Expert Systems With Applications 163 (2021) 113807

Contents lists available at ScienceDirect

Expert Systems With Applications

journal homepage: www.elsevier.com/locate/eswa

Detection of common risk factors for diagnosis of cardiac arrhythmia using machine learning algorithm

Samir S. Yadav *, Shivajirao M. Jadhav

Dr.Babasaheb Ambedkar Technological University, Lonere, Raigad, India

ARTICLE INFO

Keywords:
Cardiac arrhythmia
Machine learning
Random forest
MIMIC-III database
Electrocardiogram

ABSTRACT

This article aims to establish an accurate and innovative objective framework for classification of cardiac arrhythmia patients by trying to measure the importance of specific factors that are potentially relevant to its diagnosis. Cardiac arrhythmia (CA) is a group of condition related to the irregular heartbeats. It is very essential to prevent a CAs, as they are the most common cause of natural death in all over the world. According to the health reports, more than 4.5 lakh cardiac patients fatalities annually in the United States alone. To diagnose cardiac diseases, patient's reported qualitative symptoms can be useful. However, this strategy may fail sometimes due to less accuracy and false positive cases. Therefore in this work, we strive to find a quantitative basis for more reliable and accurate diagnosis of cardiac arrhythmias. This research used the openly available MIMIC-III database to obtain large quantities of clinical monitoring data from patients over the age of sixteen admitted to intensive care units (ICUs). The database was processed on the Health Sciences and Technology (HEST) Cluster, filtered with in a specified time frame(24hrs, 12hrs and 6hrs) and organized into a multi-class and a single-class and finally split into train, validation, and test sets with respective weights of 0.7, 0.2, and 0.1. We used random forest classifier model for the diagnosis of cardiac arrhythmia and measure the importance of different features like respiratory rate, blood pressure, sodium, potassium, calcium, among the other features. Hyperparameter optimization techniques like grid search and genetic algorithms are compared to find the maximum number and depth of trees in the forest. The model achieved, at its best, an Area Under the Receiver Operator Curve (AUC) score of 0.9787 and, thus, confirmed the importance of several previously suggested factors in the diagnosis of cardiac arrhythmias. We substantiated claims that each of sodium, calcium, potassium, respiratory rates and blood pressure can be used for the early diagnosis of cardiac arrhythmias.

1. Introduction

Despite recent advances in patient diagnostic procedures and medical technology, cardiovascular diseases remain the most common cause of natural death in developed countries (Huikiri et al., 2001). In the United States alone, more than 450,000 people die of cardiac arrhythmias annually (Splawski et al., 2017). Cardiac arrhythmias are heart rhythm problems that occur when the electrical impulses that coordinate one's heartbeats do not work properly, causing one's heart to beat too fast, too slow or irregularly (The Cleveland Clinic, 2019). Traditionally, cardiac arrhythmias have predominantly been diagnosed based on qualitative evaluations and patient-reported symptoms. However, these are largely subjective and not always effective as they rely strongly on a patient's ability to communicate clearly with their medical professional. According to the Cleveland Clinic, more than half of sudden cardiac arrests resulting from cardiac arrhythmias occur without prior symptoms, such as chest pain, dizziness, and noticeable heart palpitations (The Cleveland Clinic, 2019).

In the past several years, medical professionals have proposed experimental and quantitative biosignals as other potential markers of cardiac arrhythmias. Even Apple^{TM} has attempted to break into this field. In it's most recent attempt to use the Apple Watch to predict arrhythmia, a total of 419,297 people's heart rhythms were tracked, of which 2161 participants (0.52 percent of the group), received arrhythmic flags (American College of Cardiology, 2019). However, Apple's attempts have been severely limited by it's reliance on self-reported data, unreliable sensors — and a resulting high number of false positives (American College of Cardiology, 2019).

False positives represent a major hurdle to successfully integrating machine-generated diagnoses into regular clinical use. False positives, or the misdiagnosis of a healthy patient, may result in treatment that

* Corresponding author.
E-mail addresses: ssyadav@dbatu.ac.in (S.S. Yadav), smjadhav@dbatu.ac.in (S.M. Jadhav).

https://doi.org/10.1016/j.eswa.2020.113807
Received 11 March 2020; Received in revised form 18 July 2020; Accepted 29 July 2020
Available online 3 August 2020
0957-4174/© 2020 Elsevier Ltd. All rights reserved.

is harmful to a healthy patient. Although machine learning algorithms are not ready to take the place of human doctors, they can still shed light on the causes of many diseases, in particular cardiac arrhythmias. Medical professionals have only been able to conjecture that certain biosignals may be predictive of an imminent cardiac arrhythmia; machine learning algorithms can support these claims with statistical data and the identification of patterns that are not apparent to the human observer.

Therefore, in this paper, our aim is to find the predictive power of various biosignals, medications, preconditions, and congenital defects for the diagnosis of cardiac arrhythmias with machine learning algorithm. Using the MIMIC-III database, we need access to the model's inputs, such as patient chart events, doctor's notes, prescriptions, lab results, and outputs, arrhythmia diagnoses. Therefore, we used a supervised learning method, specifically the Random Forest method. To be precise, the problem that we attempt to solve in this work is a binary classification problem as each patient either has an arrhythmia or does not have an arrhythmia. There are multiple possible classifiers, or machine learning techniques that could have been used to solve this problem. We selected Random Forest because of it's accuracy and relatively quick training-time (Microsoft Azure, 2019)

Reminder of this paper is organized as follows: Section 2 discuses the related work on cardiac arrhythmia classification. Section 3 presents the medical and technical background of this research work. Section 4 describes the methods and data used in this work. Section 5 presents the results of the experiments carried out in this work, Section 6 discusses the result and limitations of the research work. Finally 7 section gives the conclusion and future work in this direction.

2. Related work

Multiple researchers have used machine learning algorithms, including recurrent neural networks and multi-task Gaussian process models, to predict a variety of patient information, including heart failure onset, patient mortality, and future prescriptions. Their work has paved the way for the work in this paper by (1) showing the value of machine learning algorithms in the medical field, (2) by demonstrating that specific machine learning algorithms can be suited to specific datasets, and (3) by suggesting that including temporal and trend data improves models. Their results strongly influenced the design of this research and also inspired us to use a machine learning algorithm, Random Forest, that was particularly suited to our dataset. Choi et al. (2016b) used Recurrent Neural Network (RNN) models to detect heart failure onset. They used data from 3,884 heart failure cases against 28,903 control cases from May 16, 2000 to May 23, 2013. They adapted a recurrent neural network model to detect relationships between time-stamped diagnoses, prescriptions, and procedures for all cases. They concluded that the use of time-stamped data improved the performance of deep-learning models for the early detection of heart failure during an observation window of 12 to 18 months.

Choi et al. (2016a) later developed $DoctorAI$. $DoctorAI$ used recurrent neural networks and time stamped electronic heath records from 260,000 patients and 2,128 physicians over 8 years to predict the diagnoses and medications for a subsequent visit based on the treatments administered during previous visits. The data used for $DoctorAI$ was extracted from the Medical Information Mart for Intensive Care (MIMIC) database, which can be read about more extensively in Section 3.1.3. They achieved 79.58 recall and also demonstrated $DoctorAI$'s adaptability by testing the model on another institution's database without losing substantial accuracy. They noted that a limitation of $DoctorAI$ is that its incorrect diagnoses, or false positives, can be severely damaging to a healthy patient's health if acted upon and $DoctorAI$ therefore should not be used without human supervision.

Ghassemi et al. (2015) used the MIMIC database to collect "noisy, incomplete, sparse, heterogeneous and unevenly-sampled clinical data, including both physiological signals and clinical notes". They used multi-task Gaussian process models to assess and predict patient acuity, which is the measurement of the intensity of nursing care that a patient requires. The majority of acuity scores rely on a single moment of the patient's data and do not include developing clinical data such as doctors' notes, chart events or lab values. They strove to refine a patient's acuity score by incorporating developing clinical data into the prediction. They first attempted to estimate cerebrovascular pressure reactivity, which often indicates the extent of secondary brain damage, such as cerebral edema or altered cerebral blood flow, in traumatic brain injury patients. They also attempted to use clinical progress notes to predict patient mortality. They concluded that, despite its great computation cost, this approach is effective and with some enhancements could be used regularly in a clinical setting.

Churpek et al. (2016) studied how the addition of a patient's vital sign trends to the patient's previously-measured momentary vital signs affected the model's predictive power. They used data on vital sign trends from five hospitals over a five-year period to predict cardiac arrest, hospital transfer, and death. They ultimately determined that including trends increased accuracy compared to a model containing only momentary vital signs (Area under the Curve (AUC) scores of: 0.78 vs. 0.74).

Ghassemi et al. (2014) examined the use of latent variable models, which relate a set of observable variables to inferred ones, to translate hospital notes into meaningful features, and the predictive power of these features of patient's mortality. Hospital mortality predictions are typically based on gender, age, scores which measure the severity of disease for patients admitted to intensive care units and scores which predict ICU mortality based on lab results and clinical data. In this work, they attempted to enhance previously-made predictions by adding text information extracted from the MIMIC database. They concluded that their models could ultimately be well-integrated into a clinical setting (Remme & Bezzina, 2010).

In the recent years, few researchers have used different database like, MIT-BIH database (Moody & Mark, 2001) for arrhythmia classification and obtained excellent performance to help medical professional.

Pławiak and Acharya (2019) designed a novel three-layer deep genetic ensemble of classifiers(DGEC) for the detection of cardiac arrhythmia using ECG signal. They used imbalanced data of 744 segments of ECG signal of 29 people obtained from the MIT-BIH Arrhythmia database. Their model obtained an accuracy of 99.37% with classification time of single sample of 0.8736 s in the detection of 17 types of ECG arrhythmia. The disadvantages of this DGEC model is complex structure and requires to perform features extraction compare to the other deep learning models.

Tuncer et al. (2019) employed a simple and less complex novel hexadecimal ternary pattern method for automatic detection of cardiac arrhythmia. They used Multilevel wavelet feature extraction for classification of 17 types of ECG arrhythmia and obtained classification accuracy of 95% on ECG signal obtained from MIT-BIH database. They concluded that their classification accuracy is better than other machine learning algorithms. However, they tested their methodology with very smaller set of data in each classification.

Kandala et al. (2019) presented an automatic heartbeat classification based on nonlinear morphological features and majority voting based scheme called ICEEMED, using MIT-BIH Arrhythmia database. In this method they extracted important features of ECG signals which then they used for classification of different classes of arrhythmia. Their method achieved classification accuracy of 90.4% and 100% on fusion and unknown classes. Despite these significant results their model, the performance of few heartbeats like atrial premature contraction, Aberrated atrial, supra ventricular and junctional premature beats class is still low and needs to be improved compared to the other classes. They also concluded that their system can be used for real-time monitoring system in healthcare.

Pławiak and Abdar (2020) presented a novel evolutionary-neural system using SVM classifier. In this work they analyzed longer fragments (10-second) of the ECG signal to classify 17 classes of ECG arrhythmia on MIH-BIH Arrhythmia database(only for one lead–MLII). Their proposed method showed remarkable performance with high sensitivity (90.19%), specificity (99.39%), and accuracy(98.85%). The classification time per single sample of 0.0018(s) was also very best for their model. However, their work has few limitations as they analyzed small number of ECG signal fragments (744 from 29 patients) due to insufficient number of appropriate ECG signals in MIT-BIH database.

