IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received June 5, 2019, accepted June 20, 2019, date of publication July 1, 2019, date of current version July 31, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2925847

An Automated Strategy for Early Risk Identification of Sudden Cardiac Death by Using Machine Learning Approach on Measurable Arrhythmic Risk Markers

DAKUN LAI[1], (Member, IEEE), YIFEI ZHANG[1], XINSHU ZHANG[1], YE SU[2], AND MD BELAL BIN HEYAT[1]

[1]School of Electronic Science and Engineering, University of Electronic Science and Technology of China, Chengdu 610054, China
[2]Department of Cardiovascular Ultrasound and Cardiology, Sichuan Academy of Medical Sciences & Sichuan Provincial People's Hospital, Chengdu 610072, China

Corresponding author: Dakun Lai (dklai@uestc.edu.cn)

This work was supported by the National Nature Science Foundation of China under Grant 61771100.

ABSTRACT Early risk identification of an unexpected sudden cardiac death (SCD) in a person who is suffering malignant ventricular arrhythmias is highly significant for timely intervention and increasing the survival rate. For this purpose, we have presented an automated strategy for prediction of SCD with a high-level accuracy by using measurable arrhythmic markers in this paper. The set of arrhythmic parameters includes three repolarization interval ratios, such as T_pT_e/QT, JT_p/JT_e, and T_pT_e/JT_p and two conduction-repolarization markers, such as T_pT_e/QRS and $T_pT_e/(QT\times QRS)$. Each of them is calculated directly from the detected QRS complex waves and T-wave of electrocradiogram (ECG) signals. Then, all calculated markers are used for the automatical classification of normal and SCD risk groups by employing machine learning classifiers, such as k-nearest neighbor (KNN), decision tree (DT), Naive Bayes (NB), support vector machine (SVM), and random forest (RF). The effectiveness and usefulness of the proposed method is evaluated using a database of measured ECG data acquired from 28 SCD and 18 normal patients. For the automated strategy, the set of five arrhythmic risk markers can predict SCD in less than one second with an average accuracy of 98.91% (KNN), 98.70% (SVM), 98.99% (DT), 97.46% (NB), and 99.49% (RF) for 30 minutes before the occurrence of SCD. Moreover, a practical and straightforward SCD index (SCDI) through a judicious integration of these markers is also proposed by using the Student's t-test. The obtained SCDIs are 1.2058 ± 0.0795 and 1.7619 ± 0.1902 for normal and SCD patients, respectively, which provide a sufficient discrimination degree with a p-value of 6.5061e-35. The present results show that both the automated classifier and the integrated SCDI can predict the SCD up to 30 minutes earlier, and that these predictions could be more practical and efficient if applied in portable smart devices with real-time requirements in hospital settings or at home.

INDEX TERMS Arrhythmic risk markers, electrocardiogram (ECG), machine learning, sudden cardiac death (SCD), SCD prediction.

I. INTRODUCTION

Sudden cardiac death (SCD) is defined as death due to cardiovascular causes in a patient with or without known preexisting heart disease, in whom the mode and time of death are unexpected [1], [2]. The generally accepted temporal definition is bracketed by a period of up to 1 hour between the onset of an abrupt change in clinical status and loss of consciousness [2]. It is well known that SCD is the manifestation of a fatal heart rhythm disorder, such as ventricular tachycardia (VT) and

The associate editor coordinating the review of this manuscript and approving it for publication was Yonghong Peng.

ventricular fibrillation (VF) [3], [5], or a severe bradyarrhythmia [6]. These arrhythmias often lead to sudden cardiac arrest (SCA), which renders the heart unable to pump out the blood effectively [7]. Then, unattended SCA leads to SCD. Most malignant ventricular arrhythmias are caused by coronary heart disease, cardiomyopathy, valvular diseases or genetically determined disorders, and if not accurately diagnosed nor treated, immediate death occurs [1], [8], [9]. Several studies have shown that the development of effective targeted therapeutic interventions such as implantable cardioverter defibrillators (ICD) reduce SCD mortality [10], [12]. However, their cost-effectiveness is limited because of a relatively small number of patients receiving appropriated ICD shocks during follow-up. On the other hand, most sudden deaths occur in individuals who do not have high-risk profiles [13]. Therefore, much attention has recently been drawn to the public access procedure with an automatic electrical defibrillation (AED) as a way of rescuing patients without an ICD from impending death after cardiac arrest [14]. Nevertheless, even in the United Stated where public AEDs are readily available in populated area nationwide and there have been extensive improvements in resuscitation methodology and deployment of first responder systems, average survival from SCA remains below 5 [15]. As such, an early risk identification of an unexpected SCD in a person that is suffering malignant ventricular arrhythmias is highly significant for timely intervention and increasing the survival rate.

Studies have demonstrated that an underlying electrophysiological substrate, representing dispersion of refractoriness, and increased sympathetic tone in the ventricles of the heart are critical for the arrhythmic SCD [16], [17]. In the past few years, research worldwide has focused on this severe health problem with the goal of developing an efficient way of predicting the risk of SCD using invasive and non-invasive techniques, including electrophysiological testing [14], invasive hemodynamic evaluation [18], left ventricular ejection fraction (LVEF) [19], and non-invasive electrocradiogram (ECG) [20], [21]. In comparison with the first three invasive or imaging modalities mentioned above, ECG is an inexpensive and readily available method commonly used by physicians in clinical practice, which is a recording of the electric potential, generated by the electric activity of the heart, on the surface of the thorax. Recently, a systematic meta-analysis demonstrated that certain parameters on the ECG signal and other ECG-based investigations can provide important information on the underlying cardiac substrate abnormality that may predispose to ventricular arrhythmias and SCD. These include pathophysiological control mechanisms, mediated through autonomic nervous system functions, such as heart rate variability or turbulence (HRV, HRT) [22]–[26], measures of electrocardiographic conduction and repolarization interval [27]–[29], such as QRS (Q, R, and S wave in electrocardiogram) duration [30] and QT (Q and T wave) interval and dispersion [31], [32], and T-wave alternates [33], [34]. Among of them, HRV or HRT derived from the ECG signal is defined as evaluation of beat to beat variability of the R-R interval, which have been extensive studied for SCD detection and prediction. Many time domain [23], frequency domain [24] and nonlinear methods [25] have been proposed to reveal the inherent features of an HRV signal for the purpose of detection and prediction of SCD, which is mainly served through feature selection in different processing domains and subject classification by means of classifiers that are based on the extracted features. HRV or HRT initially demonstrated promise but was later shown not be predictive of arrhythmic mortality [35], [37]. There is still debate regarding the best timing for performance of HRV measurements. The yield of HRV measurement in the first days or weeks following myocardial infarction has been criticized, and poor predictive value of such measurement is reported [35]. The use of HRV to predict SCD risk in patients with coronary heart disease is less well established [36], [37], which also cannot be evaluated in other patients with atrial fibrillation or frequent arrhythmias [37]. In addition, the insufficient advance prediction time as reported from 2 min to 13 min would also limit the use of HRV for risk prediction of SCD in a clinical situation [23]–[25].

Meanwhile, in the last two decades a large number of clinical studies have discovered that it is feasible to predict the development of ventricular arrhythmias by analyzing and measuring several markers related electrocardiographic conduction and repolarization alterations [27]–[29], [31], [32]. Some of the most explored predictors in clinical practice are based on repolarization, including QT interval and its correction by heart rate [31], QT dispersion [32], interval from the peak to the end of the T-wave (T_pT_e) [29], and T-wave alternants (TWA) [30], [31]. Recently, novel repolarization interval ratios such as T_pT_e/QT, JT_p/JT_e [29, 29, 39, 40], and T_pT_e/JT_p and conduction-repolarization markers such as T_pT_e/QRS and T_pT_e/(QT×QRS) [28], [41] were proposed, and clinical findings suggested that abnormalities of repolarization and conduction should all be taken into consideration for accurate prediction of an individual's arrhythmic potential [20], [21]. Although these ECG based markers have been reported promising for indicating patients with a high risk to develop malignant ventricular arrhythmias, most work so far has been focused on clinical investigations and they lack of effective way to present pattern of various ECG parameters prior to the occurrence of SCD. Two most recent works reported attempts to automatically extract risk features directly from ECG signal by using complex wavelet transforms [38], [39], while the absence of electrophysiological marker, the insufficient advance prediction time no more than 12 min, and a higher computational requirement would all limit its applications in prediction of SCD. Therefore, there is a need for efficiently automated methods to ensure a practical prediction of SCD before establishing its clinical applicability.

In this study, our focus was on automatically indentifying early risk markers of sudden cardiac death by using enhanced machine learning techniques on measurable arrhythmic parameters. The set of arrhythmic risk parameters includes

the repolarization interval ratios such as T_pT_e/QT, JT_p/JT_e and T_pT_e/JT_p and conduction-repolarization markers such as T_pT_e/QRS and $T_pT_e/(QT\times QRS)$, which all are calculated directly from the detected QRS complex waves and T-wave of ECG signals. Then, these markers are used for classification of normal and SCD risk groups by employing automatical classifiers such as k nearest neighbor (KNN), decision tree (DT), Naive Bayes (NB), support vector machine (SVM) and random forest (RF). Moreover, the informative markers are selected from these calculated markers by using the Student's *t*-test, and a novel integrated sudden cardiac death index (SCDI) is presented through a judicious combination of informative markers. Compared to the prior strategies that used HRV signal or that derived non-clinical features from ECG signal, the present study provides a more practical and straightforward methodology for SCD prediction with measurable arrhythmic risk markers to achieve greater accuracy and longer prediction time without the need for significant additional complex transforms of ECG waveforms. We expect the presented method to be used widely in clinical practice, especially for real-time requirements and embedding on wearable devices in hospital settings or at home.

