94%	0.95
3	Support Vector Machine (SVM)	92%	91%	0.91	91.14%	0.98
4	Logistic Regression	94%	94%	0.94	93.67%	0.99
5	Decision Tree Classifier	97%	96%	0.96	96.20%	1.00
6	Naïve Bayes Classifier	89%	89%	0.89	88.61%	0.97
7	Stochastic Gradient Descent Classifier (SGDC)	95%	95%	0.95	94.94%	0.97
8	Voting Classifier	95%	95%	0.95	94.94%	0.99

Abbreviations: AUC- Area Under the ROC Curve; F1- Harmonic Mean of Precision and Recall

Table 6: Prediction results for the 3-minute recordings

S/N	MODEL	PRECISION	RECALL	F1-SCORE	ACCURACY	AUC
1	Random Forest Classifier	96%	96%	0.96	95.96%	0.98
2	K Nearest Neighbours (KNN)	94%	94%	0.94	94.17%	0.95
3	Support Vector Machine (SVM)	96%	96%	0.96	95.74%	0.98
4	Logistic Regression	93%	92%	0.92	93.49%	0.96
5	Decision Tree Classifier	91%	90%	0.90	90.36%	0.92
6	Naïve Bayes Classifier	90%	90%	0.90	89.69%	0.94
7	Stochastic Gradient Descent Classifier (SGDC)	95%	95%	0.95	94.84%	0.95
8	Voting Classifier	95%	95%	0.95	95.07%	0.98

Abbreviations: AUC- Area Under the ROC Curve; F1- Harmonic Mean of Precision and Recall

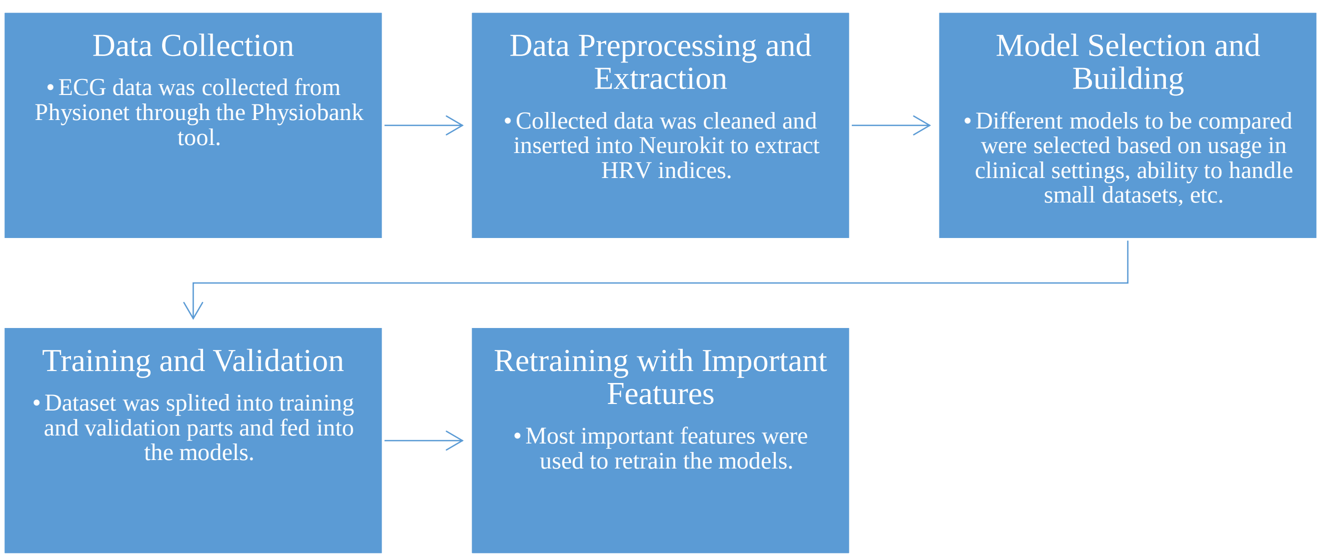

Figure 9: Block diagram summarising the training, validation, and testing processes