Pławiak (2018a) used MIH-BIH Arrhythmia database of ECG signal for classification of 17 types of arrhythmia using novel Genetic Ensemble of Classifiers model based on SVM classifier. Their model achieved a highest classification sensitivity of 91.40%, accuracy of 98.99%, specificity = 99.46%, and time for classification of one sample = 0.0186(s) on 17 types of arrhythmia. Further, it is also possible to implement this model on mobile devices due to its less computational complexity. However, due to insufficient suitable signals in MIT-BIH database, this model is also tested with less number of ECG signal fragments like model of Pławiak and Abdar (2020).

Pławiak (2018b) presented a novel methodology for efficient classification of 17 classed of cardiac arrhythmia using 1000 fragments of ECG signals from the MIH-BIH Arrhythmia database for one lead, MLII, from 45 patients. Their model also achieved higher classification sensitivity of 90.20%, accuracy = 98.85%, specificity = 99.39% and time of classification for one sample = 0.0023 [s]. However, due to insufficient appropriate signals in MIT-BIH database, this model is also tested with less number of ECG signal fragments, similar to the proposed models of Pławiak and Abdar (2020) and Pławiak (2018a).

3. Background

3.1. Medical background

In this section, we dive deeper into the medical background and the possible predictors of cardiac arrhythmias upon which this work is based.

3.1.1. Cardiac arrhythmia

There a four primary categories of cardiac arrhythmia: premature beats, supraventricular arrhythmias, ventricular arrhythmias, and bradyarrhythmias (Keating & Sanguinetti, 2001; Pinnell et al., 2007). Each of these results from a dysfunction of the heart's electrical impulses, or cardiac action potentials (Keating & Sanguinetti, 2001).

Premature beats are the most common type of arrhythmia and they are typically harmless (Pinnell et al., 2007). Patient-reported symptoms, though uncommon, include fluttering in the chest and feelings of a skipped beat. Premature beats typically require no treatment, especially in healthy people (Pinnell et al., 2007).

Supraventricular arrhythmias are tachycardias that start in the atria or the atrioventricular (AV) node (Jogla et al., 2015). The AV node is a group of cells located between the atria and the ventricles. Tachycardia is a condition that makes one's heart beat more than 100 times per minute (Jogla et al., 2015). This happens when the electrical signals in the organ's upper chambers misfire and cause the heart rate to speed up. It then beats so fast that it does not fill with blood before it contracts. Supraventricular arrhythmias require medical attention, but are not typically fatal.

Bradyarrhythmias are arrhythmias in which the heart rate is slower than normal. If the heart rate is too slow, not enough blood reaches the brain and results in loss of consciousness (Jogla et al., 2015). Bradyarrhythmias can be caused by heart attacks, conditions that harm or change the heart's electrical activity, such as an underactive thyroid gland, aging, or an imbalance of chemicals or other substances, such as potassium or beta blockers (Jogla et al., 2015). Bradyarrhythmias require medical attention, but are not typically fatal. Ventricular arrhythmias, such as ventricular tachycardia and ventricular fibrillation, start in the ventricles (Al-Khatib et al., 2018). Coronary heart disease, heart attack, weakened heart muscle, and other problems can cause ventricular arrhythmias. Ventricular arrhythmias require immediate medical attention and are often fatal (Al-Khatib et al., 2018). The previously mentioned electrical impulse, or cardiac action potential, that governs a heart's beating is a change in voltage across the cell membranes of the heart's cells. The cardiac action potential encompasses a series of events that produce a change in voltage the heart's cells. The change in voltage is a result of charged atoms moving between the inside and outside of the cell, through proteins called ion channels (Pinnell et al., 2007).

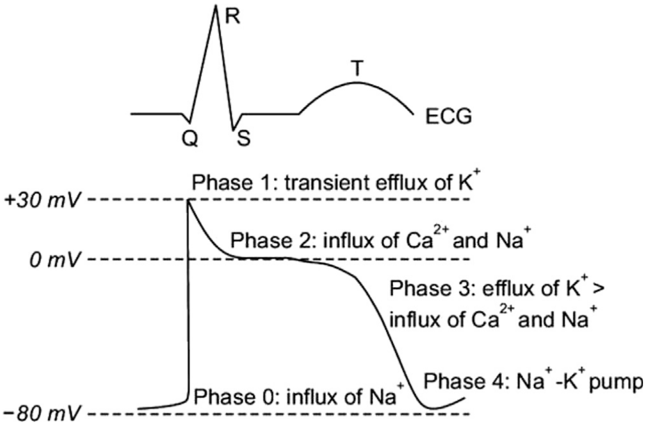

Fig. 1. Cardiac action potential (Cardiac Arrhythmias, 2013).

Though the entire process lasts only about 300 ms, there are five phases of the cardiac action potential, as depicted in Fig. 1 (Pinnell et al., 2007):

0. Potassium levels decrease and fast sodium channels open resulting in rapid depolarization
1. Fast sodium channels close and sodium levels decreasing, causing partial repolarization
2. Repolarization continues as calcium ions leave the cell (plateau phase)
3. Sodium and calcium channels close, causing membrane potentials to return to baseline
4. Rest state: ATPase[1] break down phosphate bonds, which releases energy that ATPase use to drive other cellular or chemical reactions. As three sodium ions convert to just two potassium ions, a negative intracellular potential is created.

During the vast majority of the process described above, the cell is resistant to other stimulation; another action potential will not occur until the whole previous process is complete and the cell is repolarized. If a supramaximal stimulus, a stimulus with strength significantly above that required to initiate another action potential, occurs before the previous action potential is complete, the resulting action potential is stunted (Pinnell et al., 2007).

The cardiac action potential drives the proper functioning of the heart and its impairment can have far-reaching effects.

3.1.2. Important factors for cardiac arrhythmias

Because the proper functioning of the heart depends on the exact balance and timing of sodium-, calcium-, and potassium-ion channel

[1] ATPases are a group of enzymes that decompose ATP into ADP and an extra phosphate bond (or vice versa) (ATPases, 2019).

function, measurements related to those may be a potential avenue to find more reliable, measurable predictors. As mentioned in Section 1, most cardiac arrhythmias today go undiagnosed before a cardiac event because diagnosis depends almost entirely on a patient's ability to recognize and communicate their own symptoms. A reliable and quantitative predictor of cardiac arrhythmia could represent a significant improvement in patient diagnosis.

Several groups of researchers have highlighted some of the lesser-known cardiac arrhythmia symptoms and trends, multiple of which are discussed as follows. Additionally, the Clinical Guidelines for Patients with Cardiac Arrhythmias provide useful insight into what triggers a cardiac arrhythmia and sudden cardiac death (Al-Khatib et al., 2018; Jogla et al., 2015). These features discussed are also tested in the later part of this paper.

Potassium, calcium, and sodium

The flow of calcium, potassium, and sodium through ion channels is integral to heart function. Multiple gene mutations, medicines, and acquired diseases are known that cause disruptions of this process (Keating & Sanguinetti, 2001).

Dysfunction in the heart's potassium ion channels is one cause of cardiac arrhythmia. The human ether-a-go-go (hERG) gene codes for a protein called $K_v11.1$, which is a sub-unit of the potassium ion channel (Sanguinetti & Tristani-Firouzi, 2006). This protein is one of two primarily responsible for the conclusion of the plateau phase of cardiac action potential (Keating & Sanguinetti, 2001). As shown in Fig. 2, the prolongation of the plateau phase is a sign of long-QT syndrome, which greatly increases a patient's risk of ventricular fibrillation (Sanguinetti & Tristani-Firouzi, 2006). Delayed repolarization also increases a patient's risk of Torsades de Pointes (TdP), which can also devolve into a lethal ventricular fibrillation. Mutations of other potassium-related genes such as KVLQT1, KCNQ1, and minK are also known to cause disruptions of the cardiac action potential by prolonging the plateau phase (Sanguinetti & Tristani-Firouzi, 2006). Long-QT syndrome is estimated to affect approximately 1 in 5,000 to 10,000 people worldwide and can be both hereditary and acquired (Sanguinetti & Tristani-Firouzi, 2006). There are multiple known causes of acquired long-QT syndrome, many of which are long-term, age-related diseases, but a metabolic abnormality, such as reduced serum potassium concentration (potassium deficit in the blood), is the one of the most common (Sanguinetti & Tristani-Firouzi, 2006). Dysfunction in the heart's calcium channels can also cause cardiac arrhythmia. However, these calcium channels are distributed unevenly throughout the heart and, as a result, channel dysfunction can have varying effects (Keating & Sanguinetti, 2001). One possible result is a prolonged action potential, which makes the heart muscle cells particularly resistant to re-stimulation. Dispersion of this resistance can result in a unidirectional block of electrical excitement, meaning that some muscle cells will be resistant to beginning the next heart beat (Keating & Sanguinetti, 2001). If an arrhythmia is triggered while the heart is in this state, the arrhythmia continues through a regenerative circuit of electrical activity around the relatively inexcitable tissue. This cycle, known as reentry, can quickly cause ventricular fibrillation and sudden death (Keating & Sanguinetti, 2001).

One possible cause of calcium channel dysfunction is a mutation of the L-type calcium channel CaV1.2 (Splawski et al., 2005). This protein is necessary for the correct coupling of excitation and contraction in the heart (Splawski et al., 2005). Mutations of the CACNA1C gene, which encodes the CaV1.2 protein, can also have effects in the brain, smooth muscle, immune system, teeth, and testis (Splawski et al., 2005). CaV1.2 dysfunction has been shown to cause Timothy Syndrome, a disease characterized by syndactyly (in which one's fingers and toes are abnormally united) and life-threatening cardiac arrhythmias (Splawski et al., 2005). Reduced CaV1.2 channel function, like the hERG mutation, causes prolongation of the plateau phase of the cardiac action potential and results in long-QT syndrome (Splawski et al., 2005).

The SCN5A gene, responsible for the encoding of the sodium channels that initiate cardiac action potential, is also implicated as a potential cause of cardiac arrhythmia (Splawski et al., 2017). More than two decades ago, the first SCN5A mutation was discovered in multiple families with hereditary long-QT syndrome. The same mutation was later found also in patients suffering from Brugada Syndrome, which expresses itself in specific electrocardiogram abnormalities and sudden death (Remme & Bezzina, 2010). Since then, more than 150 SCN5A mutations have been reported, some of which are thought to cause not only long-QT syndrome and Brugada Syndrome but also cardiac conduction defect, sick sinus syndrome, atrial standstill, and susceptibility to dilated cardiomyopathy and atrial fibrillation (Remme & Bezzina, 2010). Mutations of the SCN5A gene can cause both "loss-of-function" and "gain-of-function" effects, both of which negatively impact the heart's precise functioning. "Gain-of-function" mutations can cause prolongation of cardiac action potential and increase risk of long-QT syndrome, while "loss-of-function" mutations can cause shorten the cardiac action potential and higher risk of hypertension (Keating & Sanguinetti, 2001; Remme & Bezzina, 2010).

Furthermore, in the failing heart, an increased number of sodium channels fail to become deactivated and continue with an inward sodium current during what should be the plateau phase (Remme & Bezzina, 2010). This has an effect similar to that of the "gain-of-function" SCN5A mutation (Remme & Bezzina, 2010). In summary, the successful heartbeat requires the correct balance and timing of potassium, sodium, and calcium inflow and outflow. Mutations of any of the three channels can quickly throw off the heart's delicate balance and trigger a cardiac arrhythmia.