II. MATERIAL

A. DATASET USED

In the process of the SCD prediction, ECG data collection and pre-processing are the first important step. In this work, ECG signals were collected from three databases provided by AHA database (AHADB) from people at risk of SCD, the MIT-BIH Database [42] entitled Sudden Cardiac Death Holter (SDDB) from people at risk of SCD and Normal Sinus Rhythm (NSRDB) from normal people. Our SCD group consisted of 10 recordings with no. 8001-8010 from AHA database and 18 recordings from SDDB, for each ECG recording, the ECG signal of the first lead in the 30 minutes prior to VF onset was collected Noted that recordings with no. 40, 42, 49 in the SDDB were not used in this work because of the absence of VF onset, and two more recordings with no. 38 and 41 also were not included owing to unknown lead that is of a low amplitude R wave. Moreover, a normal group including 18 half-hour recordings was built from the database of NSRDB. As summarized in Table 1, the age of SCD group is of 60.65 $\pm$ 19.96 years (range, 30-89 years), and 34.33 $\pm$ 8.44 years (range, 20-50 years) for the normal group. Except the unknown gender of 11 SCD patients, the gender of the remaining 35 subjects is given, including 9 male and 8 female for the SCD group, and 5 male and 13 female for the normal group. Importantly, detailed descriptions of each recording including record name and number, length before the ventricular fibrillation onset, and related heart diseases and arrhythmia categories are given in Table 2. Most patients in the SCD group were found to be a malignant ventricular arrhythmias with an underlying cardiac substrate abnormality and heart disease. Moreover, one example of the collected ECG signal in 30 minutes and its corresponding 1 minutes prior to SCD, and one example of 1 minutes in sinus rhythm are illustrated in Fig. 1(A), Fig. 1(B), and Fig. 1 (C), respectively.

TABLE 1. The age and gender of the SCD and normal groups.

Groups	Total	Gender			Age	
		Male	Female	Unknown	Range	Mean ± SD
SCD	28	9	8	11	30-89	60.65 ± 19.96
Normal	18	5	13	0	20-50	34.33 ± 8.44

B. ECG PRE-PROCESSING

Noise reduction can greatly increase the precision and efficiency of prediction algorithm. In this work, All ECG data collected from a total of 46 recordings was pre-processed firstly. For each recording, a continuous signal segmentation with a one-minute length was carried out and 30 one-minute ECG fragments were built either into the SCD dataset or the normal dataset. Furthermore, the baseline drift and the noise of each ECG fragment were removed by using a median filter and a band-pass filter (0.5-100 Hz), respectively. As a summary, there are 840 one-minute fragments in the SCD group and 540 one-minute fragments in the normal group, which are used to train and test our automated strategy of SCD prediction and evaluate the proposed practical SCD index on the basis of the following ECG analysis and measurement.

III. METHODOLOGY

A. OVERVIEW

Taking the clinical considerations of the arrhythmic risk markers, an attempt of the VF caused SCD prediction at its earlier stage combing with five arrhythmic risk markers of ECG signal is presented in the present paper. The schematic diagram of the proposed methodology on measurable arrhythmias risk marker for SCD prediction is illustrated in Fig. 2. First, two ECG datasets were built with a series of signal pre-processing such as data segmentation and noise reduction as described above, which include one SCD group collected from the SDDB and AHADB and one normal group collected from the NSRDB. Second, an morphology based detection algorithm of the QRS-T waves was developed so as to accurately determine all key ECG parameters of each heart beat, such as Q onset, R peak, S offset, T peak, T offset, and various intervals. Third, five arrhythmic risk markers suggested by many clinical studies are extracted and calculated individually, including the T_pT_e/QT, the JT_p/JT_e, the T_pT_e/JT_p, the T_pT_e/QRS and the $T_pT_e/(QT\times QRS)$. Finally, automated prediction strategies with various machine learning methods were assessed and then compared with each other for a better classification between of the normal group and the SCD group, such as the KNN, the DT, the SVM, the RF and the NB. Furthermore, followed by ECG data preprocessing and QRS-T wave detection, a straightforward and

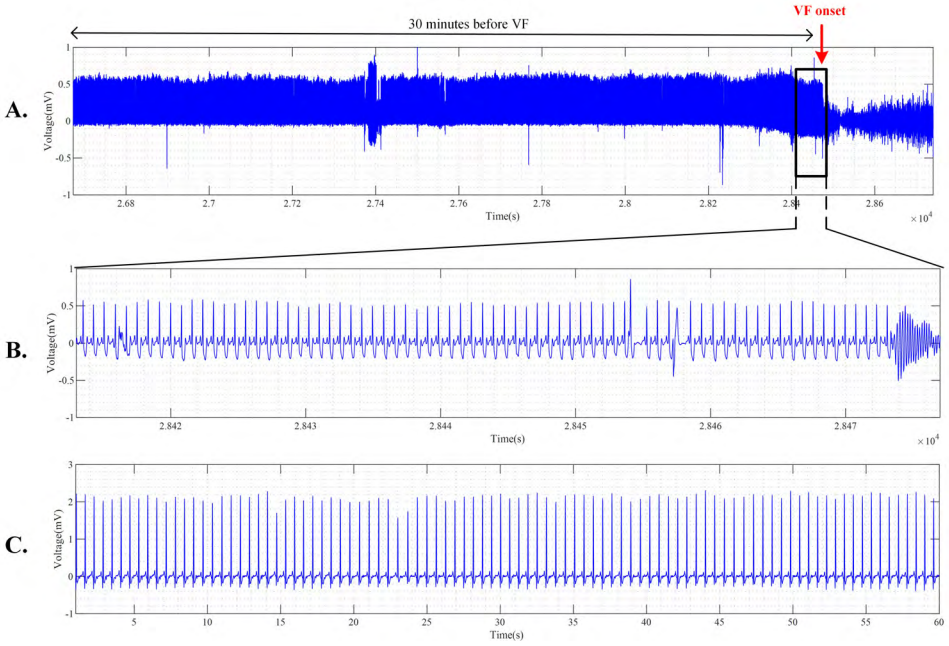


FIGURE 1. The time frame of two examples data of the built SCD group. (A). the first half hour in advance of a VF onset, (B). 1 minute before VF onset, and (C). 1 minute from normal ECG signal.

TABLE 2. A summary of the ECG recordings of SCD patients provided by MIT-BIH database and American Heart Association (AHA) databases.

Database	Heart diseases	Number of records	Record name	Length before the ventricular fibrillation onset	Arrhythmia categories
NSRDB	NA	18	Whole database	NA	Unknown
SDDB	Cardiac surgery	4	32,35,36,50	30 min	Ventricular tachycardia; ventricular fibrillation; ventricular flutter
	Coronary artery disease	1	43		
	Unknown	9	30,33,34,37,44,46,47,48,51		
	Heart failure	2	31,52		
	ventricular ectopy	1	45		
	Acute myelogenous leukemia	1	39		
AHADB	Unknown	10	8001-8010	30 min	Ventricular rhythms; ventricular fibrillation or ventricular flutter

Note: TP = true positive, TN = true negative, FP = false positive, FN = false negative, Se = sensitivity, Sp = specificity, Acc = accuracy.

practical method with an novel SCDI also was proposed for SCD prediction by using three informative markers selected from such five calculated arrhythmias markers as described above, which was generated in terms of a formulation with a series mathematical process, as shown in Fig. 2 and discussed in detail as following.

B. ECG SIGNAL ANALYSIS AND INTERVAL CALCULATION

ECG works on the principle of measuring the projection of the heart polarization vector. Because of the anatomical difference of the atria and the ventricles, their sequential activation, depolarization, and repolarization produce clearly differentiable deflections. Exploration of cardiac electrical activity is based on the search for unambiguous patterns describing beat-to-beat processes called P-wave, QRS complex and T-wave. The P-wave represents atrial depolarization, the ventricular depolarization causes the QRS complex, and repolarization is responsible for the T-wave. For the purpose of the extraction of SCD markers with physiological significance, a series of time-based ECG signal analysis algorithms were designed first for 5 ECG intervals in this study, such as QRS complex detection, T-wave delineation, and intervals calculations. For each heart beat, five points of interest are located including Q onset, R peak, S offset, T peak, and T offset, and five corresponding intervals are calculated including QT interval, QRS duration, JT_e interval, JT_p interval, and T_pT_e interval, as illustrated in Fig. 3(A).

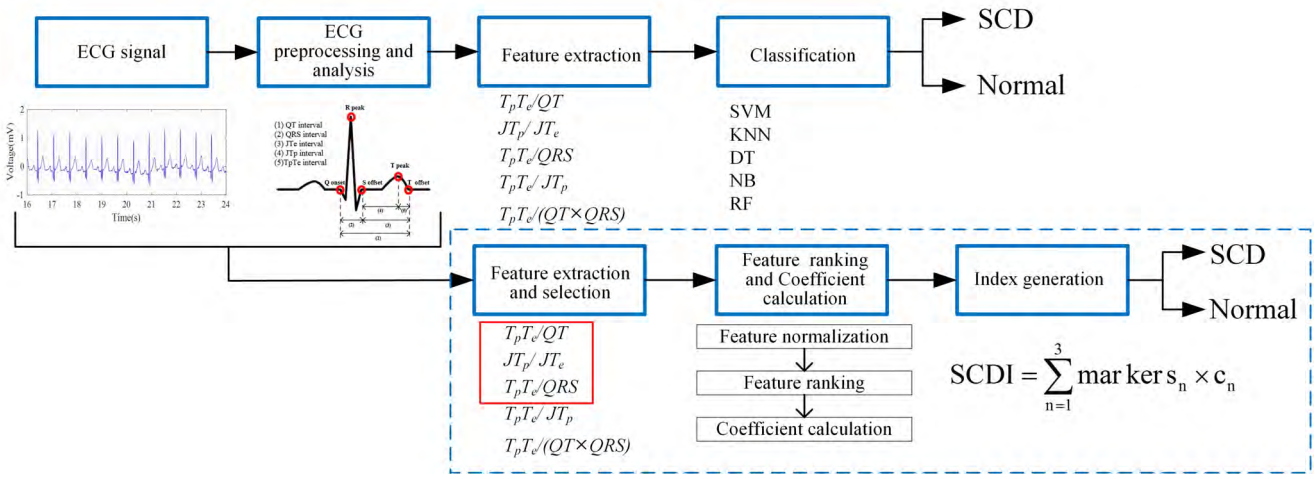


FIGURE 2. Schematic diagram of the proposed methodology on measurable arrhythmias risk marker for SCD prediction.