mean importance score of 1.45 as shown in Figure 11. The most important indices for the 10-minute models were found to be:

i. VHF (Very High Frequency)
ii. CSI (Cardiac Sympathetic Index)
iii. SD1SD2 (Ratio of SD1 to SD2)

iv. HFD (Higuchi Fractal Dimension)
v. DFA_alpha1 (Detrended Fluctuation Analysis corresponding to short-term correlations)
vi. ShanEn (Shannon Entropy)
vii. LZC (Lempelziv Complexity)
viii. CSI_Modified (Modified Cardiac Sympathetic Index)
ix. DFA_alpha2_ExpMean (Detrended Fluctuation Analysis corresponding to long-term correlations)
x. HF (High Frequency)
xi. LnHF (Log transformed High Frequency)
xii. LFn (Normalized Low Frequency)
xiii. LFHF (Ratio of Low Frequency to High Frequency)

The VHF index (Very High Frequency) was found to be the most important feature in training and prediction. VHF had the highest score in the 10-minute dataset as shown in Figure 11.

The Receiver Operating Characteristic (ROC) curves of the all models for both datasets are represented below in Figure 12 and Figure 13.

detection and arrhythmia classification (Albaladejo-González *et al*., 2023, Udawat and Singh, 2022, Malik *et al*., 2022, Pandey *et al*., 2020). The current study explored classification by analyzing different lengths of ECG signal recordings, trained the models using the extracted HRV indices from those signals, and deployed the final model on a platform easy to understand and use by healthcare professionals. The paper by Vyas and Pandit (2018) used only the time domain measurements of HRV as inputs for the machine learning algorithms, while this current study initially makes use of 77 indices across the time, frequency, and non-linear domains, before selecting the most important indices across all three domains to retrain the model.

This current study also achieved the highest overall accuracy of 96.2%, in comparison to the highest accuracy of 94.54% from previous studies (Vyas and Pandit, 2018, Pandey *et al*., 2020, Udawat and Singh, 2022). Another related study by Pandey *et al.*, (2020) also proposed a system of using