Respiratory rates and blood pressure

Respiratory rate is a major predictor of cardiac and pulmonary disorders and events (F.Fieselmann et al., 1992). Many cardiac arrhythmias result from hypoxia, a condition in which some part of the body is not receiving enough oxygen, and hypercarbia, a condition characterized by abnormally elevated carbon dioxide levels in the blood (Cretikos et al., 2002). The human body attempts to correct hypoxia and hypercarbia by pushing more oxygen into the bloodstream (Churpek et al., 2016). This process, of breathing more and pumping more blood, elevates the respiratory rate and systolic blood pressure. Experiments conducted by Churpek et al. (2016), in their attempt to foresee clinical deterioration in hospital patients, suggested that systolic blood pressure and respiratory rates are the two most predictive biosignals (Churpek et al., 2016).

Others

As discussed about Potassium, Calcium, Sodium, Respiratory Rates and Blood Pressure ion channel dysfunction and irregular cardiovascular rates can cause acquired long-QT syndrome. Other features were included in this work, but either seemed less influential or had less extensive documentation. Many medications, including some antibiotics, antihistamines, and anti-arrhythmics can also cause cardiac arrhythmia (Keating & Sanguinetti, 2001). Examples of such drugs include: terfenadine, cisapride, erythromycin, amiodarone, phenothiazines, tricyclic antidepressants, and certain diuretics (Keating & Sanguinetti, 2001). Most of the previously listed medications facilitate long-QT syndrome by blocking hERG channels, the effects of which can be read about in Section 3.1.2. Another drug, Quinidine, induces Torsades de Pointes (TdP) in an estimated 2 to 9 percent of treated patients (Sanguinetti & Tristani-Firouzi, 2006). The recognition of the arrhythmia-inducing effects of some of these drugs has resulted in their removal from the market or relegation to restricted use, and their effects were therefore unable to be measured in this project because they are no longed prescribed in most American hospitals (Sanguinetti & Tristani-Firouzi, 2006).

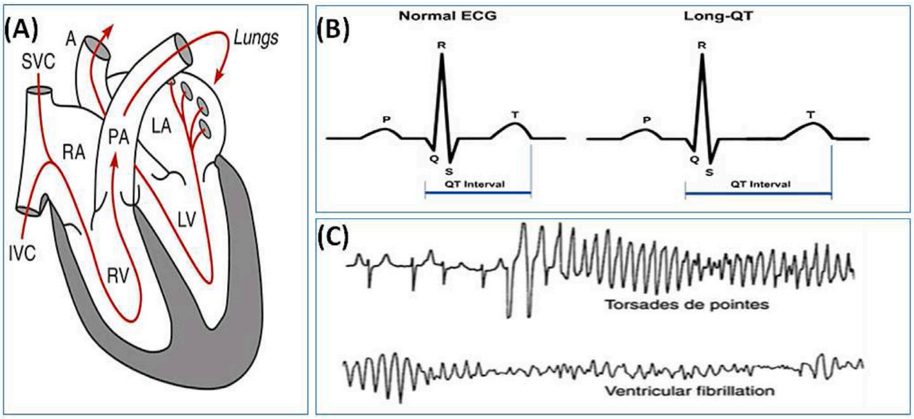

Fig. 2. Fig. 2 depicts the electrocardiograms (ECG) of multiple life-threatening arrhythmias that can be caused by one or a combination of the factors described in this section. (A): A representation of a four-chambered heart and conducting tissues. (B): A normal ECG and one showing long-QT syndrome. (C): Normal breathing rhythms transforming into Torsades de Pointes ventricular tachycardia and ventricular fibrillation.
Source: Sourced from Keating and Sanguinetti (2001).

Finally, according to the Clinical Guidelines (Al-Khatib et al., 2018; Jogla et al., 2015), symptoms, such as syncope, dyspnea, chest pain, cardiac arrest, dyspnea and edema, should also be considered flags for cardiac arrhythmia. Doctors should also be wary of patients with histories of cocaine abuse, alcohol abuse, thyroid disease, kidney disease, lung disease, epilepsy, and hypertension (Al-Khatib et al., 2018; Jogla et al., 2015).

Multiple sources argue that multiple events are required to induce a cardiac arrhythmia and that attention should be paid to a host of biosignals in order to accurately predict a cardiac arrhythmia (Al-Khatib et al., 2018; Keating & Sanguinetti, 2001; Sanguinetti & Tristani-Firouzi, 2006). Additionally, there are more biological predictors, known and unknown, that are not considered in this project, but that may be useful in the prediction of cardiac arrhythmias.

3.1.3. The MIMIC-III database

In order to measure these factors as accurately as possible, the models needs to train, validate, and test on very large quantities of data. This large amount of data was extracted from the MIMIC-III database which is shown in Table 1

The Medical Information Mart for Intensive Care III, or MIMIC III, “is a large, freely-available database comprising de-identified health-related data associated with over forty thousand patients who stayed in critical care units of the Beth Israel Deaconess Medical Center between 2001 and 2012” (Johnson et al., 2016). MIMIC III is primarily operated by the MIT Laboratory for Computational Physiology and associated research groups, and is notable for its large and diverse population of ICU patients, as well as detailed and complete lab test and chart event reporting. The MIMIC III database uses patient-unique identifiers, admission-unique codes, ICD-9 and -10 codes, as well as diagnostic codes to organize information. Any diagnosis, patient, chart event, or admission can be accessed or queried with a unique code (Johnson et al., 2016). All data referenced and used in this work were extracted from the MIMIC-III database.

3.2. Technical background

In the previous section, we discussed the medical background that underlies this paper and previous projects similar to this one. In this next section, we will cover the technical and computational background required to understand the significance of the results discussed in result and discussion section

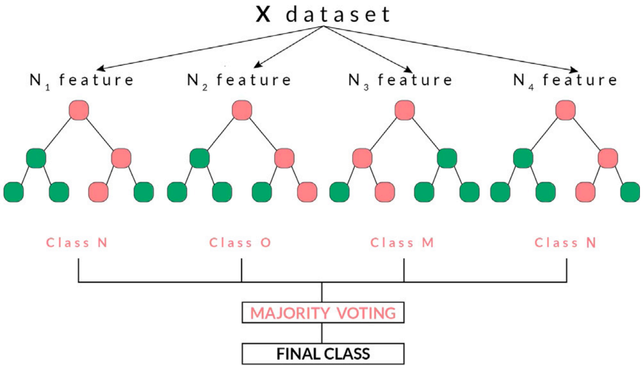

Fig. 3. The image is a graphical representation of the splits made by the many decision trees in the Random Forest.
Source: Image sourced from Quantinsti (Tahsildar, 2019).

3.2.1. Random forest classification

The Random Forest algorithm is used for both classification and regression. As illustrated in Fig. 3, random forests construct many individual decision trees, whose predictions are aggregated to make a final prediction. To predict a target value, decision trees split data into increasingly small subsets. Each leaf, or node, is represented by a condition or a subset of data and each branch, or edge, as a decision (Breiman & Cutler, 2002). The splitting process continues indefinitely until a preset maximum depth is reached or no more can be learned from further splitting. When used for classification, the mode of the classes is taken for the final prediction. The Random Forest algorithm is as follows (Breiman & Cutler, 2002):

1. Given N cases, select N bootstrap samples, meaning select N cases with replacement.
2. Given M input variables, an m≪M is selected and held constant throughout the growing of the forest. At each node, rather than choosing the best split among all M variables, a random group of size m is selected and the best split from within this subset is used to split the node.
3. Each tree is grown unpruned, meaning that it is grown to the largest extent possible.
4. New predictions are made by aggregating the predictions of the previous N trees

Table 1
MIMIC-III, a freely accessible critical care database (Johnson et al., 2016), Where: CCU is Coronary Care Unit; CSRU is Cardiac Surgery Recovery Unit; MICU is Medical Intensive Care Unit; SICU is Surgical Intensive Care Unit; TSICU is Trauma Surgical Intensive Care Unit.

Critical care unit	CCU stays, No. (% by unit)	CSRU stays, No. (% by unit)	MICU stays, No. (% by unit)	SICU stays, No. (% by unit)	TSICU stays, No. (% by unit)	Total stays, No. (% by unit)
Infectious and parasitic diseases, i.e., septicemia, other infectious and parasitic diseases, etc., (001–139)	305 (4.2%)	72 (0.8%)	3,229 (16.7%)	448 (5.6%)	152 (2.8%)	4,206 (8.6%)
Neoplasms of digestive organs and intrathoracic organs, etc., (140–239)	126 (1.8%)	287 (3.2%)	1,415 (7.3%)	1,225 (15.3%)	466 (8.6%)	3,519 (7.2%)
Endocrine, nutritional, metabolic, and immunity (240–279)	104 (1.4%)	36 (0.4%)	985 (5.1%)	178 (2.2%)	54 (1.0%)	1,357 (2.8%)
Diseases of the circulatory system, i.e., ischemic heart diseases, diseases of pulmonary circulation, dysrhythmias, heart failure, cerebrovascular diseases, etc., (390–459)	5,131 (71.4%)	7,138 (78.6%)	2,638 (13.6%)	2,356 (29.5%)	684 (12.6%)	17,947 (36.6%)
Pulmonary diseases, i.e., pneumonia and influenza, chronic obstructive pulmonary disease, etc., (460–519)	416 (5.8%)	141 (1.6%)	3,393 (17.5%)	390 (4.9%)	225 (4.1%)	4,565 (9.3%)
Diseases of the digestive system (520–579)	264 (3.7%)	157 (1.7%)	3,046 (15.7%)	1,193 (14.9%)	440 (8.1%)	5,100 (10.4%)
Diseases of the genitourinary system, i.e., nephritis, nephrotic syndrome, nephrosis, and other diseases of the genitourinary system (580–629)	130 (1.8%)	14 (0.2%)	738 (3.8%)	101 (1.3%)	31 (0.6%)	1,014 (2.1%)
Trauma (800–959)	97 (1.3%)	494 (5.4%)	480 (2.5%)	836 (10.5%)	2,809 (51.7%)	4,716 (9.6%)
Poisoning by drugs and biological substances (960–979)	50 (0.7%)	2 (0.0%)	584 (3.0%)	58 (0.7%)	11 (0.2%)	705 (1.4%)
Other	565 (7.9%)	739 (8.1%)	2,883 (14.9%)	1,204 (15.1%)	563 (10.4%)	5,954 (12.1%)
Total	7,188 (14.6%)	9,080 (18.5%)	19,391 (39.5%)	7,989 (16.3%)	5,435 (11.1%)	49,083 (100%)

Under-and overfitting

Underfitting refers to the case when the model, or machine learning algorithm, is unable to learn the underlying pattern, usually as a result of not having enough data. In this case, the model usually underestimates the complexity data it is being trained on. The opposite problem is called "overfitting". Overfitting occurs when the model not only learns the pattern, but also the noise in the training set. Every dataset has a certain amount of irreducible error, which effectively measures the amount of noise in the data. More simply, overfitting occurs when the model begins to memorize the training data. Overfitting, to which complex models such a Randoms Forests are especially susceptible, are symptomized by high variance and low bias (Koehrson, 2018). Random Forests, if not limited, are prone to overfitting because the forest can continue growing until it has perfectly learned all the data, including the data's noise. In this case, the forest will grow until it has only one leaf node for every tree and has perfectly classified all of the data. While this may sound ideal, it only benefits the training set and diminishes the model's performance on the test set. Overfitting can be avoided with cross-validation, pruning, early stopping and regularization (Pedregosa et al., 2019). By limiting the depth of each tree, we can reduce the model's variance at the expense of increasing the model's bias (Singh, 2018; Singh et al., 2018).