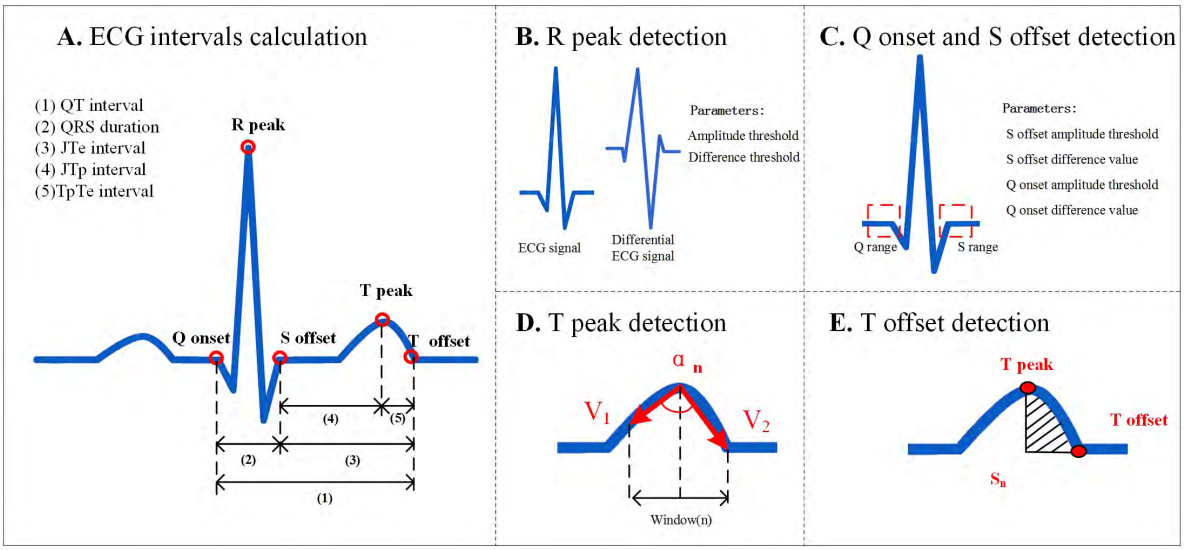


FIGURE 3. Schematic diagram of ECG signal analysis and measurement. (A). Five intervals calculation of ECG signal, (B, C) QRS complex wave detection, (D, E) T wave delineation.

1) QRS COMPLEX DETECTION

R wave is the most significant waveform on the electrocardiogram. So prior to other analysis, R peak is detected by using Pan-Tompking method [43], which is based on self-adapting amplitude threshold and difference threshold that extracted from ECG signal and differential ECG signal, respectively. The schematic diagram of R peak detection is shown in Fig. 3(B).

Once the position of R peak was detected, the detection range is determined. These detection windows are the range for detection Q onset and S offset whose detection algorithm are similar except the slightly difference in threshold. More specifically, these two detection algorithms are both carried out by a moving window with amplitude threshold (Q onset amplitude threshold and S offset amplitude threshold). S offset is taken as an example to explain the detection method. The amplitude threshold is provided by doctor's experience. The moving window moves in the range until it satisfied the condition that the maximum amplitude in the window should be higher than amplitude threshold. Then, a minimum amplitude is found to calculate the differentiation between it and the maximum amplitude. Thus, the least significant differentiation in the window whose location of the maximum amplitude is the S offset. The schematic diagram of Q onset and S offset detection is shown in Fig. 3(C).

2) T-WAVE DELINEATION

With calculations of ECG intervals of the proposed method, a ECG morphology based algorithm for T-wave delineation was developed. The schematic diagram of T peak and T offset detection are shown in Fig. 3(D) and Fig. 3(E), respectively. In detail, T peak detection utilizes the

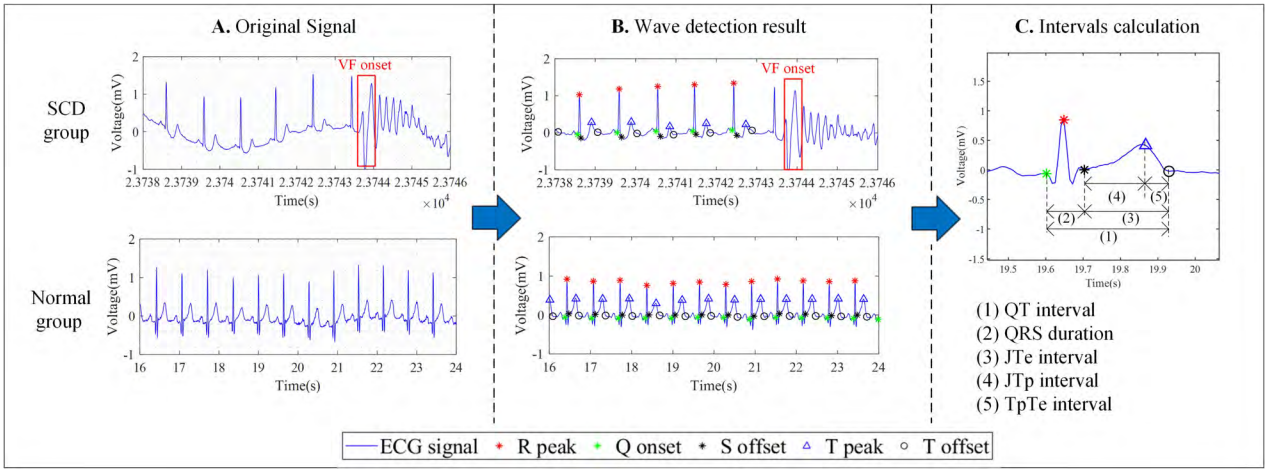


FIGURE 4. Examples of ECG processing of one normal segment and one SCD segment: (A). the original ECG siganl from SCD group and Normal group respectively, (B). the wave detection result of ECG siganl from SCD group and Normal group respectively, and (C). five intervals of one heart beat in an ECG signal.

morphological characteristics of T-wave, whose the searching range started from S offset, and last 0.7 times the length of current RR interval. In this area, a moving window moves following the rules of one point at a time from start to end (window(1), window(2)…window(n)), where window(n) contains two vectors ($\overrightarrow{V}_1$ and $\overrightarrow{V}_2$) with a common starting point. The point in the ECG signal corresponding to the midpoint of the window is taken as the common starting point of the two vectors. Thus, the point of (x_m, y_m) is defined as the coordinates of the common starting point. The points in the signal corresponding to the two ends of the window are taken as the end points of the two vectors respectively. The point of (x_1, y_1) is the coordinate of the end point of $\overrightarrow{V}_1$, and (x_2, y_2) is the coordinate of the end point of $\overrightarrow{V}_2$. As such, both vectors of $\overrightarrow{V}_1$ and $\overrightarrow{V}_2$can be obtained by the following equations:

$$\overrightarrow{V}_1 = (x_1 - x_m, y_1 - y_m) \tag{1}$$
$$\overrightarrow{V}_2 = (x_2 - x_m, y_2 - y_m) \tag{2}$$

As the window moved, a sequence of angles between two vectors obtained made up a new array(α_1, α_2 ..... α_n), α can be obtained through the following equation:

$$\alpha = \arccos\left(\frac{\overrightarrow{V}_1 \cdot \overrightarrow{V}_2}{\left|\overrightarrow{V}_1\right|\left|\overrightarrow{V}_2\right|}\right) \tag{3}$$

in which the smallest angel α_T was picked for choosing the window we need. And α_T can be obtained by the following equation:

$$\alpha_T = \min\{\alpha_1, \alpha_2, \alpha_3, \ldots, \alpha_n\} \tag{4}$$

For purpose of increasing the algorithm robustness, the final location of T peak of this current RR interval was chosen as the point which was furthest from the equipotential line of ECG in the window in which the α_T exist.

Meanwhile, T offset detection is also developed on the basis of morphological method which grants the low space complexity and time complexity, as shown in Fig. 3(E). Positive T-wave detection is considered first. And the n-th point on the right of T peak is defined as the temporary T offset. Then, a time range is used to restrict the location of temporary T offset. An area S_n is bounded by two straight line and T-wave of ECG signals. S_n can be calculated by the following equation.

$$S_n = \sum_{i=0}^{n} (Signal(tp_1 + i) - te_2) \tag{5}$$

In which te_2 is the ordinate of the temporary T offset. tp_1 is the abscissa of the T peak. Signal($tp_1 + i$) means the magnitude corresponding to the location of tp_1 + i in ECG signal. As the temporary T offset moved to the direction which time went, the $S_n$was becoming bigger and bigger. Until this area stopped getting bigger or the temporary T offset reach the boundary of time range, it stops moving. S_T is picked out as the maximum value by the following equation.

$$S_T = \max\{S_1, S_2, S_3, \ldots, S_n\} \tag{6}$$

The final location of T offset is the temporary T offset corresponding to S_T. The processing for positive T-wave is similar to negative T-wave except the S_n was upon T-wave.

3) INTERVAL CALCULATION

Based on the QRS-T wave detection, five ECG intervals of one heart beat can be calculated. Examples of ECG processing of one normal segment and one SCD segment is shown in Fig. 4. The original ECG signals and the wave detection results for the SCD group and Normal group are shown in top and bottom of Fig.4 (A) and Fig. 4(B), respectively. Fig. 4 (C) shows five ECG intervals calculated as below.

Following the QRS complex detection and T-wave delineation, It is necessary to remove the abnormal beat

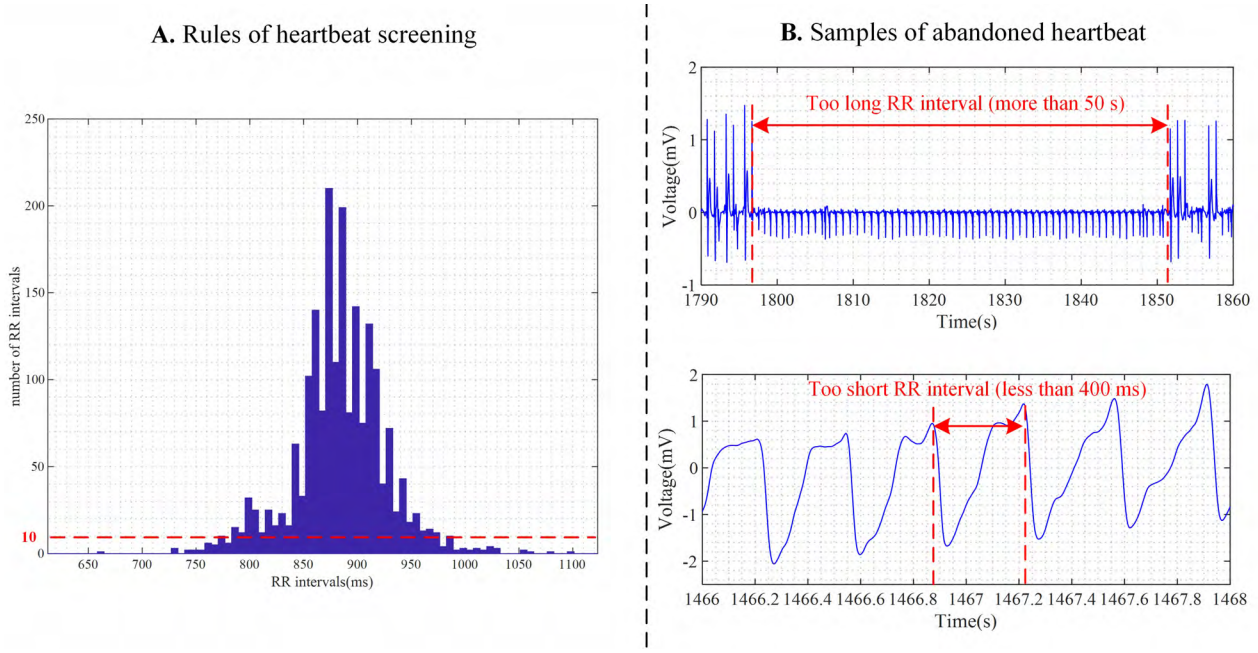


FIGURE 5. Rules of screening of a heartbeat used in this study (A) and two examples (B) a too long length of RR interval (top) and a too short length of RR interval (bottom).

by screening RR intervals so as to effectively increase the precision and efficiency of the prediction algorithm. Based on R peak detection, a frequency histogram of RR interval of each ECG recording could be obtained, as shown in Fig. 5(A). In this work, RR intervals whose number is less than ten were selected as a condition for abandoning improper heartbeat. It means the improper heart beat (with too long or short RR interval) was abandoned. Two examples with abandoned heartbeat are presented in Fig. 5(B) for a too long length of RR interval (top) and a too short length of RR interval (bottom).