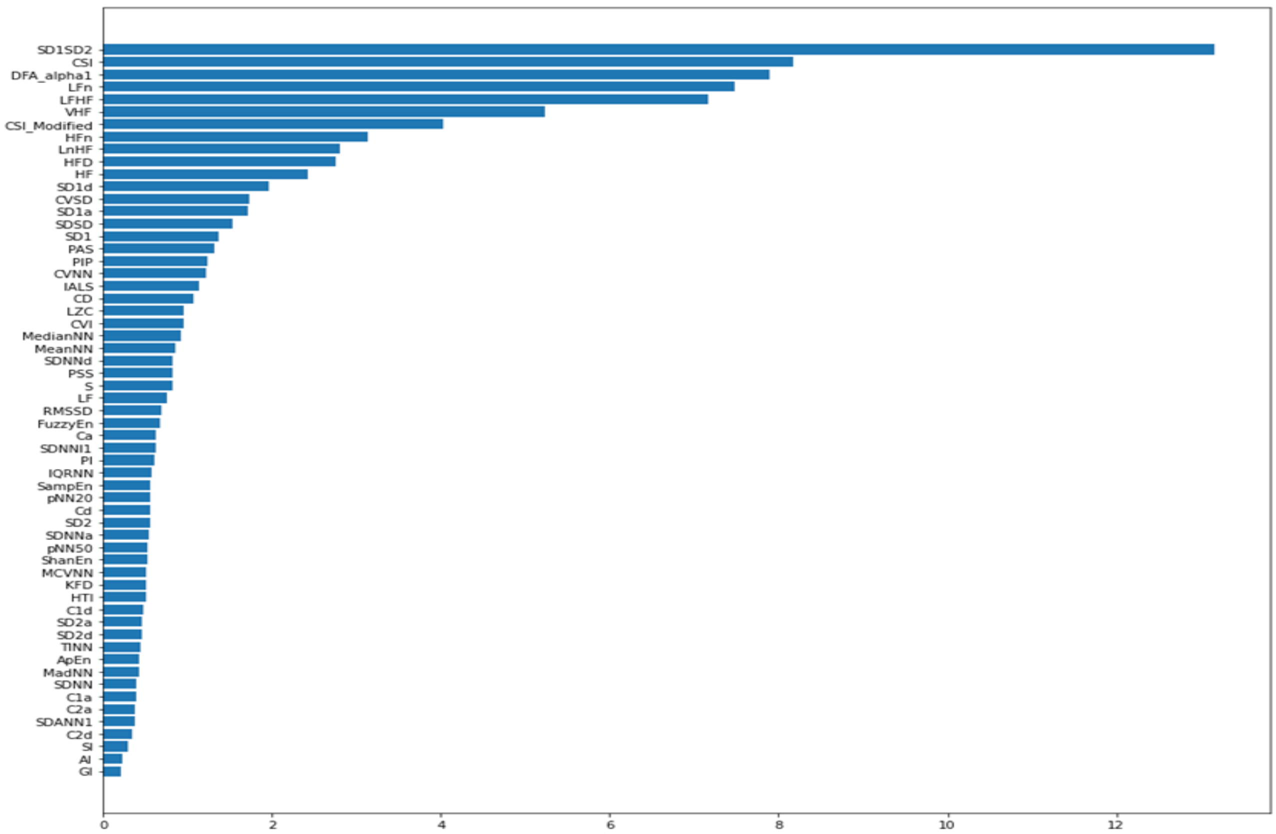

Figure 10: Feature importance chart for the 3-minute dataset

IV. DISCUSSION OF RESULTS

There is a developing industry in the use of machine learning models for the prediction and detection of cardiac abnormalities using ECG signals and classification techniques. Several research studies have been published relating to the classification of heartbeats for beat-to-beat abnormality

ensemble classifiers to detect the presence of arrhythmia in an ECG signal recording. Although their study focused on classifying different types of heartbeats, the models were not deployed onto any user device or platform such as a Raspberry Pi or a software interface. Our study also compares the results of HRV extraction and arrhythmia detection between two sets of ECG recordings of different lengths, 3 and 10 minutes, using the same models, which has not been done by previous studies to our knowledge.