3.3. Gini impurity

Node impurity, or Gini Impurity, is the probability that a new instance would be incorrectly classified if the new instance were randomly classified by the distribution of labels in the dataset (Koehrson, 2018). The Gini Impurity represents the probability of a new data point being incorrectly classified based on the training we have already observed.

The Gini Impurity of a node n is calculated by the following formula (Breiman & Cutler, 2002; Koehrson, 2018):

$$I_G(n) = \sum_{i=1}^{J} (p_i) \times (1 - p_i) = 1 - \sum_{i=1}^{J} (p_i)^2 \tag{1}$$

where J are all classes (in a binary classification J = 2) and p_i is the probability of a classification i. At each node, each decision tree uses the Gini Impurity to choose which feature to split the node on. The decision tree chooses the feature that most decreases the Gini Impurity for the node in question (Koehrson, 2018). It repeats this process recursively on each node until the tree reaches its preset maximum depth or no further information can be gained. If each node contains only samples from one class, no further information can be gained (Koehrson, 2018). At the last possible layer, the Gini Impurity goes to zero, meaning that there is no chance that a new instance would be misclassified (Koehrson, 2018). While this may sound positive, the model could be overfitted.

Gini importance

Random Forests allow for the calculation of feature importances, otherwise known as Gini Importances (Breiman & Cutler, 2002). Feature importance is generally measured by the mean decrease in node impurity offset by the probability of reaching the same node.

In Sci-Kit Learn, the feature importance functions uses the Gini Importance for a binary tree:

$$ni_j = w_j C_j - w_{left(j)} C_{left(j)} - w_{right(j)} C_{right(j)} \tag{2}$$

where ni_j is the node importance of node j, w_j is the number of weighted samples reaching node j, and C_j is the impurity of node j (Breiman & Cutler, 2002). The importance of each feature is then calculated by:

$$fi_i = \frac{\sum_{j:nodejsplitsonfeaturei} ni_j}{\sum_{n\in allnodes} ni_n} \tag{3}$$

where fi_i is the importance of feature i and ni_j is the importance of node j. These values can be normalized and averaged over all trees to find feature importance at the forest level.

3.4. Metrics

Metrics, or scores, are used to evaluate the quality of the model's predictions. The majority of the metrics are based on a confusion matrix. Confusion matrix is combination of four different values: True Positive (TP) means model correctly predicts patients of cardiac arrhythmia, False Positive (FP)means model predicts normal patients as cardiac arrhythmia, True Negative (TN) means model correctly predicts healthy patients as healthy, and False Negative (FN) means model predicts cardiac arrythmia patients as normal or healthy. We have used several different metrics to measure model performance as follows:

3.4.1. Recall

The recall score is calculated by dividing the total number of TPs by the sum of the TPs and FNs (Pedregosa et al., 2019). In a perfect model, a recall score would be 1.0, as there would be no false negatives. The closer the model's recall score is to 1.0, the better the model is.

3.4.2. Precision

The precision score, similar to the recall score, is calculated by dividing the total number of TPs by the sum of the TPs and FPs (Pedregosa et al., 2019). Again, one hopes for as few false positives as possible and aims for a precision score of 1.0.

3.4.3. Average precision

The Average Precision score encapsulates the precision–recall curve by using the weighted mean of precisions at each threshold weighted against the increase in recall from the previous threshold (Pedregosa et al., 2019):

$$AP = \sum_n (R_n - R_{n-1})P_n \tag{4}$$

where P_n and R_n represent the precision and recall, respectively, at threshold n (Pedregosa et al., 2019).

3.4.4. F1 score

The f1 score, or the F measure, is useful to compare models with dissimilar precision and recall scores. The F1 score attempts to measure both precision and recall simultaneously, using the following equation (Pedregosa et al., 2019):

$$f1 = \frac{2 \times Recall \times Precision}{Recall + Precision} \tag{5}$$

The f1 score emphasizes the outlying smaller values and suppresses outlying larger values by using the Harmonic Mean, rather than the Arithmetic Mean.

3.4.5. Hamming loss

The Hamming Loss computes the Hamming Distance between two sample sets in order to determine percentage of the sample mislabeled (Pedregosa et al., 2019). A simple example follows: Given that p_j is the predicted value of the label j, y_j is the true value of label j, and n is the number of labels, the Hamming Loss between two samples is defined as (Pedregosa et al., 2019):

$$L_{Hamming}(y, p) = \frac{1}{n}\sum_{j=0}^{n-1} 1(p_j \neq y_j) \tag{6}$$

where $1(x)$ is an indicator function.

3.4.6. The area under the curve score

The Area Under the Curve Score, or AUC score, is a performance measurement for classification problem at various thresholds settings. The AUC score measures the area under the ROC curve, which stands for the Receiver Operator Characteristic Curve. The ROC curve is a probability curve and is a graphical representation of the diagnostic ability of a binary classifier system (Narkhede, 2018; Pedregosa et al., 2019). More simply, it depicts how well the model is able to differentiate between classes. ROC curves typically have the Recall, otherwise called the True Positive Rate (TPR) or Sensitivity, on the Y axis and the False Positive Rate (FPR), or probabilistic inverse of the specificity, on the X axis (Pedregosa et al., 2019).

3.4.7. Specificity, sensitivity and threshold in the ROC curve

Sensitivity is another word for Recall, which again is the number of true positives divided by the sum of the true positives and false negatives. Specificity is 1 - FPR, the false positive rate. Specificity is calculated by dividing the number of true negatives by the sum of the false positives and true negatives (Narkhede, 2018; Pedregosa et al., 2019). Threshold is the limit at which we classify a non-binary prediction as a Positive or Negative. By manipulating the threshold, we can minimize or maximize Type 1 Error, the number of False Positives, and Type 2 Error, the number of False Negatives (Huang, 2018).

3.5. Optimization, training, and classification time

Optimization time is time needed to find optimal parameter for a given classifier. Training time is the sum of the training times for all training sets for a given variant of cross-validation method. Whereas, the classification time is the average time for a single classification of a 10-s fragment of an ECG signal after pre-processing and feature extraction and selection.

3.5.1. Software and hardware

Python Libraries: For machine learning developments Python is the most popular language. Python is freely usable and distributable as it is developed under an OSI-approved open source license (Foundation, 2018). The Python Package Index (PyPI) hosts thousands of third-party modules and libraries for Python (Foundation, 2018) is used extensively in this work and are elaborated upon below. For this project, we used Python 2.7.

1. **Argparse**: The argparse library allows for the creation user-friendly command-line interfaces and was used extensively in *ProgramArguments* file. This program defines what arguments a program requires and parses those arguments into an easily-usable Python dictionary (Foundation, 2018). The argparse module als+o automatically generates help and usage messages and issues errors when users give the program invalid arguments (Foundation, 2018).
2. **Datetime**: The datetime module allows users to easily manipulate dates and times in both simple and complex ways (Foundation, 2018). The package supports date and time arithmetic, parsing, and intelligent formatting (Foundation, 2018). This module was used primarily for the parsing of time stamps on chart events and lab values and for manipulating patient time frames.
3. **Num-Py**: Num-Py is a submodule of the Sci-Py library which is sponsored by Entthought (SciPy Organization, 2019). NumPy is the primary package used for scientific computing with Python (SciPy Organization, 2019). This package contains a powerful N-dimensional array object, basic linear algebra functions, basic Fourier transforms, sophisticated random number capabilities, tools for integrating Fortran code, and tools for integrating C/C++ code (SciPy Organization, 2019). The Numpy package allows for the easy conversion of Python dictionaries into C arrays, which the Sci-Kit libraries require for training.

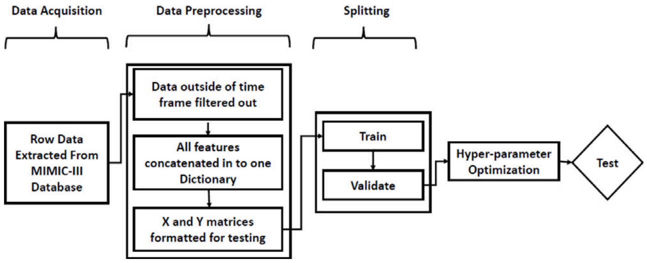
Data Acquisition
Data Preprocessing
Splitting
Row Data Extracted From MIMIC-III Database
Data outside of time frame filtered out
All features concatenated in to one Dictionary
X and Y matrices formatted for testing
Train
Validate
Hyper-parameter Optimization
Test

Fig. 4. Model design.

Sci-Kit Learn: We have also used Sci-Kit Learn, formerly scikits.learn, is a free machine learning library with a Python interface, though c-libraries are sometimes used for arrays and matrix operations (Pedregosa et al., 2011). The library includes various tools for classification, regression and clustering and is designed to seamlessly operate with NumPy and SciPy.
Three of such libraries were **SQLite:** It is a library based in C that implements a miniature SQL that generally follows standard PostgreSQL syntax (sql, 2019).
Hardware: For this work, we have used an Ubuntu machine with Xeon E5-2697, 64 GB memory, and Nvidia Tesla K80 was used to train model.

4. Methodology

In this section, we will discuss how the data from the MIMIC-III database was prepared for training, validation, and testing. The following flow chart in Fig. 4 shows a simplified version of the process and Table 2 shows the detail information about the proposed methodology.

4.1. Data acquisition

In this work, we used SQLite to extract data from the MIMIC III database. SQLite is the most used database in the world and was founded in 2000 (sql, 2019). The data was cleaned and processed with Python 3.0 and the model was trained with a variety of Sci-Kit Learn libraries. In order to utilize the MIMIC's by-patient-id search optimizations, we optimized SQLite queries in many functions by looping through patient identifiers, rather than creating more expansive searches. In a test done with getting the patient's admit times, a loop query completed the task in 17.9274 s, while the batch query needed 102.4190 s. A single function, using mutable ICD-9 codes to specify the desired feature or diagnosis, was used for most queries through. According to the Centers for Disease Control and Prevention (Center for Disease Control and Prevention, 2015): "The International Classification of Diseases (ICD) is designed to promote international comparability in the collection, processing, classification, and presentation of mortality statistics. This includes providing a format for reporting causes of death on the death certificate. The reported conditions are then translated into medical codes through use of the classification structure and the selection and modification rules contained in the applicable revision of the ICD, published by the World Health Organization". We used similar functions to query for specific drugs, diagnoses, and chart events.