After QRS-T wave detection and every heartbeat screening, five ECG intervals calculation can be computed directly by the following equations:

$$\text{QT interval} = \text{T offset - Q onset} \tag{7}$$

$$\text{QRS duration} = \text{S offset - Q onset} \tag{8}$$

$$\text{JT}_\text{p}\text{interval} = \text{T peak - S offset} \tag{9}$$

$$\text{JT}_\text{e}\text{interval} = \text{T offset - S offset} \tag{10}$$

$$\text{T}_\text{p}\text{T}_\text{e}\text{interval} = \text{T offset - T peak} \tag{11}$$

C. ARRHYTHMIC RISK OF ECG SIGNAL

An ECG is one of the best features to describe heart diseases, which carries valuable information about electrophysiological properties of the heart. Mechanisms of VT/VF have been comprehensively studied during the last 20 years, and data suggest that common underlying electrophysiological substrates could be found in all arrhythmic SCD [20], [21]. As described in Introduction of this paper, traditional clinical markers for risk prediction have largely focused on abnormal repolarization, of which QT and corrected QT are archetypal examples. And, the limitations of them in predicting arrhythmogenicity led to the recent development of novel repolarization markers such as T_pT_e, T_pT_e/QT ratio, JT_p/JT_e and T_pT_e/JT_p ratios, and conduction-repolarization markers such as T_pT_e/QRS and $T_pT_e/(QT\times QRS)$, which all are used as our arrhythmic risk markers for early identification of SCD in this study, as summarized in Table 3.

As shown in Table. 3, five risk markers are derived from different ECG signal components, such as QRS complex duration, which is used for deriving conduction-repolarization markers such as T_pT_e/QRS and $T_pT_e/(QT\times QRS)$ ratios. The other repolarization markers such as T_pT_e/QT, JT_p/JT_e and T_pT_e/JT_p ratios are easily calculated on the basis of the accurate delineation of T-wave and relevant interval. As suggestions from clinical findings, arrhythmias risk markers might have different prediction result in patients with different diseases. Therefore, these five arrhythmic risk markers extracted from both electrophysiological conduction and repolarization all are taken into consideration and fused together to expect a good SCD prediction results in this work.

D. MACHINE LEARNING BASED CLASSIFICATION

For the propose of getting a precise result of classifying ECG signals from normal group and SCD group and selecting the best classifier with the highest accuracy. Five currently frequently studied classifiers such as the KNN, the DT, the SVM, the NB and the RF were trained and assessed in this work for automatically distinguishing between the SCD group and the normal group . The general description of the different classifier are given below.

TABLE 3. A summary of arrhythmias risk markers based on measurements of electrocardiographic conduction and repolarization used in this study.

Risk ECG Features	Definition	Derived from	Prediction of SCD for patients with certain diseases	Reference	Classification
T_pT_e/QT	Interval from the peak to the end of the T-wave divided by QT interval	T wave and QT interval	Long QT syndrome, short QT syndrome, myocardial ischemia, heart failure, Brugada syndrome	[28,29]	Repolarization
JT_p/JT_e	Interval from J-point to peak of the T-wave divided by interval from J-Point to end of T-wave	T wave and ST segment	Myocardial ischemia	[39,40]	Repolarization
T_pT_e/JT_p	Interval from the peak to the end of the T-wave divided by Interval from J-point to peak of the T-wave	T wave and ST segment	Diabetes mellitus, myocardial ischemia	[39,40]	Repolarization
T_pT_e/QRS	Interval from the peak to the end of the T-wave divided by QRS duration	T wave and QRS duration	Hypertension	[28,41]	Conduction and Repolarization
$T_pT_e/(QT \times QRS)$	Interval from the peak to the end of the T-wave divided by QT interval multiplied by QRS duration	T wave, QRS duration and QT interval	Long QT syndrome, short QT syndrome, Brugada syndrome,	[28,41]	Conduction and Repolarization

1) K-NEART NEIGHBOR (KNN)

KNN is a simple and effective method for classification.

KNN classifies each unlabeled example according to most marks between k-nearest neighbors in the training set [44]. Therefore, the performance of KNN rules depends critically on the distance metric used to identify the nearest neighbor. In this study, k = 15, 20 and 25 are used.

2) DECISION TREE (DT)

The decision tree classification algorithm is an instance-based induction learning method, which can extract the tree classification model from a given disordered training sample. The decision tree classification algorithm is relatively simple [45].

3) SUPPORT VECTOR MACHINE (SVM)

Support vector machines are trained using the idea of classification intervals. It relies on the preprocessing of data to express the original pattern in a higher dimensional space [46]. In this work, to distinguish ECG between the normal and SCD, SVMs with Radial Basis Function (RBF) kernel function are explored.

4) NAÏVE BAYES (NB)

The naive Bayes classifier is a series of simple probability classifiers based on Bayesian theorem based on strong and simple independence between hypothetical features [47].

5) RANDOM FOREST (RF)

A random forest is a classifier containing multiple decision trees. A training forest is used to generate a random forest composed of multiple classification trees. The classification result of the test data is determined by the score formed by the classification tree voting [48].

E. INTEGRATED SCD INDEX (SCDI) BASED ON ELECTROCARDIOGRAPHIC FEATURE

On the basis of the complex electrophysiological mechanism of ventricular arrhythmias induced SCD, a common limitation of individually using the above mentioned arrhythmic risk markers is that they may not provide a high enough sensitivity, specificity, or either both of them. This concept of integrated index for SCD prediction was conceived and advanced recently by Acharya's Lab [38] and Ghista's Lab [49], while both of them were with only the features extracted from the complex transforms of ECG signal and without any considerations of the electrophysiological characteristics. Thus, our further hypothesis is that, if such markers reflect different underlying physiological phenomena, they might add complementary information to each other and, consequently, a combined and integrated index might improve the capability for risk stratification of patients. Therefore, a novel SCD index integrated from above mentioned arrhythmias risk markers was proposed without any additionally complex signal transforms and machine learning analysis. Although the machine leaning based classifier can perform automatic diagnosis as discussed above, the SCDI could provide clinicians in hospital settings or at home with a convenient, intuitive and simple way to initially determine whether a patient has a risk of sudden cardiac death, where there usually are a low computation resource and a high algorithm robustness requirement. Based on arrhythmic risk markers, the proposed SCDI is formulated in this paper by using the following mathematical methods. Initially, three informative markers (the T_pT_e/QT, the JT_p/JT_e

and the T_pT_e/QRS) are picked for generating SCDI, which are selected from such five risk markers as discussed above with the method of Student's t-test. The p-value of each risk marker is calculated and then employed to identify the significance of the marker. Then, the SCDI is preliminarily formulated by the equation below:

$$\mathrm{SCDI} = \sum_{n=1}^{3} \mathrm{marker_n} \times \mathrm{c_n} \tag{12}$$

Moreover, there selected three markers were normalized individually; and an average difference calculated for each normalized marker is formulated by the equation below:

$$\mathrm{ave\_diff_n} = \mathrm{ave\_nmarker_n} - \mathrm{ave\_dmarker_n} \tag{13}$$

where the $\mathrm{ave\_nmarker_n}$ is the average value of normalized $\mathrm{marker_n}$ in normal group, the $\mathrm{ave\_dmarker_n}$ is the average value of normalized $\mathrm{marker_n}$ in SCD group, the $\mathrm{ave\_diff_n}$ is the average difference between normal groupand SCD group of $\mathrm{marker_n}$. According to the value of $\mathrm{ave\_diff_n}$, three markers were ranked from high to low as the $\mathrm{marker_1}$ of T_pT_e/QRS,, the $\mathrm{marker_2}$ of T_pT_e/QT, and the $\mathrm{marker_3}$ of JT_p/ JT_e. It means T_pT_e/QRS has the most significant difference between the SCD group and the normal group. Moreover, the $\mathrm{ave\_diff_n}$ is also used to calculate the coefficient of $\mathrm{c_n}$ which can be obtained by the following equations:

$$\mathrm{c_n} = \left(\frac{\mathrm{ave\_diff_n}}{\mathrm{ave\_diff_1}}\right) \times \left(\frac{1}{\mathrm{max\_marker_n}}\right) \tag{14}$$

where the $\mathrm{max\_marker_n}$ is the maximum value of $\mathrm{marker_n}$ from these three extracted markers in this work. Finally with our 46 ECG recordings, the obtained mathematical formulation of this integrated SCDI is given by:

$$\begin{aligned}\mathrm{SCDI} = {} & T_pT_e/\mathrm{QRS} \times 0.6669 \\ & + T_pT_e/\mathrm{QT} \times 3.2787 + JT_p/JT_e \times 1.2084\end{aligned} \tag{15}$$

IV. PERFORMANCE AND RESULTS

A. PERFORMANCE OF INDIVIDUAL ECG MARKERS

For both the SCD group and the normal group, the performance of detection of QRS complex and T wave is summarized in Table 4. The parameter of μ means the average value of error, and the σ means the standard deviation of error. The result is obtained by testing the same database (SDDB and NSRDB) in QT database and the golden standard is the annotation of the first expert from the public database. Moreover, Table 5 shows the averaged results of calculation of each arrhythmic risk marker at different time thirty minutes before the SCD occurs. The time lasts from the first minute to the thirtieth minute. For the first five minutes prior to SCD onset, the results is given as averaged values in one minute; and for the subsequent second twenty-five minutes, the results is shown as averaged values in five minutes. Furthermore, Table 6 shows the comparison of averaged results of 5 arrhythmic risk markers between the normal group and thirty minutes before SCD onset of the SCD group, where the p-value indicate that all markers are significant enough for classification.