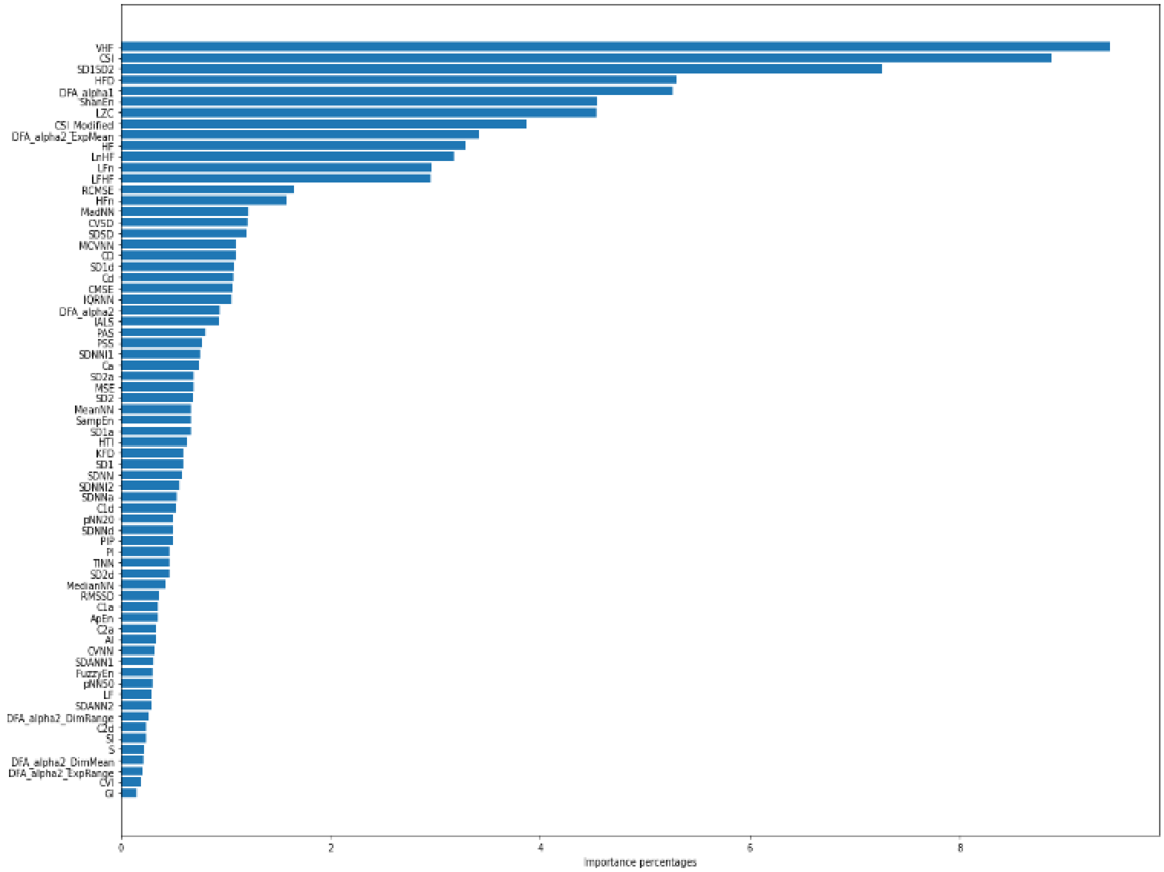

Figure 11: Feature importance chart for the 10-minute dataset

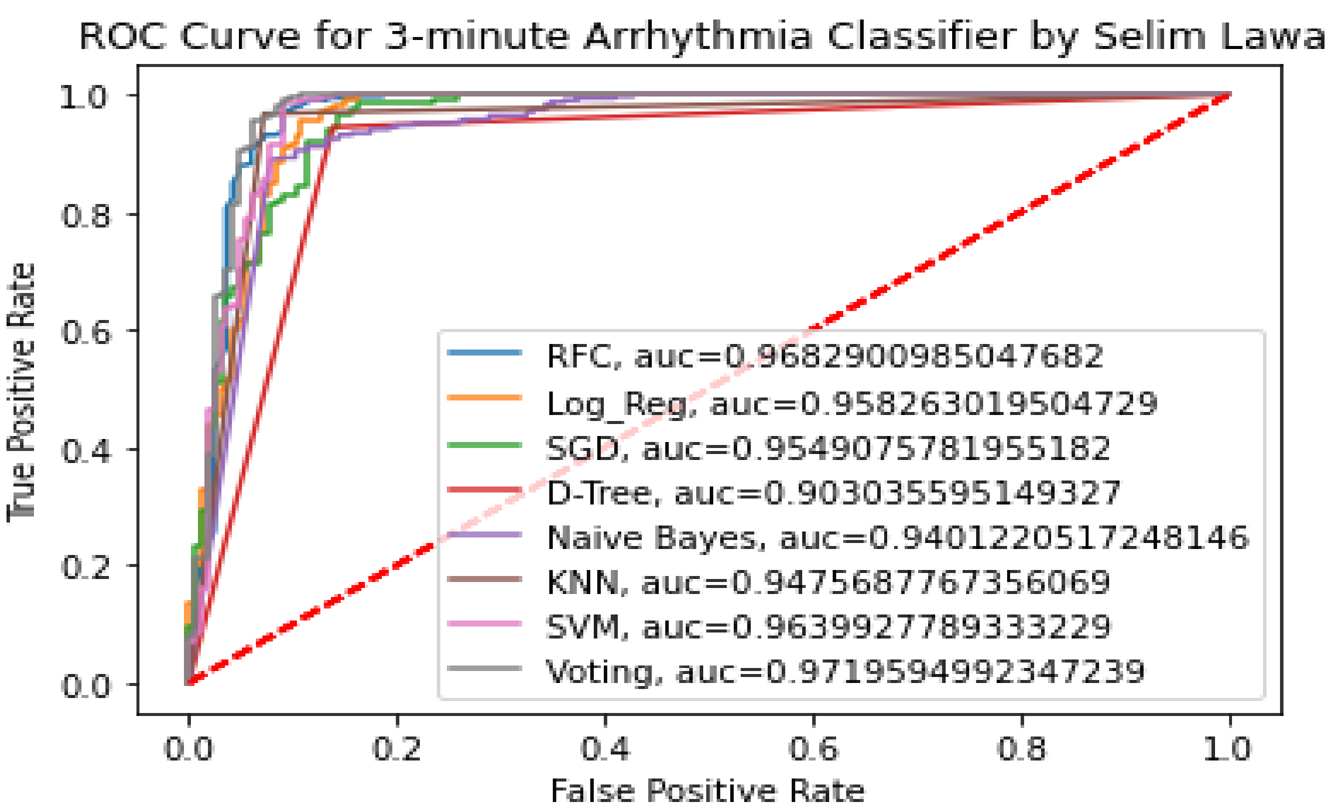

Figure 12: ROC Curve for the 3-minute dataset

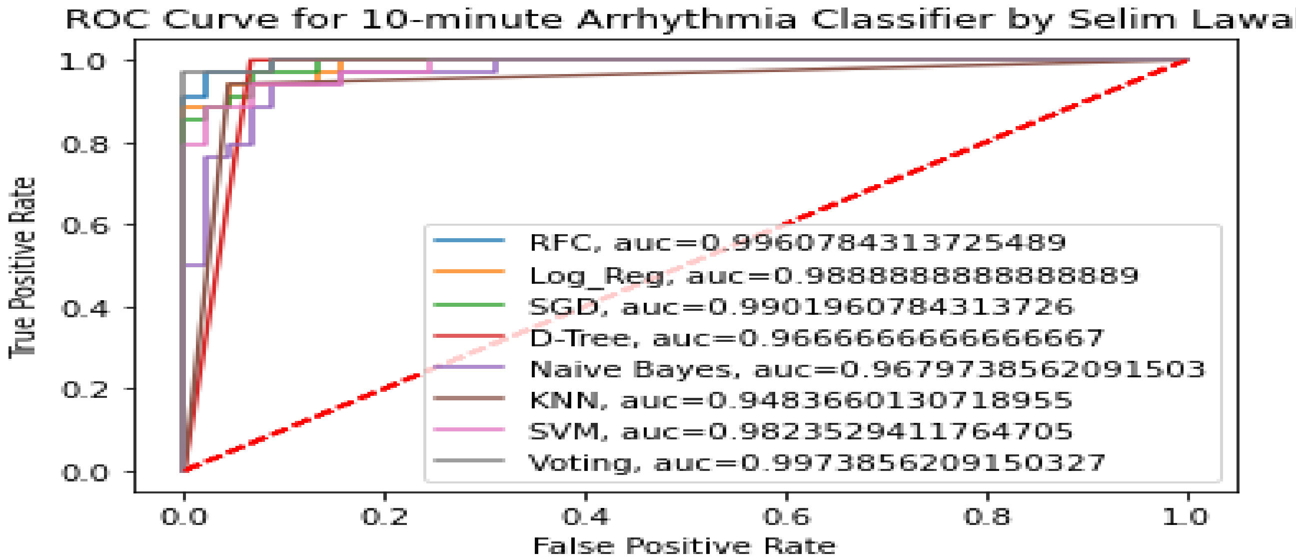

Figure 13: ROC Curve for the 10-minute dataset

The study by Udawat and Singh (2022) presented an approach for automated detection of atrial fibrillation using HRV analysis and machine learning. They were able to train the models on the MIT-BIH dataset for classification and test the proposed algorithm on unseen records from the database but were only able to achieve an accuracy of 94.43%, a specificity of 92.46%, and a sensitivity of 95.16%. In comparison, this current study has been able to achieve accuracies of 96.2% from the dataset of 10-minute intervals, and 95.96% from the dataset of 3-minute intervals. In addition, our study uniquely made use of several other performance metrics to validate the experiment results such as precision, recall, accuracy, F1 score, and AUC. These metrics were applied to the two sets of data and iterated over the 8 models used. The first dataset, with recording lengths of 10 minutes, had the highest precision of 96%, recall of 96%, F1 score of 0.96, and AUC of 1.00. The second dataset consisting of 3-minute recordings had the highest precision of 96%, recall of 96%, F1 score of 0.96, and AUC of 0.98.