4.2. Data preprocessing: Selection

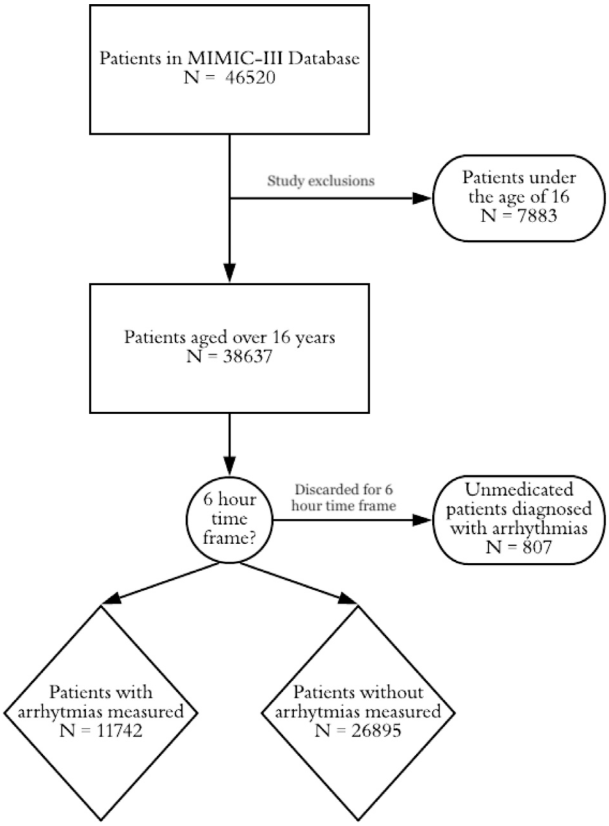
Patients in MIMIC-III Database
N = 46520
Study exclusions
Patients under the age of 16
N = 7883
Patients aged over 16 years
N = 38637
6 hour time frame?
Discarded for 6 hour time frame
Unmedicated patients diagnosed with arrhythmias
N = 807
Patients with arrhytmias measured
N = 11742
Patients without arrhytmias measured
N = 26895

Fig. 5. Flowchart for inclusion or exclusion of patients in our work.

In summary, the data was acquired from the version of MIMIC III processed on the Health Sciences and Technology (HEST) Cluster, filtered to be with in a specified time frame, organized into a multi-class X and a single-class Y, and finally split into train, validation, and test sets. Data preprocessing and cleaning comprised much of the heavy-lifting in this project. The software is mutable through a Program Arguments file and is described in more detail in the subsection Program Arguments 4.2.1. Two aspects of the timing were considered. The number of hours of data to be measured and which hours they should be. Data from non-binary biosignals (e.g. respiratory rates) was collected only within a specific time window. Each data point's time stamp was used to filter out data points outside the preset time frame. Fig. 5 depicts a flowchart that may clarify the following explanation of how cases were included or excluded from the study. The time frame was chosen differently for patients who had been diagnosed with arrhythmia and those who had not. We further differentiated between patients who had been medicated for said arrhythmia and those who had not. Despite the fact some patients had not been medicated, they were nonetheless included in some of our tests. The data of medicated patients diagnosed with cardiac arrhythmia was measured for the preset number of hours before their time of first medication. The prescription of any of the following anti-arrhythmic medications would include a patient in the "medicated patients" group: Procainamide, Feclainide, Sotalol, Metoprolol, Toprol, Verapamil, Digoxin, Amiodarone, Potassium Chloride, Torsemide, or Sodium Chloride. The data of unmedicated patients diagnosed with cardiac arrhythmia was measured for the preset number of hours after the admission in which they were diagnosed with cardiac arrhythmia (as a patient could have multiple distinct admissions). Patients diagnosed with cardiac arrhythmia were medicated on average within 11 h 16 min and 17 s of admission (Pollard & Johnson, 2016). As a result, patients diagnosed with cardiac arrhythmia but not prescribed one of the above listed medications were included in tests in which time frames were 24 or 12 h (but not for those in which time frame was 6 h).

The data of patients who were not diagnosed with cardiac arrhythmia was measured for the preset number of hours after their earliest admission. Patients under the age of 16 were not included in any of the

Table 2
Detailed information about the optimal proposed method.

Steps		Description
Data acquisition	Extraction	-The raw data extracted from the MIMIC III database and stored in an SQLite database. -Data optimized by SQLite queries in many functions by looping through patient identifiers. -This data then translated into medical codes through the use of the classification structure and the selection and modification rules contained in the applicable revision of the ICD, published by the WHO.
Data preprocessing:	Selection	-The MIMIC database processed on the HEST Cluster. -Data selection is done by using the following constraints: (i) Age: Patients aged over 16 years selected only (ii) Time frame: (6 h, 12 h, and 24 h) (iii) Patients diagnosed with arrhythmia or not. -Only one value per feature was associated with each patient when the model was trained on the data. -For each feature, we used the median of all the patient's data points over the selected time frame. -Two types of arguments considered while selecting data: Biosignal selection and Model Modification
	Organization	-Two matrices built for training: Ytrue and Xdata matrix -Ytrue matrix(one dimensional) stored true outcomes for patients (Normal and with Arrhythmia) -Xdata (m × n) matrix stored selected biosignals for all the loaded patients.
Splits	Train and validate	-To avoid an imbalanced distribution stratified shuffle Split is used with respective weights of 0.7, 0.2, and 0.1 for the train, validate, and test data.
Hyperparameter optimization	No .of trees and Depth of trees	-When the split is set to validate, hyperparameters optimization is used to find the maximum number and depth of trees in the forest. -AUC metric was used to judge the quality of the tested hyperparameters.
Testing		-The model tested a total of 16 numbers of biosignals. -We ran tests with 24-, 12-, and 6-hour time frames.

training, validation, or testing processes, as a child's normal respiratory rate is very different than that of an older patient (American Heart Association, 2017; Pinnell et al., 2007). It should also be made clear that the time of medication does not necessarily denote the time of a cardiac event, but was the best way to approximate the time of a cardiac event.

After the number of hours was selected, each patient's lab values and chart events were filtered so that only those in the specified time frame were measured. Though each patient could potentially have multiple measurements for a single feature, only one value per feature was associated with each patient when the model was trained on the data. For each feature, we used the median of all the patient's data points.

Features were not normalized because, unlike many other algorithms, outliers do not affect the Random Forests ability to make decisions. The Random Forest makes splits depending on what most decreases the Gini Impurity and can make those splits no matter the scale of variations in data.

4.2.1. *Program arguments*

The program includes a variety of mutable run-time arguments that generally fall into one of two categories: biosignal selection or model modification. Biosignal selection allows the user to select any combination of the following biosignals for prediction: respiratory rates, potassium levels, sodium levels, calcium levels, and blood pressure. It also allows the user to select histories of the following: quinine, terfenadine, astemizole, cocaine use, alcohol abuse, muscular dystrophy, renal failure, heart failure, epilepsy, lung disease, and pulmonary circulation disorder. Biosignal selection options also include qualitative symptoms, such as dyspnea, labored breathing, and angina, chest pain.

Parameters for model modifications include choices for: database, validation set fraction, test set fraction, various output files, maximum depth of tress in the random forest, maximum number of trees in the random forest, inclusion/exclusion of child patients, the number of hours of chart events to measure, and the number of patients to be included. The number of patients included can also be offset, so that an exact range of patient identifiers can be selected. Data was collected and measured on exclusively the patient identifiers that appear in the list of "loaded patients".

4.3. *Data preprocessing: Organization*

Two tables of information are necessary for training. The y_{true} matrix, in which the true outcomes for each patient are stored, and the x_{data} matrix, in which the selected biosignals for all the loaded patients are stored. The y_{true} matrix is a one-dimensional matrix that contains only zeros and ones, zeros for patients who have not been diagnosed with cardiac arrhythmia and ones for patients who have been diagnosed with cardiac arrhythmia. The matrix is ordered by patient identifier, so that the rows of the x and y matrices line up correctly. The x_{data} matrix is an m by n matrix, where m is the number of biosignals measured and n is the number of patients loaded. The x and y matrices are both converted, from the Python dictionaries that originally stored the un-ordered information, with the same pre-processing function that effectively just removes the patient identifiers (as the model should not be trained on random identifiers) and converts the data type to Num-Py array. Creating the y_{true} dictionary is done by loading the list of patients that have been diagnosed with cardiac arrhythmia, creating a dictionary with all patient identifiers from loaded patients as keys, and assigned ones or zeros for diagnosis or not of cardiac arrhythmia. The x_{data} dictionary was a bit more complicated. For each selected biosignal, an individual dictionary is made with either zeros and ones for binary biosignals or median values for non-binary biosignals. These dictionaries are concatenated per patient to make a multi-value dictionary, which is also eventually converted into a Num-Py array.

4.3.1. *Splits*

The final step before training is the splitting stage. Splitting data is an important aspect of machine learning and is used to verify that the model generalizes well to new data. Splitting allows us to test on data that the model has not yet seen, which in turn prevents overfitting.

In the splitting stage, the data is split into three categories: train, validate, and test. The data is split stratified to avoid a large imbalance in the distribution of the target class (Pedregosa et al., 2019). In this work, we split the data into the three splits, train, validate, and test, with respective weights of 0.7, 0.2, and 0.1.

4.4. Hyperparameter optimization

The choice of hyperparameters, such as the maximum number of trees in the random forest, significantly affects the model's runtime during training and its performance on the validation set. The goal in optimizing hyperparameters is to find the approximately optimal values for hyperparameters, such as the maximum number and depth of trees in the forest, that will yield the best results. In this work, the fine-tuning of hyperparameters is only done when the "split" program argument is set to "validate". When configuring multiple sets of hyperparameters, the Area Under the Curve, or AUC metric, was used to judge the quality of the tested hyperparameters. The hyperparameters that maximized the AUC score should then used for testing. However, each model could have its own hyperparameters, of which there are also many. In order to efficiently find appropriate hyperparameters, we compared grid Search and genetic algorithms. Grid Search implemented in Python using scikit-learn GridSearchCV() function whereas, TPOT auto machine learning library was used to implement genetic algorithm. For both implementation We performed a 5-fold cross validation on the validation set to select model's hyperparameters. The important aspect of hyperparameter optimization is finding the optimal number of trees in the forest and the optimal depth of each of those trees. For each model, the number and depth of trees was tested with GridSearchCV and Tpot functions. As a rule, the greater number of trees the more accurate the model becomes, however one should also be wary overfitting. As seen in Fig. 6, the model's performance peaks at 64 trees.

4.5. Testing

The finished model tested 16 biosignals, the selection of which is described in Section 3.1.2, and a variety of combinations of those 16. However, in order to test all possible combinations, we would have had to do: $\sum_{k=1}^{n=16}\binom{n}{k} = 65,535$ tests. This is neither time-efficient nor useful, so instead we used the following method to determine which combinations to test. Upon running the first test, it was discovered that some features had importances of exactly 0.0 because no patient in the database had a recording for those features; they were subsequently stripped from further tests. We ran tests with 24-, 12-, and 6-hour time frames. An initial test was run with all features with non-zero importances for each time frame. The combination for each subsequent run was selected by systematically removing the biosignal with the highest feature importance. In repeatedly removing the most important feature, we were eventually left with a set of relatively useless features and tests were cut off after the AUC score fell under .530. A few other tests were run to collect other data, as seen in the Results Section 5

5. Results

5.1. Comparison between hyperparameter optimization techniques

As discussed in Section 4.4, we implemented and compared grid search and genetic algorithm techniques for optimizing random forest model. From Figs. 6 and 7, it can be seen that both the techniques are performed well in optimizing the model. When the model tested with 19 biosignals, the maximum AUC scores obtained were 0.9775 (grid search) and 0.9686 (genetic algorithm) using maximum number of 64 trees as in Fig. 6. The results obtained, are highly dependent on the chosen grid space and dataset used. Therefore, in different situations, different optimization techniques will perform better than others.