TABLE 4. Performance evaluation of the QRS complex detection and T wave delineation.

Parameter	*Q onset*	*R peak*	*S offset*	*T peak*	*T offset*
μ (ms)	13.1468	3.3752	14.3053	14.0275	13.6501
σ (ms)	6.7885	2.1199	6.6700	6.6241	8.5385

B. PERFPORMANCE OF THE MACHINE LEANRING BASED CLASSIFICAITON WITH FIVE ECG MARKERS

In this study, five-fold cross validation method is employed to build and evaluate the performance of various classifiers. In this study, a total of 1380 one-minute segments from 46 signals were divided into five parts equally, One part was used to test the classifier and four parts were used to training the classifier. This procedure was repeated five times using a different test set and train set each time The final performance is the average value of every test set result. As a consequence, the ability of the proposed method for prediction of sudden cardiac death was evaluated using sensitivity (Sen), specificity (Spe), accuracy (Acc) which are computed by following equations below:

$$\mathrm{Sensitivity} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}} \tag{16}$$

$$\mathrm{Specificity} = \frac{\mathrm{TN}}{\mathrm{TN} + \mathrm{FP}} \tag{17}$$

$$\mathrm{Accuracy} = \frac{\mathrm{TN+TP}}{\mathrm{TN} + \mathrm{TP} + \mathrm{FP} + \mathrm{FN}} \tag{18}$$

where the TP means the number of SCD that are recognized as SCD; FN means the number of SCD that are recognize as not SCD; TN means the number of not SCD that are recognized as not SCD and FP means the number of not SCD that are recognized as SCD.

All five markers are used as input to KNN, SVM, DT, NB, RF classifiers for classification in this work. The classifier of RF shows the best performance with an accuracy 99.49% in comparison with the other four classifiers, As shown in Table 7 and 8. Table 7 shows the result of five-fold cross validation with the RF. Table 8 summarize performances of various classifiers. Noted that the SVM classifier with kernel of RBF has been tested. Other performance of classifiers are slightly lower than RF. Moreover, the performance of the RF based method for SCD prediction is given in detail at different time before the SCD occurs, as shown in Fig. 6. Similarly, the time lasts from the first minute to the thirtieth minute. The first five minutes shows the results in one minute and the second twenty-five minutes show the results in five minutes.

TABLE 5. The mean±SD of features extracted from 1 Min to 5 Min, in 1-min intervals, and from 1 o 30 Min in 5-mins intervals, before SCD ECG signals data.

Minutes before SCD	T_pT_e/QT		JT_p/JT_e		T_pT_e/JT_p		T_pT_e/QRS		$T_pT_e/(QT\times QRS)$	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
1st	0.2145	0.0368	0.7014	0.0555	0.4688	0.1463	0.9054	0.2796	0.0073	0.0026
2nd	0.2137	0.0373	0.6995	0.0611	0.4721	0.1630	0.8608	0.2448	0.0068	0.0022
3rd	0.2146	0.0387	0.6961	0.0689	0.4858	0.1942	0.8893	0.2488	0.0073	0.0025
4th	0.2098	0.0393	0.7081	0.0591	0.4552	0.1483	0.9448	0.3589	0.0078	0.0037
5th	0.2116	0.0350	0.7047	0.0596	0.4602	0.1643	0.9057	0.2870	0.0072	0.0024
1st-5th	0.2128	0.0369	0.7020	0.0603	0.4684	0.1621	0.9012	0.2839	0.0073	0.0027
6th-10th	0.2105	0.0408	0.7041	0.0651	0.4637	0.1750	0.8664	0.2719	0.0068	0.0024
11th-15th	0.2088	0.0390	0.7069	0.0650	0.4588	0.1781	0.8561	0.2652	0.0068	0.0023
16th-20th	0.2053	0.0388	0.7147	0.0561	0.4370	0.1284	0.8664	0.2892	0.0068	0.0024
21th-25th	0.2097	0.0382	0.7117	0.0538	0.4421	0.1269	0.9554	0.4567	0.0075	0.0035
26th -30th	0.2157	0.0390	0.7018	0.0575	0.4695	0.1449	0.9812	0.5626	0.0078	0.0045

TABLE 6. The mean±SD of features extracted from the normal and thirty minutes before SCD ECG signals data.

Feature	Normal		SCD		p-Value
	Mean	SD	Mean	SD	
T_pT_e/QT	0.1682	0.0246	0.2105	0.0388	7.0373e-12
JT_p/JT_e	0.6573	0.0436	0.7069	0.0598	2.3410e-10
T_pT_e/JT_p	0.5390	0.0999	0.4566	0.1540	7.5806e-04
T_pT_e/QRS	0.3359	0.0624	0.9045	0.3747	4.6497e-15
$T_pT_e/(QT\times QRS)$	0.0051	0.0010	0.0072	0.0031	0.0113

TABLE 7. The statistical measures and the average performance in percentage of the random forest by using the five-fold cross method on the normal and thirty one-min intervals before SCD ECG signals data.

Folds	TP	TN	FP	FN	Sen (%)	Spe (%)	Acc (%)
1	176	98	2	0	100	98	99.28
2	166	110	0	0	100	100	100
3	165	110	0	1	99.40	100	99.64
4	173	101	2	0	100	98.06	99.28
5	158	116	1	1	99.37	99.15	99.28

Note: TP = true positive, TN = true negative, FP = false positive, FN = false negative, Sen = sensitivity, Spe = specificity, Acc = accuracy.

C. VALIDAITON OF OUR INTEGRATED SCDI EFFICACY

On the other hand, the efficacy of the proposed SCDI for SCD prediction is validated with the same datasets as used in the machine learning based method above. The calculated boxplot of the proposed SCDI during the thirty-minute prior to the SCD onset for SCD group and normal group is

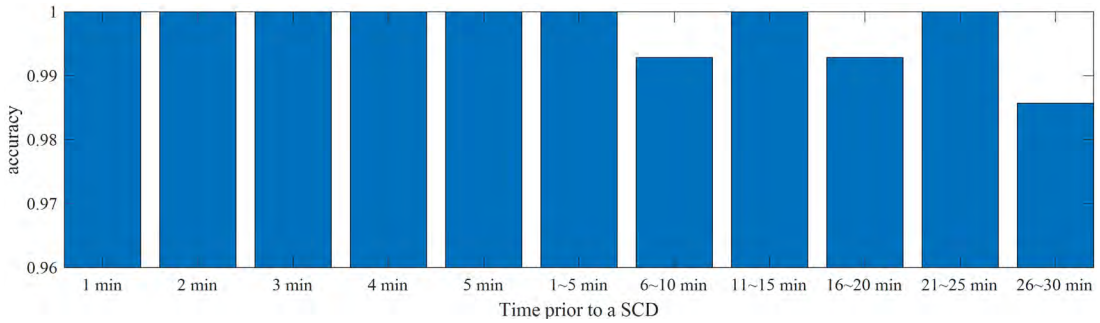


FIGURE 6. Accuracy of the proposed method by using the random forest classifier during the thirty-minute prior to the SCD onset. The first five minutes are the average of the accuracy per minute, and the last twenty-five minutes is the average of the accuracy every five minutes.

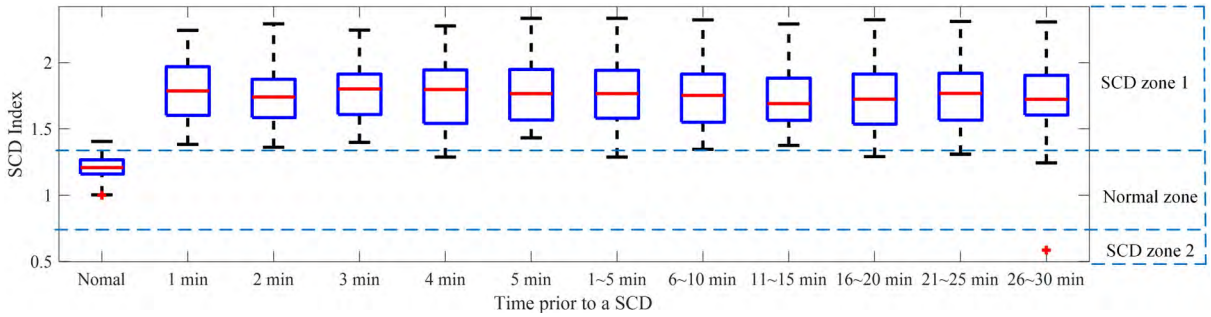


FIGURE 7. Results of the proposed method with the integrated SCDI during the thirty-minute prior to the SCD onset from SCD group and Normal group. SCD zone 1 and SCD zone 2 means the distribution of SCDI in SCD group and normal group means the distribution of SCDI in normal group.

TABLE 8. Average performance of various machine learning classifiers on the normal and thirty one-min intervals before SCD ECG signals data.

Classifiers	Sen (%)	Spe (%)	Acc (%)
KNN_25	98.56	99.46	98.91
KNN_20	98.45	99.45	98.84
KNN_15	98.23	98.96	98.48
SVM_ RBF_0.8	97.87	99.79	98.62
SVM_ RBF_1.2	98.00	99.81	98.70
DT	99.17	98.68	98.99
NB	98.67	95.58	97.46
RF	99.75	99.04	99.49

Note: KNN = k nearest neighbor, DT = decision tree, NB = Naive Bayes, SVM = support vector machine, RF = random forest, Se = sensitivity, Sp = specificity, Acc = accuracy.