It may be noted that the results for the 3-minute and 10-minute datasets had a number of similarities. For example, the most important features for both datasets were more or less the same with the exception of a couple of indices and the order of importance. The model with the highest accuracy for both datasets was also the same. The random forest classifier had the highest accuracies and AUC for both datasets. This may be because the random forest classifier is an excellent model for performing classification due to its robustness, feature importance property, and scalability (Krittanawong et al., 2017). The HRV data was normalized before being fed into the models to create a balanced dataset in each case. As a result, we have been able to effectively compare the performance of different models over the two recording lengths and also found an efficient way to make these models easy to use and available to healthcare professionals.

V. CONCLUSION

In exploring the possibility of detecting arrhythmia in ECG signal recordings, accurate and effective medical diagnosis may be possible within seconds and a large number of patients could be screened accurately. This study developed an ensemble model for the detection of CVDs from ECG's HRV and created a medium for the deployment of the model for easy access to clinicians for seamless classification exercise and patient diagnosis. This study compared the arrhythmia classification based on the ECG recording length used to extract HRV indices and compared the results of several models in achieving this classification. In order to improve usability of these results, the optimal model was deployed using Raspberry Pi and Streamlit.

To further improve this work in the future, more data is required to account for an even wider variety of the types of arrhythmias that may occur in medicine. Arrhythmias occur in different forms, and by gathering more datasets, the model will be exposed to a larger sample size containing more examples reflecting this variety. Although the performance accuracy recorded in this study is high, the model can still perform better and reflect an improvement in performance and accuracy with a larger sample size. An upgraded model that can detect other types of cardiac abnormalities can also be developed by incorporating 3-minute recordings of other conditions into the original dataset.

AUTHOR CONTRIBUTIONS

S. K. Lawal: Conceptualization, Methodology, Software, Validation, Writing – original draft, Writing – review & editing. **I. O. Muniru:** Methodology, Software, Validation, Writing – review & editing. **S. A. Yahaya:** Methodology, Validation, Writing – review & editing, **M. O. Ibitoye:** Conceptualization, Supervision, Validation, Writing – original draft, Writing – review & editing.

REFERENCES

Albaladejo-González, M.; J. A. Ruipérez-Valiente and F. Gómez Mármol. (2023). Evaluating different configurations of machine learning models and their transfer learning capabilities for stress detection using heart rate. Journal of Ambient Intelligence and Humanized Computing, 14 (8): 11011-11021.

Ali, N.; D. Neagu and P. Trundle. (2019). Evaluation of k-nearest neighbour classifier performance for heterogeneous data sets. SN Applied Sciences, 1 (12): 1559.

Chuduc, H.; Nguyenphan, K. and Nguyenviet, D. (2013). A Review of Heart Rate Variability and its Applications. APCBEE Procedia, 7: 80-85.

Connell, P. S.; J. F. Price; C. G. Rusin; T. S. Howard; J. A. Spinner; S. O. Valdes; T. D. N. Pham; C. Y. Miyake and J. J. Kim. (2023). Decreased Heart Rate Variability in Children with Acute Decompensated Heart Failure is Associated with Poor Outcomes. Pediatric Cardiology, 10.1007/s00246-023-03279-7.

Education, I. C. (2020). Supervised learning. IBM. Available online: https://www.ibm. com/cloud/learn/supervisedlearning (accessed on 23 December 2021).

El Naqa, I. and Murphy, M. J. (2015). What Is Machine Learning? *In:* El Naqa, I., Li, R. & Murphy, M. J. (ed.)^(eds.) *Machine Learning in Radiation Oncology: Theory and Applications.* Cham: Springer International Publishing.

Fletcher, G. and Rea, T. (2008). Sudden Cardiac Arrest. *In:* Heggenhougen, H. K. (ed.)^(eds.) *International Encyclopedia of Public Health.* Oxford: Academic Press.