Fig. 7 shows the total number of tree splits i.e the depth of the tree; model peaks at a depth of approximately 25.

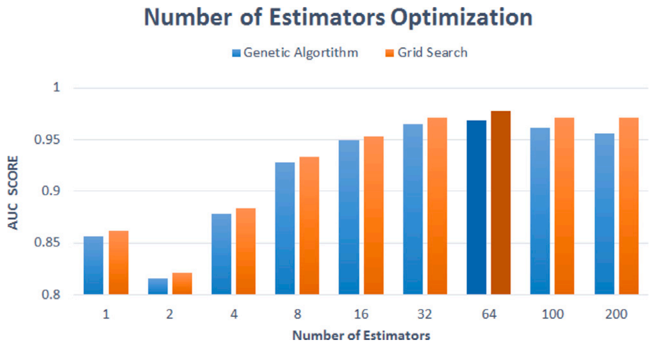

Fig. 6. Graphic depicts results from validation run with 19 biosignals, where a maximum number of trees of 64 begets an AUC scores of 0.9775 and 0.9686 for grid search and genetic algorithm techniques respectively.

5.2. Metric scores for different time frame

The following five features, Terfenadine, Quinidine, Astemizole, Epilepsy, and Dyspnea, had feature importances of exactly 0.0 and were removed after the first test because they only confused the model with useless data (Pollard & Johnson, 2016). The model's AUC score increased once these features were removed.

With only the five most predictive features, calcium, sodium, blood pressure, potassium, and respiratory rates, the model still achieved an AUC score 0.96797 on the 24 h model (Pollard & Johnson, 2016). As seen in Tables 3, 4, and 5, the model achieved AUC scores of 0.979, 0.962, and 0.947 with times frames of 24, 12, and 6 h respectively when all of the features with non-zero importances were included.

As seen in Figs. 8, 9, 10, the feature importances varied slightly depending on the length of the time frame, but generally feature importances remained within a margin of error.

As is evident from Tables 3, 4, and 5, the model predominantly used the same five features to split nodes: Sodium, Potassium, Calcium, Respiratory Rate, and Blood Pressure. The difference in the median values for those features between patients diagnosed with arrhythmia and those not diagnosed is graphically depicted in Fig. 11. P-values were calculated to better understand the statistical significance of the results displayed in Fig. 11. To better understand these values, most authors refer to statistically significant as a *p*-value of less than 0.05 and highly statistically significant as *p*-value of less than 0.001 (SciPy Organization, 2019). The following two-tailed p-values were calculated from a t-test, a two-sided test for the null hypothesis that two independent samples have identical expected averages (SciPy Organization, 2019). The p-values for calcium, sodium, potassium, respiratory rate, and blood pressure were 7.84×10^{-37}, 0.01, 3.53×10^{-14}, $3.47\times$, 10^{-39}, and 5.85×10^{-88} respectively (Pollard & Johnson, 2016). The differences between all of these features were statistically significant.

To contextualize Fig. 11, we discuss typical values for each of these features. According to the American Heart Association, a blood pressure[2] of above 80 means that a patient has Stage 1 Hypertension (American Heart Association, 2017). While the median blood pressure of a patient with arrhythmia is not over 80, it is closer than those without arrhythmias. A normal respiratory rate falls within 12 to 20 breaths per minute (American Heart Association, 2017). In adults, the healthy range of potassium is from 3.5 to 5.1 mmol/L; A serum potassium concentration greater than the upper limit of the normal range is dangerous and can quickly cause lethal cardiac arrhythmia (Lerma et al., 2018). The healthy reference ranges for serum sodium and serum calcium are from 136 to 145 mmol/L and from 8.9 to 10.1 mg/dL respectively (Lerma et al., 2018). It should be noted that the vital signs and serum values of other patients in the ICU are also likely different from that of the average healthy person.

[2] Diastolic Blood Pressure.

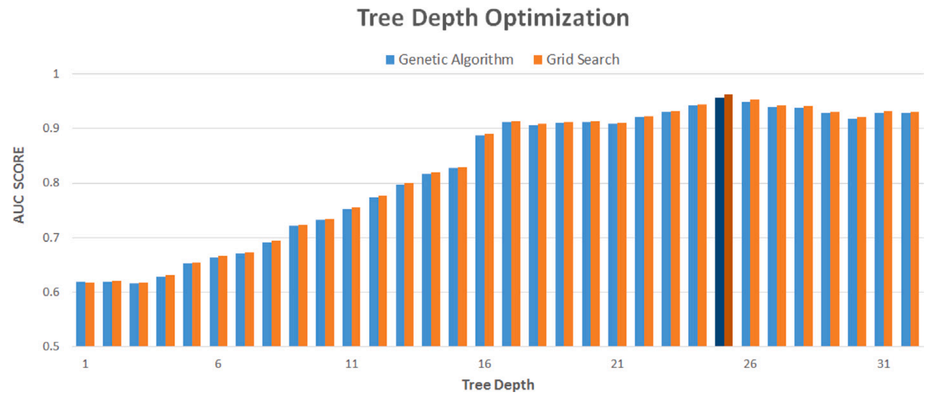

Fig. 7. Graphic depicts results from validation run with 14 biosignals, where a maximum depth of 25 begets an AUC scores of 0.9634 and 0.9571for grid search and genetic algorithm techniques respectively.

Table 3
This table shows the score results when the model was run with a time frame of 24 hours (Pollard & Johnson, 2016). After each test, the most important feature in that test was removed before running a new test. For example, "Sodium" denotes that a patient's sodium level removed from the set of features from the previous line.

Metric scores for 24 h time frame

	AUC	F1	Average precision	Recall	Accuracy	Hamming loss	Specificity	Sensitivity
All	0.9768	0.9664	0.9433	0.9681	0.9803	0.01967	0.9854	0.9681
Unimportant factors*	0.9787	0.9687	0.9466	0.9717	0.9816	0.0184	0.9857	0.9716
Sodium	0.9385	0.9247	0.8882	0.8944	0.9557	0.0443	0.9825	0.8944
Potassium	0.8028	0.7461	0.6851	0.6350	0.8686	0.1314	0.9706	0.6350
Blood pressure	0.6830	0.5433	0.5052	0.4140	0.7885	0.2115	0.9520	0.4140
Respiratory rates	0.5753	0.2908	0.3786	0.1827	0.7293	0.2707	0.9678	0.1827
Calcium	0.5386	0.1885	0.3352	0.1124	0.7058	0.2942	0.9649	0.1124
Lung disease	0.5307	0.1583	0.3281	0.0920	0.7027	0.2973	0.9693	0.09120
Heart failure	0.5002	0.0009	0.3042	0.0004	0.6963	0.3037		

Terfenadine, Epilepsy, Quinidine, Astemizole, Dyspnea.

Table 4
This table shows the score results when the model was run with a time frame of 12 hours (Pollard & Johnson, 2016). After each test, the most important feature in that test was removed before running a new test. For example, "Sodium" denotes that a patient's sodium level removed from the set of features from the previous line.

Metric scores for 12 h time frame

	AUC	F1	Average precision	Recall	Accuracy	Hamming loss	Specificity	Sensitivity
All	0.9624	0.9451	0.9077	0.9523	0.9664	0.0336	0.9725	0.9523
Potassium	0.9319	0.9095	0.8588	0.8965	0.9458	0.05423	0.9673	0.8965
Sodium	0.7957	0.7411	0.7036	0.5992	0.8728	0.1272	0.9922	0.5992
Blood pressure	0.6653	0.5059	0.4920	0.3663	0.7826	0.2174	0.9643	0.3663
Calcium	0.5519	0.2287	0.3494	0.1397	0.7136	0.2864	0.9641	0.1397
Respiratory rates	0.5116	0.0622	0.3135	0.0328	0.6994	0.3006	0.9903	0.0328

Terfenadine, Epilepsy, Quinidine, Astemizole, Dyspnea.

Table 5
This table shows the score results when the model was run with a time frame of 6 hours (Pollard & Johnson, 2016). After each test, the most important feature in that test was removed before running a new test. For example, "Sodium" denotes that a patient's sodium level removed from the set of features from the previous line.

Metric scores for 6 h time frame

	AUC	F1	Average precision	Recall	Accuracy	Hamming loss	Specificity	Sensitivity
All	0.95	0.91	0.85	0.94	0.95	0.05	0.95	0.94
Potassium	0.92	0.88	0.81	0.88	0.93	0.07	0.95	0.88
Blood pressure	0.85	0.80	0.70	0.76	0.89	0.11	0.94	0.76
Sodium	0.71	0.59	0.54	0.46	0.82	0.18	0.97	0.46
Calcium	0.5676	0.2722	0.3454	0.1725	0.7391	0.2609	0.9626	0.1725
Respiratory rates	0.5107	0.0597	0.2914	0.0316	0.7188	0.2812	0.9899	0.0316

Terfenadine, Epilepsy, Quinidine, Astemizole, Dyspnea.

6. Discussion and limitations

The fact that Terfenadine, Quinidine, and Astemizole had no influence is likely because their potential to cause arrhythmias has long been known and they were therefore not prescribed to patients in the MIMIC database (Keating & Sanguinetti, 2001; Pollard & Johnson, 2016; Sanguinetti & Tristani-Firouzi, 2006).

As is seen in Figs. 8, 9, and 10, the top five features are significantly more predictive than any of the others. It should be noted, that the Clinical Guidelines do not suggest to practicing help. Features outside of the top five are those given in the Al-Khatib et al. (2018) and Jogla

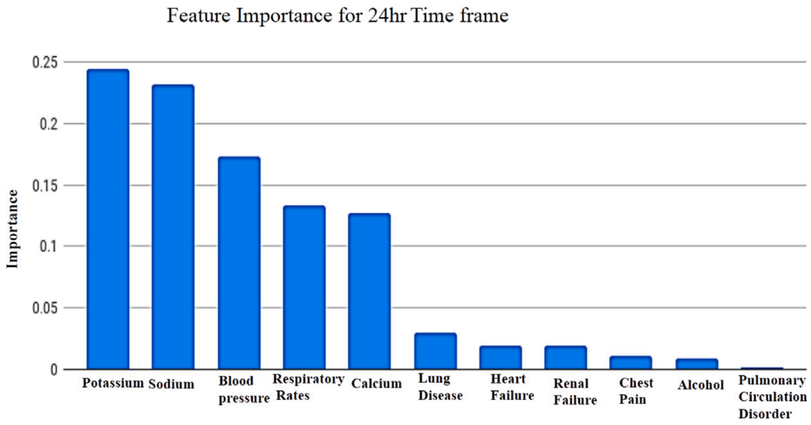

Fig. 8. This histogram depicts the distribution of feature importance when the time frame was 24 h.