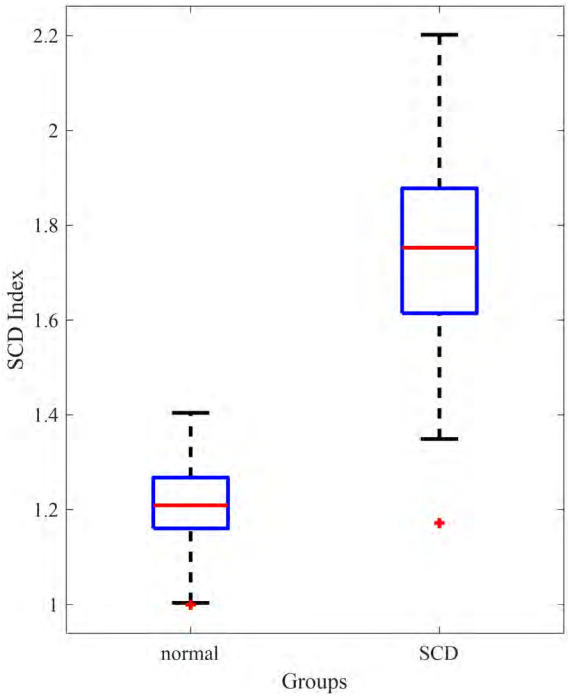


FIGURE 8. Variation of ECG index with the integrated SCDI for normal and SCD classes.

presented in detail at different time before the SCD occurs in Fig. 7. The box chart shows the difference in the range of data distribution between the normal group and the SCD group. Meanwhile, Fig. 8 shows the boxplot of averaged value of the SCDI for the thirty minutes prior to SCD onset in the SCD group and in normal group. According to the global distribution of SCDI from normal and SCD group, the figure was divided into three zones (normal zone, SCD zone 1, SCD zone 2). Table 10 shows the significant difference of

TABLE 9. The comparison of our algorithm and other recent methods for Predicting sudden cardiac death based on ECG parameters and HRV signal.

Author (year)	Material			Methodology		Best performance		
	Data type	Dataset used	Length of signal (min)	Feature extraction (No. of features)	Classifier	Acc (%)	Sen (%)	Spe (%)
Acharya (2015) [38]	ECG	SDDB NSRDB	4 minutes before SCD	Nonlinear features(18) and SCDI	DT, SVM	92.11%	92.5%	91.67%
Fujita (2016) [25]	HRV	SDDB NSRDB	4 minutes before SCD	Nonlinear features (4) Nonlinear heart rate variability analysis	SVM, KNN	94.7%	95.00%	94.40%
Sanchez (2018) [39]	ECG	SDDB NSRDB	20 minutes before SCD	Nonlinear methods HI Wave packet transform	EPNN	95.8%	Unknown	Unknown
Khazaei (2018) [26]	HRV	SDDB NSRDB	6 minutes before SCD	RQA (13) and increment entropy(2 out of 14) Nonlinear method	DT, KNN, SVM NB	95.00%	95.00%	95.00%
Ebrahimzadeh (2018) [23]	HRV	SDDB NSRDB	12 minutes before SCD	HRV Features(23) Time local subset feature selection	MLP	88.29%	Unknown	Unknown
Ebrahimzadeh (2019) [24]	HRV	SDDB NSRDB	13 minutes before SCD	HRV Features(23) Time local subset feature selection	MLP	90.18%	Unknown	Unknown
Present study	ECG	SDDB NSRDB AHADB	30 minutes before SCD	Arrhythmias risk markers (5) and SCDI	DT, KNN, SVM , NB, RF	99.49%	99.75%	99.04%

Note: KNN = k nearest neighbor, DT = decision tree, NB = naive Bayes, SVM = support vector machine, RF = random forest, EPNN = enhanced probabilistic neural network classifier, MLP = multilayer perceptron classifier, HI = homogeneity index, RQA = recurrence quantification analysis, HRV = heart rate variability, Sen = sensitivity, Spe = specificity, Acc = accuracy.

TABLE 10. Range of SCD index for normal and SCD classes.

Normal		SCD		*p*-Value
Mean	SD	Mean	SD	
1.2058	0.0795	1.7619	0.1902	6.5061e-35

SCDI between normal group and SCD group. The p-value indicate that SCDI are significant enough for classification. Compared with previous study [38], our SCDI is directly based on arrhythmic risk markers which are easier for doctors to understand. Meanwhile our SCDI was generated by using mathematical methods which based on the significance difference of 5 arrhythmic risk markers. This SCDI is developed by the database only. It can be further improved by optimizing the c_n with more clinic ECG data and using arrhythmic markers with more clinic significance.

D. COMPARISON TO PREVIOUS WORKS

Table 9 shows that the comparison of our RF based algorithm on measurable arrhythmic markers with other recent methods for predicting SCD on ECG parameters and HRV signal. We have listed six research work from 2015 to 2019, giving detailed information of work from three aspects: data material, methodology and performance. In terms of materials, we separately listed the types of signals, databases and the length of signals used in each study. Concerning about methods, we separately list the methods used by each work, the number of markers, and the classifier. In regard to performance, we give results of sensitivity, specificity, accuracy.

V. DISUSSION AND CONCLUSION

A. MAIN CONTRIBUTION OF THIS STUDY

For more than 20 years, the incidence of SCD in the United States has been estimated to be around 350,000 events each year [50], and the total annual incidence of SCD was estimated to be the range of 544,000 in China. Meanwhile, the population pool is continuing to grow with better therapies for heart disease, the currently fast-paced modern society, and the boom of the aging population (noted that the risk of SCD increases with advancing age, peaking in the 75-84 years age group) [1], [2]. These alarming statistics emphasize the importance of this societal challenge and the need for attempts to find effective techniques for SCD prevention. Clinical arrhythmias risk markers derived from electrocardiography are important for this purpose. In past decades, a number of clinical investigations demonstrated that the ECG derived features have been shown to be useful for prediction of a malignant arrhythmias and subsequent SCD in certain clinical situations [27], [29] and [39], [40]. As a significant challenge to the clinical use of the complex ECG parameters owing to the time-consuming procedure of manually indentifying them from large amounts of

ECG data, two automated predictor of SCD that directly analyzed ECG signal have been proposed recently by Sanchez's group in 2018 [39] and Acharya's group and 2015 [38]. Besides these complex algorithms on the basis of wavelet transform for extracting informative features from ECG, the recent proposed machine learning techniques would be a good opportunity for exploring an automatic SCD Predictor [43], [48], which were used to differentiate between ECG of normal subjects and those of subjects at risk of developing SCD in an end-to-end manner.

In this study, we tried to automatically indentify a developing SCD in patients by using machine learning approaches directly on arrhythmic risk markers, without any complex transforms for extracting features from ECG. These electrocardiographic markers used in this work have been intensively demonstrated in clinical practices as mentions above, including repolarization interval ratios such as T_pT_e/QT, JT_p/JT_e and T_pT_e/JT_e and conduction-repolarization markers such as T_pT_e/QRS and $T_pT_e/(QT\times QRS)$. Moreover, we introduced one more much efficient and practical approach based on simple calculation of SCDI, which is integrated from these informative markers and would simplify the whole prediction strategy. With experiments in the MIT-BIH and AHA databases [42] from 28 patients at risk of SCD and 18 health subject, the performance both of the machine learning based and SCDI based prediction methods presented in this work were evaluated. Compared to those automated prediction methods by using complex signal processing techniques and without any considerations of electrophysiological indicators of SCD, our proposed method by deriving arrhythmias risk markers directly from ECG has a higher accuracy, a relative longer prediction time prior the occurrence of an SCD, and lower computational consuming. Besides that, the proposed practical and straightforward strategies with arrhythmias risk markers offers a considerable benefit when the need to interpret ECG abnormalities at the clinical and electrophysiological level. To our best knowledge, this is the first attempt where measurable arrhythmias risk markers from ECG signal were employed for the automated identification of an developing SCD at an earlier stage by using machine learning approaches and their correspondingly integrated index, thereby it provides practical advantages in analyzing large amounts of ECG data in a short time while assuring a relative high accuracy.

B. PERFORMANCE OF OUR ECG BASED PREDICTOR OF SCD

Compared with some studies using HRV signal [23], [26], we innovatively used clinically markers to analyze ECG signals directly, rather than HRV signals. This is more simple and fast. All work's dataset contain: NSRDB and SDDB which makes the results can compare with each other. Almost all of them using nonlinear methods and time and frequency analysis methods, only the proposed method use the clinic markers which are more simple and reasonable. Furthermore, we tested many mainstream classifiers and adjusted the parameters of the classifiers for the purpose of getting the better classification results. Compared with the performance of previous work on the ECG signals reported recently by other researchers [38], [39], our research has two main advantages. Firstly, our method has the highest sensitivity (99.75%) and specificity (99.04%). Secondly, our method predicts the risk of SCD up to 30 minutes before its onset using ECG signals. Meanwhile, based on the clinic markers, we proposed a simple and efficient formulation of SCDI using mathematical method. SCDI can differentiate the SCD and normal group significantly. Additionally, the algorithm shows good real-time performance, For 30 minutes, the marker extraction and subsequent classification only costs 0.8721 second. And the entire process including extraction and classification costs less than one second. The proposed method runs on matlab2018a which deployed on a computer with Inter(R) Core(TM) i5-3470 CPU with 8G memory.

C. CHALLENGES OF EARLY SCD RISK PREDICTION

Actually, the clinical usefulness and impact of SCD risk predictors are complex and challenging. Although there is usually an underlying substrate that puts an individual at risk, this factor may not be discovered in advance, especially in general population [4], [7], [13]. In fact, 40-50% of all SCDs cases have a cardiac arrest with no prior symptoms or warnings [13], [51], [52]. In other words, the ability to effectively predict cardiac arrest among a high percentage of potential victims is difficult if one predictor is derived only from risk markers both in an invasive or a non-invasive modality, and without any considerations of underlying disease behind these markers [53]. As reported in epidemiologic studies, nearly two thirds of cardiac arrests occur as a first clinically manifested event or in the clinical setting of known disease in the absence of strong risk predictors [1]. Among them, coronary heart disease is the most common etiologic basis for SCD, approximately 75% - 80% of which are due to this one underlying etiology [7], [49]. The evolution and expression of SCD due to coronary heart disease involves a multitiered cascade of pathophysiology, operating in different time domains and generating various clinical symptoms. Less than 25% of the victims of SCD have high-risk markers based on only arrhythmic (5%-10%) or hemodynamic parameters (7%-15%) [2]. Given these considerations, the current concept of sudden cardiac death should embrace electrical, ischemic, mechanical (pump failure), and more recently, myocardial intracellular genetic mutation mechanisms. From the perspective of individual risk, there is a broad range of predictive power, depending upon the specific measures of risk used. Much of the progress that has been made to date in profiling risk of SCD has been based on clinical markers, which primarily indentify the extent of disease, either at a myocardial level or at a vascular level [4]. One of approach appear promising for the future is prediction of arrhythmias risk, which is intended to identify those individuals who are at risk for the events that rigger fatal arrhythmias in a shorter time frame [3]–[5]. There include

measures of cardiac repolarization used in this study, pathophysiological control mechanisms liking heart rate variability studied intensively before [23], [25], [35], [36], [54]–[56], or inflammatory markers such as familial or genetic profiling which may provide higher resolution of SCD risk in specific individual [50], [57], [58]. Nevertheless, much deeper discussions on the causes and clinical expression, pathological and physiological mechanism, and even relevant epidemiology of SCD are beyond the scope of this paper.