Goldberger, A. L.; L. A. Amaral; L. Glass; J. M. Hausdorff; P. C. Ivanov; R. G. Mark; J. E. Mietus; G. B. Moody; C.-K. Peng and H. E. Stanley. (2000). PhysioBank, PhysioToolkit, and PhysioNet: components of a new research resource for complex physiologic signals. Circulation, 101 (23): e215-e220.

Haines, D. E.; S. Beheiry; J. G. Akar; J. L. Baker; D. Beinborn; J. F. Beshai; N. Brysiewicz; C. Chiu-Man; K. K. Collins; M. Dare; K. Fetterly; J. D. Fisher; R. Hongo; S. Irefin; J. Lopez; J. M. Miller; J. C. Perry; D. J. Slotwiner; G. F. Tomassoni and E. Weiss. (2014). Heart Rhythm Society Expert Consensus Statement on Electrophysiology Laboratory Standards: Process, Protocols, Equipment, Personnel, and Safety. Heart Rhythm, 11 (8): e9-e51.

Handa, B. S.; Li, X.; Baxan, N.; Roney, C. H.; Shchendrygina, A.; Mansfield, C. A.; Jabbour, R. J.; Pitcher, D. S.; Chowdhury, R. A.; Peters, N. S. and Ng, F. S. (2020). Ventricular fibrillation mechanism and global fibrillatory organization are determined by gap junction coupling and fibrosis pattern. Cardiovascular Research, 117 (4): 1078-1090.

Harris, P. and Lysitsas, D. (2015). Ventricular arrhythmias and sudden cardiac death. BJA Education, 16 (7): 221-229.

Ibitoye, M. O.; N. A. Hamzaid; A. K. Abdul Wahab; N. Hasnan; S. O. Olatunji and G. M. Davis. (2020). SVR modelling of mechanomyographic signals predicts neuromuscular stimulation-evoked knee torque in paralyzed quadriceps muscles undergoing knee extension exercise. Computers in Biology and Medicine, 117: 103614.

Jalife, J. (2000). Ventricular Fibrillation: Mechanisms of Initiation and Maintenance. Annual Review of Physiology, 62 (1): 25-50.

Kleiger, R. E.; J. P. Miller; J. T. Bigger and A. J. Moss. (1987). Decreased heart rate variability and its association with increased mortality after acute myocardial infarction. The American Journal of Cardiology, 59 (4): 256-262.

Krittanawong, C.; H. Zhang; Z. Wang; M. Aydar and T. Kitai. (2017). Artificial Intelligence in Precision Cardiovascular Medicine. Journal of the American College of Cardiology, 69 (21): 2657-2664.

Kumar, U. K.; M. B. S. Nikhil and K. Sumangali. Prediction of breast cancer using voting classifier technique. 2017 IEEE International Conference on Smart Technologies and Management for Computing, Communication, Controls, Energy and Materials (ICSTM), 2-4 Aug. 2017 2017. 108-114.

Lipshultz, S. E.; Y. M. Law; A. Asante-Korang; E. D. Austin; A. I. Dipchand; M. D. Everitt; D. T. Hsu; K. Y. Lin; J. F. Price; J. D. Wilkinson and S. D. Colan. (2019). Cardiomyopathy in Children: Classification and Diagnosis: A Scientific Statement from the American Heart Association. Circulation, 140 (1): e9-e68.

Loyse, G. (2017). raspberry-pi Documentation https://www.raspberrypi.com/documentation/ [20 December 2022].

Makowski, D.; T. Pham; Z. J. Lau; J. C. Brammer; F. Lespinasse; H. Pham; C. Schölzel and S. H. A. Chen. (2021). NeuroKit2: A Python toolbox for neurophysiological signal processing. Behavior Research Methods, 53 (4): 1689-1696.

Malik, H.; U. Bashir and A. Ahmad. (2022). Multi-classification neural network model for detection of abnormal heartbeat audio signals. Biomedical Engineering Advances, 4: 100048.

Mehra, R. (2007). Global public health problem of sudden cardiac death. Journal of Electrocardiology, 40 (6, Supplement 1): S118-S122.