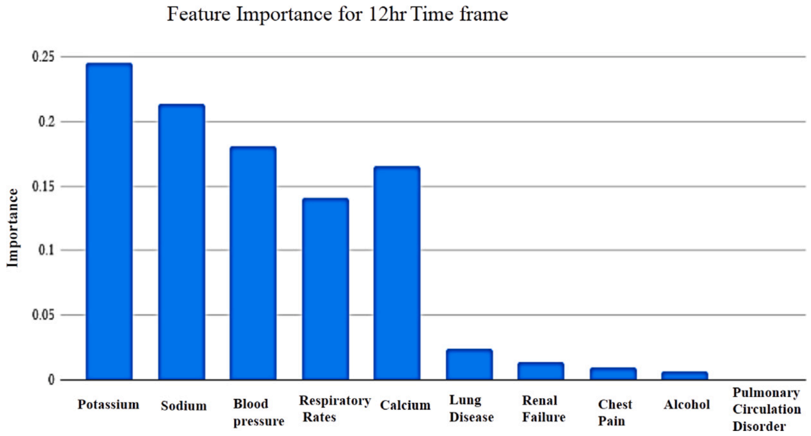

Fig. 9. This histogram depicts the distribution of feature importance when the time frame was 12 h.

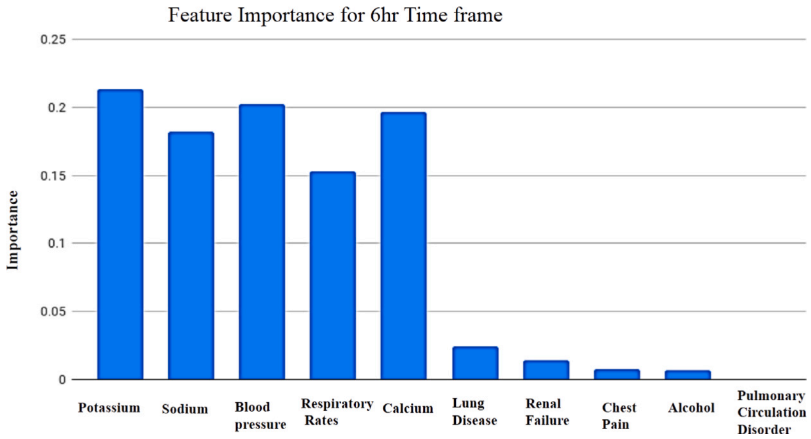

Fig. 10. This histogram depicts the distribution of feature importance when the time frame was 6 h.

et al. (2015). However, the features that were most important, Sodium, Potassium, Calcium, Blood Pressure, and Respiratory Rates, are also those that were most commonly observed in patients. The numbers of patients with data for serum potassium, serum sodium, serum calcium, respiratory rate, and blood pressure were 34,999, 34,843, 13,597, 26,166, and 26,146 respectively. In contrast, the numbers of patients with lung disease, renal failure, alcohol abuse history, and chest pain were (5200), (2139), (1224), and (1094) respectively. The greater quantity of patients with data for serum potassium, serum sodium, serum calcium, respiratory rate, and blood pressure could have skewed feature importances. Reformat so that numbers are in parenthesis next to features names.

It is also possible that effects of serum potassium, sodium and calcium were exaggerated by the Clinical Guidelines' recommendations to in some cases treat patients by repleting a patient's blood with these serums (Al-Khatib et al., 2018). Specifically, for patients diagnosed with Torsades des Pointes, the Clinical Guidelines recommend using potassium or magnesium repletion to treat patients (Al-Khatib et al., 2018). Furthermore, the Clinical Guidelines state that treating patients with Premature Ventricular Contractions with sodium channel blockers, which are used to treat some other types of arrhythmias, actually

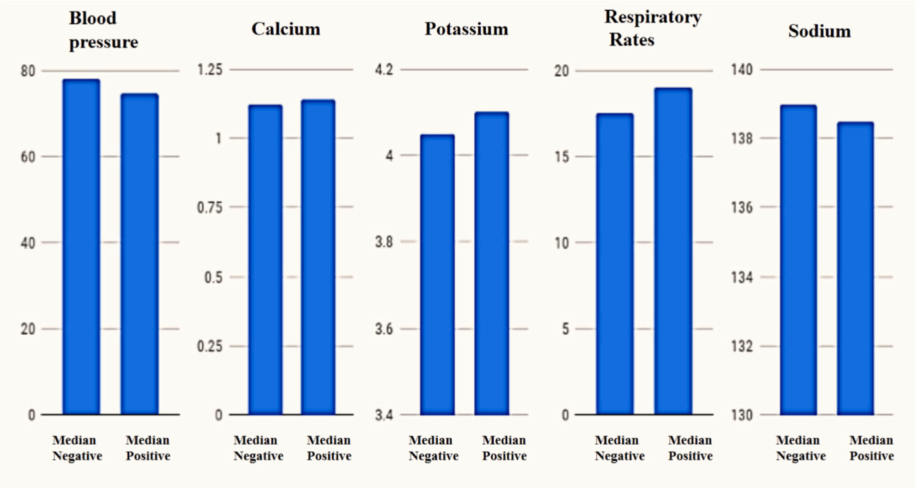

Fig. 11. Comparison of median blood pressure, calcium, potassium, respiratory rate, and sodium values, where positive represents patients diagnosed with arrhythmia and negative represents those without (Pollard & Johnson, 2016).

increases a patient's risk of death (Al-Khatib et al., 2018). Calcium channel blockers are also thought to be harmful in patients with complex tachycardias and the Clinical Guidelines suggest that increasing a patient's serum calcium levels may be helpful in treating tachycardias (Jogla et al., 2015). Given that, in some cases, the repletion of serum potassium, sodium, and calcium is recommended, it is possible that the serums that were injected as treatment for these diagnoses were recognized by the model as causes of the diagnosis. Make clear that real issue is timing and we do not know when event/treatment is. "could be a result of the treatment rather than a cause of the diagnosis".

Additionally, many of the medications used for determining the time frame of data that should be used for a diagnosed patient were antiarrhythmic medications that act as sodium-, calcium-, and potassium-channel blockers. These medications are typically administered when a patient's sodium, potassium, or calcium levels are too high. Since we measured only the hours immediately before drug administration, it could be that during these hours patients had an especially high amount of serum sodium, potassium, or calcium in their blood, further exaggerating their importances.

The model's accuracy is also highly dependent on the amount and quality of input data (Singh et al., 2018). Specifically, the accuracy of the model is reliant on a reliable and consistent ground truth (Singh et al., 2018). A pitfall of using the MIMIC-III database is that only hospital patients are included which does not facilitate the comparison of patients diagnosed with arrhythmias against completely healthy people. The resulting discrepancies are difficult to identify and measure.

6.1. Performance comparison

As discussed in Section 2, many researcher worked on development of computer aided system based on machine learning algorithms. Some of them have also developed systems for classification(Binary as well as multi-class) of cardiac arrhythmias using different features set on different dataset. Table 6 shows comparison of the proposed methodology with some existing methods. While comparing these methods we have considered only those authors who have used features like various biosignals, medications, and preconditions for binary classification of cardiac arrhythmia. As per Table 6, it can be seen that our proposed methodology achieved highest accuracy of 98%. The optimization, training and classification time lasted for single sample are 6 h, 0.3040 s, and 0.7685 s respectively for our model.

7. Conclusion and future scope

In this research work, we quantified with Random Forest the predictive power of various biosignals, medications, and preconditions for the diagnosis of cardiac arrhythmias. We achieved an Area Under the Receiver Operating Curve (AUC) score of 0.98, comparable to the performances achieved in several of the published state-of-the-art studies. We substantiated claims that each of sodium, calcium, potassium, respiratory rates and blood pressure can be used for the early diagnosis of cardiac arrhythmias. Our experimental results suggest that machine learning approaches such as this one could in the future could aid in the diagnosis of cardiac arrhythmias. In a further study, we propose manipulating the threshold, to minimize Type 1 Error, or False Positives. The incorrect classification of a healthy patient is a particularly negative result, as the once healthy patient may become ill as a result of treatment for a disease they did not have. We would also recommend testing further features, such as magnesium.

CRediT authorship contribution statement

Samir S. Yadav: Conceptualization, Methodology, Software, Data creation, writing - original draft, Software, Validation, Writing - review & editing. **Shivajirao M. Jadhav:** Visualization, Investigation, Supervision.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Compliance with ethical standards

Funding

No funding was received for this study.

Research involving human participants and/or animals

This research paper does not contain any studies with human participants or animals performed by any of the authors.

Table 6
Comparison of the proposed methodology with some existing methods.

Sr. No	Author and year	Methodology	Accuracy (%)
1	Guvenir et al. (1997)	VF15 algorithm GA based VF15 algorithm	62 65
2	Gao et al. (2004)	Bayesian artificial neural network (ANN) classifier used a logistic regression model and the back-propagation algorithm	75
3	Polat et al. (2006)	Artificial immune recognition system (AIRS) with fuzzy weighted pre-processing	80.7
4	Lee et al. (2007)	Neural network with weighted fuzzy membership functions	81.32
5	Uyar and Gurgen (2007)	SVM with Gaussian kernel	76.1
6	Oveisi et al. (2011)	Tree-Structured Feature Extraction Using Mutual Information	68.5
7	Jadhav et al. (2012)	Modular neural network Generalized feedforward neural network Multilayer perceptron model	82 .22 82.35 86.67
8	Khare et al. (2012)	Spearman Rank with Correlation,PCA and SVM	85.98
9	Yılmaz (2013)	Fisher Score and Least Squares-SVM	82.09
10	Jadhav et al. (2014)	Feature elimination based random subspace ensembles learning	91.11
11	Sean et al. (2017)	SVM FDR+SVM PCA+SVM Deep Neural Network FDR+Deep Neural Network PCA+Deep Neural Network	77 .77 78.23 76.97 79.18 80.64 73.65
12	Han (2017)	Cascade architectures of fuzzy neural net After gradient decent method	82.5
13	Ayar and Sabamoniri (2018)	Genetic Algorithm and Decision Tree classifier (c4.5)	86.96
14	Kadam et al. (2018)	Soft-Margin SVM, Feature Selection using Improved Elitist GA and SVM fitness function	87.83
15	**Proposed method**	**Random forest model with feature optimization using a Grid search or genetic algorithm**	**98**

References

Al-Khatib, S. M., Stevenson, W. G., Ackerman, M. J., Bryant, W. J., Callans, D. J., Curtis, A. B., Deal, B. J., Dickfeld, T., Field, M. E., Fonarow, G. C., Gillis, A. M., Granger, C. B., Hammill, S. C., Hlatky, M. A., Jogla, J. A., Kay, G. N., Matlock, D. D., Myerburg, R. J., & Page, R. L. (2018). 2017 AHA/ACC/HRS guideline for management of patients with Ventricular arrhythmias and the prevention of sudden cardiac death: executive summary. *Journal of the American College of Cardiology*, *72*.

American College of Cardiology (2019). Apple heart study identifies AFib in small group of apple watch wearers. Available at: https://www.acc.org/latest-in-cardiology/articles/2019/03/08/15/32/sat-9am-apple-heart-study-acc-2019.

American Heart Association (2017). Heath topics. https://www.heart.org/en/health-topics/high-blood-pressure/understanding-blood-pressure-readings.

(2019). ATPases. Available at: https://www.tocris.com/pharmacology/atpases.

Ayar, M., & Sabamoniri, S. (2018). An ECG-based feature selection and heartbeat classification model using a hybrid heuristic algorithm. *Informatics in Medicine Unlocked*, *13*, 167–175.

Breiman, L., & Cutler, A. (2002). Random forests. *Machine Learning Journal*.

(2013). Cardiac Arrhythmias. Available at: https://www.textbookofcardiology.org/wiki/Cardiac_Arrhythmias.