D. LIMITATIONS OF THIS STUDY AND FUTURE WORKS

Besides general challenges of risk prediction of SCD, we are still working on few limitations of this study in our future study. First, the relatively small patient data from the public ECG database was collected in the present work. Further work would be interested to collect a great number of clinical data to train the proposed classifier and SCDI for a higher accuracy. Meanwhile, it should be noted that the problem of overfitting may appear during the training process that may reduce the accuracy of the recognition of arrhythmias risk marker of SCD. Which means the production is too close to the training set and it may therefore fail to fit additional data or predict future observations reliably. In this study, an attempt was tried to deal with this problem as the training set is increased and the complexity of the training model is reduced. Additionally, it should be cautious that these arrhythmias risk markers used in this study are currently in epidemiological studies and not routinely. Eventually, once there have proved their clinical utility in a large population, we expect our proposed methods to be used widely in clinical practice. These predictions could be more practical and efficient if applied in portable smart devices with real-time requirements in hospital settings or at home.

ACKNOWLEDGMENT

The authors would like to acknowledge Biomedical Imaging and Electrophysiology Laboratory (BMI-EP), University of Electronic Science and Technology of China for providing computational and biomedical equipments.

REFERENCES

[1] D. P. Zipes and H. J. J. Wellens, "Sudden cardiac death," *Circulation*, vol. 98, no. 21, pp. 2334–2351, 1998.

[2] R. J. Myerburg and A. Castellanos, "Cardiac arrest and sudden cadiac death," in *Braunwald's Heart Disease: A Textbook of Cardiovascular Medicine*, 6th ed. Philadelphia, PA, USA: WB Saunders, 2001, pp. 890–931.

[3] R. M. John, U. B. Tedrow, B. A. Koplan, C. M. Albert, and M. O. Sweeney, "Ventricular arrhythmias and sudden cardiac death," *Lancet*, vol. 380, no. 9852, pp. 1520–1529, Oct. 2012.

[4] A. B. de Luna, P. Coumel, and J. F. Leclercq, "Ambulatory sudden cardiac death: Mechanisms of production of fatal arrhythmia on the basis of data from 157 cases," *Amer. Heart J.*, vol. 117, no. 1, pp. 151–159, 1989.

[5] H. V. Huikuri, A. Castellanos, and R. J. Myerburg, "Sudden death due to cardiac arrhythmias," *New England J. Med.*, vol. 345, no. 20, pp. 1473–1482, 2001.

[6] T. W. Shen, H. Shen, C. H. Lin, and Y. L. Ou, "Detection and prediction of sudden cardiac death (SCD) for personal healthcare," in *Proc. IEEE 29th Annu. Int. Conf.*, Buenos Aires, Argentina, Aug. 2007, pp. 2575–2578.

[7] N. J. Pagidipati and T. A. Gaziano, "Estimating deaths from cardiovascular disease: A review of global methodologies of mortality measurement," *Circulation*, vol. 127, no. 6, pp. 749–756, 2013.

[8] G. Finocchiaro, M. Papadakis, S. Sharma, and M. Sheppard, "Sudden cardiac death," *Eur. Heart J.*, vol. 38, no. 17, pp. 1280–1282, 2017.

[9] J. C. Lopshire and D. P. Zipes, "Sudden cardiac death: Better understanding of risks, mechanisms, and treatment," *Circulation*, vol. 114, no. 11, pp. 1134–1136, 2006.

[10] R. Passman, "Prevention of sudden cardiac death in dialysis patients: Drugs, defibrillators or what else," *Blood Purification*, vol. 35, nos. 1–3, pp. 49–54, 2013.

[11] F. Braunschweig, G. Boriani, A. Bauer, R. Hatala, C. Herrmann-Lingen, S. S. Pedersen, S. Pehrson, R. Ricci, M. J. Schalij, and J. Kautzner, "Management of patients receiving implantable cardiac defibrillator shocks: Recommendations for acute and long-term patient management," *Europace*, vol. 12, no. 12, pp. 1673–1690, 2010.

[12] A. Cheng, D. Dalal, B. Butcher, S. Norgard, Y. Zhang, Z. A. Eldadah, K. A. Ellenbogen, E. Guallar, G. F. Tomaselli, and T. Dickfeld, "Prospective observational study of implantable cardioverter-defibrillators in primary prevention of sudden cardiac death: Study design and cohort description," *J. Amer. Heart Assoc.*, vol. 2, no. 1, 2013, Art. no. e000083.

[13] S. S. Chugh, "Sudden Cardiac Death with apparently normal heart: Clinical implications of progress in pathophysiology," *Cardiac Electrophysiol. Rev.*, vol. 5, no. 4, pp. 394–402, 2001.

[14] G. I. Fishman, S. S. Chugh, J. P. DiMarco, C. M. Albert, M. E. Anderson, R. O. Bonow, A. E. Buxton, P. S. Chen, M. Estes, X. Jouven, and R. Kwong, "Sudden cardiac death prediction and prevention: Report from a national heart, lung, and blood institute and heart rhythm society workshop," *Circulation*, vol. 122, no. 22, pp. 48–2335, 2010.

[15] F. Zhang, P. Li, F. Jiang, and D. Lai, "A shockable rhythm detection algorithm for automatic external defibrillators by combining a slope variability analyzer with a band-pass digital filter," in *Proc. IEEE Workshop Electron., Comput. Appl.*, Ottawa, ON, Canada, May 2014, pp. 828–831.

[16] R. Passman and J. J. Goldberger, "Predicting the future: Risk stratification for sudden cardiac death in patients with left ventricular dysfunction," *Circulation*, vol. 125, no. 24, pp. 3031–3037, 2012.

[17] E. Herzog, F. Javed, B. Pratap, E. F. Aziz, "Strategies for the prevention and treatment of sudden cardiac death," *Open Access Emerg., Med.*, vol. 2010, no. 2, pp. 99–114, 2010.

[18] R. Pacher, B. Stanek, M. Hülsmann, J. Koller-Strametz, R. Berger, E. Hartter, E. Ogris, B. Frey, G. Heinz, G. Mauerer, and M. Schuller, "Prognostic impact of big endothelin-1 plasma concentrations compared with invasive hemodynamic evaluation in severe heart failure," *J. Amer. Coll. Cardiol.*, vol. 27, no. 3, pp. 633–641, 1996.

[19] S. D. Solomon, S. Zelenkofske, J. J. McMurray, A. Harsanyi, J. L. Rouleau, A. Maggioni, L. Kober, H. White, and G. Ertl, "Sudden death in patients with myocardial infarction and left ventricular dysfunction, heart failure, or both," *New England J. Med.*, vol. 352, no. 25, pp. 2581–2588, 2005.

[20] S. Mandala and T. C. Di, "ECG parameters for malignant ventricular arrhythmias: A comprehensive review," *J. Med. Biol. Eng.*, vol. 37, no. 4, pp. 441–453, 2017.

[21] Y. Castro-Torres, R. Carmona-Puerta, and R. E. Katholi, "Ventricular repolarization markers for predicting malignant arrhythmias in clinical practice," *World J. Clin. Cases*, vol. 3, no. 8, pp. 707–720, 2015.

[22] I. Cygankiewicz, W. Zareba, R. Vazquez, J. R. Gonzalez-Juanatey, M. Valdes, J. Almendral, J. Cinca, P. Caminal, A. B. de Luna, and M. Vallverdu, "Heart rate turbulence predicts all-cause mortality and sudden death in congestive heart failure patients," *Heart Rhythm*, vol. 5, no. 8, pp. 1095–1102, 2008.

[23] E. Ebrahimzadeh, M. S. Manuchehri, S. Amoozegar, B. N. Araabi, and H. Soltanian-Zadeh, "A time local subset feature selection for prediction of sudden cardiac death from ECG signal," *Med. Biol. Eng. Comput.*, vol. 56, no. 7, pp. 1253–1270, 2018.

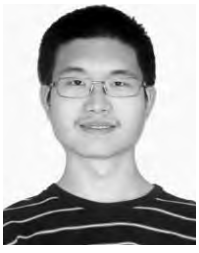
[24] E. Ebrahimzadeh, A. Foroutan, M. Shams, R. Baradaran, L. Rajabion, M. Joulani, and F. Fayaz, "An optimal strategy for prediction of sudden cardiac death through a pioneering feature-selection approach from HRV signal," *Comput. Methods Programs Biomed.*, vol. 169, pp. 19–36, Feb. 2019.

[25] H. Fujita, U. R. Acharya, V. K. Sudarshan, D. N. Ghista, S. V. Sree, L. W. J. Eugene, and J. E. W. Koh, "Sudden cardiac death (SCD) prediction based on nonlinear heart rate variability features and SCD index," *Appl. Soft Comput.*, vol. 43, pp. 510–519, Jun. 2016.