Moody, G. B. and Mark, R. G. (2001). The impact of the MIT-BIH Arrhythmia Database. IEEE Engineering in Medicine and Biology Magazine, 20 (3): 45-50.

Moody, G. B.; R. G. Mark and Goldberger, A. L. (2000). PhysioNet: a research resource for studies of complex physiologic and biomedical signals. Computers in Cardiology 2000. Vol.27 (Cat. 00CH37163), 24-27 Sept. 2000 2000. 179-182.

Noble, W. S. (2006). What is a support vector machine? Nature Biotechnology, 24 (12): 1565-1567.

Pai, S. S. (2011). Ventricular Fibrillation. *In:* (ed.)^(eds.) *Essence of Anesthesia Practice.* Elsevier.

Pandey, S. K.; R. R. Janghel and V. Vani. (2020). Patient Specific Machine Learning Models for ECG Signal Classification. Procedia Computer Science, 167: 2181-2190.

Papakyriakou, D. and Barbounakis, I. (2023). Benchmarking and Review of Raspberry Pi (RPi) 2B vs RPi 3B vs RPi 3B+ vs RPi 4B (8GB). International Journal of Computer Applications, 185: 975-8887.

Parker, A.; A. Heflin and L. C. Jones. (2021). Analyzing University of Virginia Health publications using open data, Python, and Streamlit. J Med Libr Assoc, 109 (4): 688-689.

Ramanathan, T. T.; J. Hossen and S. Sayeed. (2022). Naïve Bayes Based Multiple Parallel Fuzzy Reasoning Method for Medical Diagnosis. Journal of Engineering Science and Technology, 17 (1): 472-490.

Shaffer, F. and Ginsberg, J. P. (2017). An Overview of Heart Rate Variability Metrics and Norms. Frontiers in Public Health, 5: 258.

Song, Y. Y. and Lu, Y. (2015). Decision tree methods: applications for classification and prediction. Shanghai Arch Psychiatry, 27 (2): 130-5.

Srinivasan, N. T. and Schilling, R. (2018). Sudden Cardiac Death and Arrhythmias. Arrhythmia & Electrophysiology Review, 7 (2): 111-117.

Tang, S.-Y.; H.-P. Ma; C. Lin; M.-T. Lo; L.-Y. Lin; T.-Y. Chen; C.-K. Wu; J.-Y. Chiang; J.-K. Lee; C.-S. Hung; L.-Y. D. Liu; Y.-W. Chiu; C.-H. Tsai; Y.-T. Lin; C.-K. Peng and Y.-H. Lin. (2023). Heart rhythm complexity analysis in patients with inferior ST-elevation myocardial infarction. Scientific Reports, 13 (1): 20861.

Tsuji, H.; M. G. Larson; F. J. Venditti; E. S. Manders; J. C. Evans; C. L. Feldman and D. Levy. (1996). Impact of Reduced Heart Rate Variability on Risk for Cardiac Events. Circulation, 94 (11): 2850-2855.

Tsuruoka, Y.; J. I. Tsujii and S. Ananiadou. Stochastic gradient descent training for l1-regularized log-linear models with cumulative penalty. Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP, 2009. 477-485.

Udawat, A. S. and Singh, P. (2022). An automated detection of atrial fibrillation from single-lead ECG using HRV features and machine learning. Journal of Electrocardiology, 75: 70-81.

Variability, H. R. (1996). Standards of measurement, physiological interpretation, and clinical use. Task Force of the European Society of Cardiology and the North American Society of Pacing and Electrophysiology. European Heart Journal, 17 (3): 354-381.

Vyas, P. and Pandit, D. (2018). Heartbeat abnormality detection using machine learning models and rate variability (HRV) data. https://doi.org/10.20944/preprints201807.0488.v1.

Werner, K.; K. Kander and C. Axelsson. (2016). Electrocardiogram interpretation skills among ambulance nurses. European Journal of Cardiovascular Nursing, 15 (4): 262-268.

Zabor, E. C.; C. A. Reddy; R. D. Tendulkar and S. Patil. (2022). Logistic Regression in Clinical Studies. International Journal of Radiation Oncology*Biology*Physics, 112 (2): 271-277.