Center for Disease Control and Prevention (2015). International classification of diseases, ninth revision (ICD-9). https://www.cdc.gov/nchs/icd/icd9.htm.

Choi, E., Bahadori, M. T., Schuetz, A., Stewart, W. F., & Sun, J. (2016a). Doctor AI: Predicting clinical events via recurrent neural networks. *Journal of Machine Learning Research (JMLR)*, *56*.

Choi, E., Schuetz, A., Stewart, W. F., & Sun, J. (2016b). Using recurrent neural network models for early detection of heart failure onset. *Journal of the American Medical Informatics Association*, *24*.

Churpek, M. M., Adhikari, R., & Edelson, D. P. (2016). *The value of vital sign trends for detecting clinical deterioration on the wards*. Elsevier.

Cretikos, M. A., Bellomo, R., Hillman, K., Chen, J., Finfer, S., & Flabouris, A. (2002). Respiratory rate: the neglected vital sign. *Medical Journal of Australia*, *188*.

F.Fieselmann, J., Hendryx, M. S., Helms, C. M., & Wakefield, D. S. (1992). Respiratory rate predicts cardiopulmonary arrest for internal medicine inpatients. *Journal of General Internal Medicine*, *8*, http://dx.doi.org/10.1007/bf02600071.

Foundation, P. S. (2018). Python language reference, version 2.7. Available at: https://docs.python.org/2.7/.

Gao, D., Madden, M., Schukat, M., Chambers, D., & Lyons, G. (2004). Arrhythmia identification from ECG signals with a neural network classifier based on a Bayesian framework. In *Proceedings of the 24th SGAI International Conference on Innovative Techniques and Applications of Artificial Intelligence*. Citeseer.

Ghassemi, M., Brimmer, N., Tristan Naumann, R. J., Szolovits, P., Doshi-Velez, F., & Rumshisky, A. (2014). Unfolding physiological state: Mortality modelling in intensive care units. *KDD*.

Ghassemi, M., Pimentel, M. A., Naumann, T., Brennan, T., Clifton, D. A., Szolovits, P., & Feng, M. (2015). A multivariate time modeling approach to severity of illness assessment and forecasting in ICU with sparse. In *Heterogeneous Clinical Data, AAAI Conference on Artificial Intelligence*.

Guvenir, H. A., Acar, B., Demiroz, G., & Cekin, A. (1997). A supervised machine learning algorithm for arrhythmia analysis. In *Computers in Cardiology 1997* (pp. 433–436). IEEE.

Han, C. (2017). Detecting an ECG arrhythmia using cascade architectures of fuzzy neural networks. *Advanced Science and Technology Letters*, *143*, 272–275.

Huang, H. (2018). Multiple hypothesis testing and false discovery rate. Available at: https://www.stat.berkeley.edu/~hhuang/STAT141/Lecture-FDR.pdf.

Huikiri, H. V., Castellanos, A., & Myerburg, R. J. (2001). Sudden death due to cardiac arrhythmias. *New England Journal of Medicine*, *345*.

Jadhav, S. M., Nalbalwar, S. L., & Ghatol, A. A. (2012). Artificial neural network models based cardiac arrhythmia disease diagnosis from ECG signal data. *International Journal of Computer Applications*, *44*(15), 8–13.

Jadhav, S., Nalbalwar, S., & Ghatol, A. (2014). Feature elimination based random subspace ensembles learning for ECG arrhythmia diagnosis. *Soft Computing*, *18*(3), 579–587.

Jogla, J. A., Page, R. L., Caldwell, M. A., Calkins, H., Conti, J. B., Deal, B. J., III, N. M. E., Field, M. E., Goldberger, Z. D., Hammill, S. C., Indik, J. H., Lindsay, B. D., Olshansky, B., Russo, A. M., Shen, W. -K., Tracy, C. M., & Al-Khatib, S. M. (2015). 2015 AHA/ACC/HRS guideline for management of patients with supraventricular tachycardia. *Journal of the American College of Cardiology*, http://dx.doi.org/10.1161/CIR.0000000000000311.

Johnson, A. E. W., Pollard, T. J., Shen, L., Lehman, L., Feng, M., Ghassemi, M., Moody, B., Szolovits, P., Celi, L. A., & Mark, R. G. (2016). MIMIC-III, a freely accessible critical care database.. http://dx.doi.org/10.1038/sdata.2016.35.

Kadam, V. J., Yadav, S. S., & Jadhav, S. M. (2018). Soft-margin SVM incorporating feature selection using improved elitist GA for arrhythmia classification. In *International Conference on Intelligent Systems Design and Applications* (pp. 965–976). Springer.

Kandala, R. N., Dhuli, R., Pławiak, P., Naik, G. R., Moeinzadeh, H., Gargiulo, G. D., & Gunnam, S. (2019). Towards real-time heartbeat classification: evaluation of nonlinear morphological features and voting method. *Sensors*, *19*(23), 5079.

Keating, M. T., & Sanguinetti, M. C. (2001). Molecular and cellular mechanisms of cardiac arrhythmias. *Cell*, *104*.

Khare, S., Bhandari, A., Singh, S., & Arora, A. (2012). Ecg arrhythmia classification using spearman rank correlation and support vector machine. In *Proceedings of the International Conference on Soft Computing for Problem Solving, SocProS 2011 December 20-22, 2011* (pp. 591–598). Springer.

Koehrson, W. (2018). An implementation and explanation of the random forest in Python. https://towardsdatascience.com/an-implementation-and-explanation-of-the-random-forest-in-python-77bf308a9b76.

Lee, S. -H., Uhm, J. -K., & Lim, J. S. (2007). Extracting input features and fuzzy rules for detecting ECG arrhythmia based on NEWFM. In *2007 International Conference on Intelligent and Advanced Systems* (pp. 22–25). IEEE.

Lerma, E. V., Simon, E. E., & Sofronescu, A. G. (2018). Laboratory medicine. https://reference.medscape.com/guide/laboratory_medicine.

Microsoft Azure (2019). How to choose algorithms for azure machine learning studio. Available at: https://docs.microsoft.com/en-us/azure/machine-learning/studio/algorithm-choice.

Moody, G. B., & Mark, R. G. (2001). The impact of the MIT-BIH arrhythmia database. *IEEE Engineering in Medicine and Biology Magazine*, *20*(3), 45–50.

Narkhede, S. (2018). Understanding AUC-ROC curve. https://towardsdatascience.com/understanding-auc-roc-curve-68b2303cc9c5.

Oveisi, F., Oveisi, S., Erfanian, A., & Patras, I. (2011). Tree-structured feature extraction using mutual information. *IEEE Transactions on Neural Networks and Learning Systems*, *23*(1), 127–137.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, E. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research (JMLR)*, *12*, 2825–2830.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, E. (2019). Scikit-learn user guide. (pp. 149–648). Available at: https://scikit-learn.org/stable/_downloads/scikit-learn-docs.pdf.

Pinnell, J., Turner, S., & Howell, S. (2007). Cardiac muscle physiology. *Continuing Education in Anaesthesia Critical Care & Pain*, *7*, 85—88.

Pławiak, P. (2018a). Novel genetic ensembles of classifiers applied to myocardium dysfunction recognition based on ECG signals. *Swarm and Evolutionary Computation*, *39*, 192–208.

Pławiak, P. (2018b). Novel methodology of cardiac health recognition based on ECG signals and evolutionary-neural system. *Expert Systems with Applications*, *92*, 334–349.

Pławiak, P., & Abdar, M. (2020). Novel methodology for cardiac arrhythmias classification based on long-duration ECG signal fragments analysis. In *Biomedical Signal Processing* (pp. 225–272). Springer.

Pławiak, P., & Acharya, U. (2019). Novel deep genetic ensemble of classifiers for arrhythmia detection using ECG signals. *Neural Computing and Applications*, 1–25.

Polat, K., Şahan, S., & Güneş, S. (2006). A new method to medical diagnosis: Artificial immune recognition system (AIRS) with fuzzy weighted pre-processing and application to ECG arrhythmia. *Expert Systems with Applications*, *31*(2), 264–269.

Pollard, T. J., & Johnson, A. E. W. (2016). The MIMIC-III clinical database. http://dx.doi.org/10.13026/C2XW26.

Remme, C. A., & Bezzina, C. R. (2010). Sodium channel (dys)function and cardiac arrhythmias. *Cardiovascular Therapeutics*, *28*, http://dx.doi.org/10.1111/j.1755-5922.2010.00210.x.

Sanguinetti, M. C., & Tristani-Firouzi, M. (2006). hERG potassium channels and cardiac arrhythmia. *Nature*, *440*.

SciPy Organization (2019). Numpy reference guide, version 1.16. Available at: https://docs.scipy.org/doc/numpy/about.html.

Sean, S. X., Mak, M. -W., & Cheung, C. -C. (2017). Deep neural networks versus support vector machines for ECG arrhythmia classification. In *2017 IEEE International Conference on Multimedia & Expo Workshops (ICMEW)* (pp. 127–132). IEEE.

Singh, S. (2018). Understanding the bias-Variance tradeoff. https://towardsdatascience.com/understanding-the-bias-variance-tradeoff-165e6942b229.

Singh, G., Al'Aref, S. J., Assen, M. V., Kim, T. S., van Rosendael, A., Kolli, K. K., Dwivedi, A., Maliakal, G., Pandey, M., Wang, J., Do, V., Gummalla, M., Cecco, C. D., & Min, J. K. (2018). Machine learning in cardiac CT: Basic concepts and contemporary data. *Journal of Cardiovascular Computed Tomograph*, http://dx.doi.org/10.1016/j.jcct.2018.04.010.

Splawski, I., Timothy, K. W., Decher, N., Kumar, P., Sachse, F. B., Beggs, A. H., Sanguinetti, M. C., & Keating, M. T. (2005). Severe arrhythmia disorder caused by cardiac L-type calcium channel mutations. *Proceedings of the National Academy of Sciences of the United States of America*, *102*.

Splawski, I., Timothy, K. W., Tateyama, M., Clancy, C. E., Malhotra, A., Beggs, A. H., Cappuccio, F. P., Sagnella, G. A., Kass, R. S., & Keating, M. T. (2017). Variant of SCN5A sodium channel implicated in risk of cardiac arrhythmia. *Science*.

(2019). SQLite Release (3.27.2), https://www.sqlite.org/changes.html.

Tahsildar, S. (2019). Random forest algorithm in trading using Python. https://blog.quantinsti.com/random-forest-algorithm-in-python/.

The Cleveland Clinic (2019). Sudden cardiac death. https://my.clevelandclinic.org/health/diseases/17522-sudden-cardiac-death-sudden-cardiac-arrest.

Tuncer, T., Dogan, S., Pławiak, P., & Acharya, U. R. (2019). Automated arrhythmia detection using novel hexadecimal local pattern and multilevel wavelet transform with ECG signals. *Knowledge-Based Systems*, *186*, Article 104923.

Uyar, A., & Gurgen, F. (2007). Arrhythmia classification using serial fusion of support vector machines and logistic regression. In *2007 4th IEEE Workshop on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications* (pp. 560–565). IEEE.

Yılmaz, E. (2013). An expert system based on Fisher score and LS-SVM for cardiac arrhythmia diagnosis. *Computational and mathematical methods in medicine*, *2013*.