[26] M. Khazaei, K. Raeisi, A. Goshvarpour, and M. Ahmadzadeh, "Early detection of sudden cardiac death using nonlinear analysis of heart rate variability," *Biocybernetics Biomed. Eng.*, vol. 38, no. 4, pp. 931–940, 2018.
[27] J. Ramírez, A. Mincholé, J. Bolea, P. Laguna, and E. Pueyo, "Prediction of sudden cardiac death in chronic heart failure patients by analysis of restitution dispersion," in *Proc. Comput. Cardiol.*, 2013, pp. 1–4.
[28] G. Tse and B. P. Yan, "Traditional and novel electrocardiographic conduction and repolarization markers of sudden cardiac death," *Europace*, vol. 19, pp. 712–721, May 2017.
[29] J. C. Hevia, C. Antzelevitch, F. T. Bárzaga, F. D. Balea, M. A. Q. Pérez, Y. F. Rodríguez, and R. Z. Molina, "Tpeak-Tend and Tpeak-Tend dispersion as risk factors for ventricular tachycardia/ventricular fibrillation in patients with the Brugada syndrome," *J. Amer. Coll. Cardiol.*, vol. 47, no. 9, pp. 1828–1834, 2006.
[30] D. P. Morin, L. Oikarinen, L. Toivonen, M. S. Nieminen, S. E. Kjeldsen, B. Dahlöf, M. John, R. B. Devereux, P. M. Okin, and M. Viitasalo, "QRS duration predicts sudden cardiac death in hypertensive patients undergoing intensive medical therapy: The LIFE study," *Eur Heart J.*, vol. 30, no. 23, pp. 2908–2914, 2009.
[31] E. Locati and P. J. Schwartz, "Prognostic value of QT interval prolongation in post myocardial infarction patients," *Eur. Heart J.*, vol. 8, pp. 121–126, Mar. 1987.
[32] K. S. Spargias, S. J. Lindsay, D. C. Greenwood, J. C. Cowan, S. G. Ball, A. S. Hall, and G. I. Kawar, "QT dispersion as a predictor of long-term mortality in patients with acute myocardial infarction and clinical evidence of heart failure," *Eur. Heart J.*, vol. 20, no. 16, pp. 1158–1165, 1999.
[33] V. Monasterio, P. Laguna, I. Cygankiewicz, R. Vázquez, A. Bayés-Genís, J. P. Martínez, and A. B. de Luna, "Average T-wave alternans activity in ambulatory ECG records predicts sudden cardiac death in patients with chronic heart failure," *Heart Rhythm*, vol. 9, no. 3, pp. 383–389, 2012.
[34] R. L. Verrier and T. Ikeda, "Ambulatory ECG-based T-wave alternans monitoring for risk assessment and guiding medical therapy: Mechanisms and clinical applications," *Prog. Cardiovascular Diseases*, vol. 56, no. 2, pp. 85–172, 2013.
[35] M. T. La Rovere, J. T. Bigger, Jr., A. Mortara, P. J. Schwartz, and F. I. Marcus, "Baroreflex sensitivity and heart-rate variability in prediction of total cardiac mortality after myocardial infarction. ATRAMI (Autonomic Tone and Reflexes After Myocardial Infarction) Investigators," *Lancet*, vol. 351, no. 9101, pp. 478–484, 1998.
[36] H. Evrengul, H. Tanriverdi, B. Amasyali, A. Kilic, T. Celik, H. Turhan, and S. Kose, "The relationship between heart rate recovery and heart rate variability in coronary artery disease," *Ann. Noninvasive Electrocardiol.*, vol. 11, pp. 154–162, Apr. 2006.
[37] R. Liew, "Electrocardiogram-based predictors of sudden cardiac death in patients with coronary artery disease," *Clin. Cardiol.*, vol. 34, no. 8, pp. 466–473, 2011.
[38] U. R. Acharya, H. Fujita, V. K. Sudarshan, V. S. Sree, L. W. J. Eugene, D. N. Ghista, and R. S. Tan, "An integrated index for detection of sudden cardiac death using discrete wavelet transform and nonlinear features," *Knowl.-Based Syst.*, vol. 83, pp. 149–158, Jul. 2015.
[39] J. Amezquita-Sanchez, M. Valtierra-Rodriguez, H. Adeli, and C. A. Perez-Ramirez, "A novel wavelet transform-homogeneity model for sudden cardiac death prediction using ECG signals," *J. Med. Syst.*, vol. 42, no. 10, p. 176, 2018.
[40] C. Alvarado-Serrano, J. Ramos-Castro, and R. Pallàs-Areny, "Novel indices of ventricular repolarization to screen post myocardial infarction patients," *Comput. Biol. Med.*, vol. 36, pp. 507–515, May 2006.
[41] G. Tse and B. P. Yan, "Novel arrhythmic risk markers incorporating QRS dispersion: $QRS_d \times (T_{peak} - T_{end})$/QRS and $QRS_d \times (T_{peak} - T_{end})$/(QT × QRS)," *Ann. Noninvasive Electrocardiol.*, vol. 22, no. 6, 2016, Art. no. e12397.
[42] A. L. Goldberger, L. A. N. 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, "PhysioBank, PhysioToolkit, and PhysioNet: Components of a new research resource for complex physiologic signals," *Circulation*, vol. 101, no. 23, pp. e215–e220, 2000.
[43] J. Pan and W. J. Tompkins, "A real-time QRS detection algorithm," *IEEE Trans. Biomed. Eng.*, vol. BME-32, no. 3, pp. 230–236, Mar. 1985.
[44] T. Denoeux, "A k-nearest neighbor classification rule based on Dempster–Shafer theory," *IEEE Trans. Syst., Man., Cybern.*, vol. 25, no. 5, pp. 804–813, May 1995.
[45] J. R. Quinlan, "Simplifying decision trees," *Int. J. Man-Mach. Stud.*, vol. 27, no. 3, pp. 221–234, Sep. 1987.
[46] J. A. K. Suykens and J. Vandewalle, "Least squares support vector machine classifiers," *Neural Process. Lett.*, vol. 9, no. 3, pp. 293–300, Jun. 1999.
[47] S.-B. Kim, K.-S. Han, H.-C. Rim, and S. H. Myaeng, "Some effective techniques for Naive Bayes text classification," *IEEE Trans. Knowl. Data Eng.*, vol. 18, no. 11, pp. 1457–1466, Nov. 2006.
[48] A. Liaw and M. Wiener, "Classification and regression by random forest," *R Newslett.*, vol. 2, no. 3, pp. 18–22, 2002.
[49] D. N. Ghista, "Physiological systems' numbers in medical diagnosis and hospital cost-effective operation," *J. Mech. Med. Biol.*, vol. 4, no. 4, pp. 401–418, 2004.
[50] R. Deo and C. M. Albert, "Epidemiology and genetics of sudden cardiac death," *Circulation*, vol. 25, no. 4, pp. 620–637, 2012.
[51] M. Borjesson and A. Pelliccia, "Incidence and aetiology of sudden cardiac death in young athletes: An international perspective," *Brit. J. Sports Med.*, vol. 43, no. 9, pp. 644–648, 2009.
[52] D. Corrado, C. Basso, G. Rizzoli, M. Schiavon, and G. Thiene, "Does sports activity enhance the risk of sudden death in adolescents and young adults?" *J. Amer. Coll. Cardiol.*, vol. 42, no. 11, pp. 1959–1963, 2003.
[53] S. S. Chugh, "Early identification of risk factors for sudden cardiac death," *Nature Rev. Cardiol.*, vol. 7, no. 6, pp. 318–326, 2010.
[54] D. Lai, J. Sun, Y. Li, and B. He, "Usefulness of ventricular endocardial electric reconstruction from body surface potential maps to noninvasively localize ventricular ectopic activity in patients," *Phys. Med. Biol.*, vol. 58, no. 11, pp. 3897–3909, 2013.
[55] D. Lai, C. Liu, M. D. Eggen, P. A. Iaizzo, and B. He, "Equivalent moving dipole localization of cardiac ectopic activity in a swine model during pacing," *IEEE Trans. Inf. Technol. Biomed.*, vol. 14, no. 6, pp. 1318–1326, Nov. 2010.
[56] D. Lai, C. Liu, M. D. Eggen, P. A. Iaizzo, and B. He, "Localization of endocardial ectopic activity by means of noninvasive endocardial surface current density reconstruction," *Phys. Med. Biol.*, vol. 56, no. 13, pp. 4161–4176, 2011.
[57] Y. Friedlander, D. S. Siscovick, and S. Weinmann, "Family history as a risk factor for primary cardiac arrest," *Circularion*, vol. 97, no. 2, pp. 155–160, 1998.
[58] S. G. Priori, J. Barhanin, R. N. Hauer, W. Haverkamp, H. J. Jongsma, A. G. Kleber, W. J. McKenna, D. M. Roden, Y. Rudy, K. Schwartz, and P. J. Schwartz, "Genetic and molecular basis of cardiac arrhythmias: Impact on clinical management parts I and II," *Circulation*, vol. 99, no. 4, pp. 518–528, 1999.

DAKUN LAI (M'11) received the Ph.D. degree in biomedical engineering from Fudan University, Shanghai, China, in 2008, and the Postdoctoral Fellowship in biomedical engineering from the University of Minnesota, Minneapolis, MN, USA.

Since 2012, he has been on the faculty of at the University of Electronic Science and Technology of China (UESTC), China, where he was appointed as an Associate Professor of electrical engineering and biomedical engineering. He is also an editor and reviewer of several international journals. He is currently the Director of the Biomedical Imaging and Electrophysiology Laboratory, UESTC. He is currently involved in the field of heart disease, neurological disorder, and sleep disorders. His major research interests include bioelectromagnetism, neuroengineering, and cardiac electrophysiology. He has pioneered the development of noninvasive cardiac electric source imaging, and made significant contributions to deep learning-based bioelectrical signal analysis, and detection and prediction of severe cardiac arrhythmias and neuro disorder, and numerical application of bio electromagnetism.



YIFEI ZHANG received the bachelor's degree in electronic information science and engineering from the University of Electronic Science and Technology of China (UESTC), Chengdu, China, in 2017, where he is currently pursuing the master's degree in the electronic science and engineering. His current research interest includes ventricular fibrillation and other malignant ventricular arrhythmia using machine learning.

XINSHU ZHANG received the bachelor's degree in electronic engineering from the University of Electronic Science and Technology of China (UESTC), in 2016, where he is currently pursuing the master's degree with the School of Electronic Science and Engineering. His current research interests include biomedical signal processing, deep learning technology, and atrial fibrillation detection.

YE SU received the Ph.D. degree in biomedical engineering from the University of Electronic Science and Technology. She is currently an Attending Physician of Cardiovascular Ultrasound and Cardiology, Sichuan Academy of Medical Sciences & Sichuan Provincial People's Hospital, where she is engaged in the analysis and diagnosis of stress electrocardiogram, electrocardiogram, dynamic electrocardiogram, and dynamic blood pressure. In recent years, she mainly participated in scientific research with the Science and Technology Department of Sichuan Province and the Health Department.

MD BELAL BIN HEYAT received the B.Tech. degree in electronic and instrumentation and the M.Tech. degree in electronics circuit and system from Integral University, Lucknow, UP, India, in 2014 and 2016, respectively. He is currently pursuing the Ph.D. degree with the University of Electronic Science and Technology of China (UESTC). He is an Editor and a reviewer of several international journals. His current research interest includes developing diagnosis system of the neurological disorder and sleep disorders.