International Journal of Electrical and Computer Engineering (IJECE)
Vol. 12, No. 5, October 2022, pp. 4960~4969
ISSN: 2088-8708, DOI: 10.11591/ijece.v12i5.pp4960-4969

Automated prediction of sudden cardiac death using statistically extracted features from electrocardiogram signals

Karna Viswavardhan Reddy[1], Navin Kumar[2]
[1]Department of Electronics Engineering, Jain University, Bengaluru, India
[2]Department of Electronics and Communication Engineering, Amrita School of Engineering, Amrita Vishwa Vidyapeetham, Bengaluru, India

Article Info

Article history:

Received Jul 8, 2021
Revised May 26, 2022
Accepted Jun 14, 2022

Keywords:

Electrocardiogram
Machine learning
Sudden cardiac arrest
Sudden cardiac death
Ventricular fibrillation

ABSTRACT

Sudden cardiac death (SCD) is becoming a severe problem despite significant advancements in the usage of the information and communication technology (ICT) in the health industry. Predicting an unexpected SCD of a person is of high importance. It might increase the survival rate. In this work, we have developed an automated method for predicting SCD utilizing statistical measures. We extracted the intrinsic attributes of the electrocardiogram (ECG) signals using Hilbert-Huang and wavelet transforms. Then utilizing machine learning (ML) classifier, we are using these traits to automatically classify regular and SCD existing risks. Support vector machine (SVM), decision tree (DT), naive Bayes (NB), discriminate k-nearest neighbors (KNN), analysis (Disc.), as well as an ensemble of classifiers also utilized (Ens.). The efficiency and practicality of the proposed methods are evaluated using a standard database and measured ECG data obtained from 18 ECG records of SCD cases and 18 ECG records of normal cases. For the automated scheme, the set of features can predict SCD very fast that is, half an hour before the occurrence of SCD with an average accuracy of 100.0% (KNN), 99.9% (SVM), 98.5% (NB), 99.4% (DT), 99.5% (Disc.), and 100.0% (Ens.)

Corresponding Author:

Karna Viswavardhan Reddy
Department of Electronics Engineering, Jain University
#44/4, District Fund Road, Jayanagar 9th Block, Bengaluru, Karnataka 560069, India
Email: viswavardhanreddy@gmail.com

1. INTRODUCTION

Sudden cardiac death (SCD) arises when a patient's heart starts to pump in an unstable or abnormal rhythm (arrhythmia) and afterwards stops beating altogether. If the person survives, the condition is also known as sudden cardiac arrest (SCA) [1]. SCD is induced by cardiovascular disease patients who have or have not previously had a cardiac problem. The time and manner of death in such circumstances are unanticipated [2], [3]. From the beginning of an unanticipated variations in health condition and unconsciousness, up to a minute is typically assumed to be controlled [3]. SCD is generally the result of a deadly cardiac activity disorder such ventricular fibrillation (VF) or ventricular tachycardia (VT) [4], or a severe bradyarrhythmia [5]. Such arrhythmias typically result in SCA that decreases cardiac function and found it challenging for the heart to effectively push blood out [6]. Whenever SCA issue is left neglected for an extended period of time, it sends a message to SCD. Cardiomyopathy, coronary heart disease, valve illnesses, and hereditary abnormalities are by far the most common causes of malignant ventricular arrhythmias. Instantaneous death may occur if not detected accurately and treated quickly [2], [7]. According to research findings, the rise in actual targeted treatment involvements such as, implantable cardioverter

defibrillators (ICD) lowers SCD death [8], [9]. However, they are still expensive and only a small percent of participants continues receiving appropriate ICD shock. The majority of unintentional deaths occur who have not had qualified candidates [10]. As a result, the public viewing method via an automatic electrical defibrillator (AED) has recently attracted a lot of attention as a technique for preserving patients without an ICD against mortality after cardiogenic shock [11]. However, even now in countries when public AEDs are commonly accessible and has strengthened recovery methods as well as when the first reaction controls are in place, the median rate of survival with SCA continues to decrease [12]. Basic electrophysiological substrate, showing dispersion of refractoriness, and sympathetic stimulation activity in the chambers of the heart was all demonstrated to be harmful for cardiac arrhythmia SCD [13], [14]. As a result, early diagnosis of an unexpected SCD in a patient having aggressive ventricular arrhythmias is critical to enhancing overall survival.

Lately, research has focused on building effective models for calculating the risk of SCD utilizing invasive and non-invasive methodologies including such electrophysiological scanning [11], left ventricular ejection fraction (LVEF) [15], invasive hemodynamic assessment [16], and non-invasive electrocardiography (ECG) [17], [18], among many others. These are still not cost-effective. In comparing to the intrusive or imaging techniques mentioned above, ECG is much less expensive and much more generally sold. The electrical action of the heartbeat generates an ECG, that is, an electronic signal. An organized meta-analysis recently confirmed that certain ECG signal parameters could provide important information on the underlying cardiac substrate abnormality that can contribute to ventricular arrhythmias including SCD. Few of these metrics are pathophysiological control systems controlled by cardiac autonomic processes. Heart rate variability (HRV) or heart rate turbulence (HRT) [19]–[21], echocardiography transfer procedures, and the repolarization delay [22] are also significant. Other metrics are QRS (Q, R, and S waves in an electrophysiological) time [23], QT (Q and T waves) gap and dispersion [24], and T-wave alternative [25]. HRV and HRT is a derivation of an ECG signal which is described as the measurement of the R-R interval's beat-to-beat variation. It has been seriously evaluated for SCD prediction and diagnosis. To reveal the intrinsic features of an HRV signals for monitoring and detection of SCD, frequency domain [20], temporal domain [26], and nonlinear techniques [27] have been proposed. This is achieved mostly through the use of classifications to organize topics and the number of features in various processing domains. HRV or HRT first provided impressive outcomes, but instead were eventually discovered to be unpredictive of arrhythmic mortality [28], [29]. HRV measurement's output in the first several days after a myocardial disease has already been challenged and, its prognostic value has also been shown to really be low [28]. The efficacy of HRV-based SCD risk stratification in cardiovascular events is unknown [30]. It cannot be evaluated in people who have atrial fibrillation or to have a lot of arrhythmias [29]. As a consequence, these Electrocardiography markers are convincing in identification of patients at higher risk of having malignancy ventricular arrhythmia. The majority of the effort has really been centered on clinical studies. Using complex wavelet transforms, statistical calculations, and/or electrophysiological indicators, a few research have achieved automatic extraction of ECG characteristics immediately from ECG data [31], [32]. While, in some works, an SCD index (SCDI) introduced technique of combining some of the features in such a way that they could predict the SCD [33]–[35].

In this paper, our focus is to predict SCD automatically. Towards this, the contribution in this work is at multiple levels: i) an algorithm is developed for constructing a labelled database by segmenting the datasets; ii) data cleaning algorithm is developed to deal with missing values and removal of noise [36]; iii) feature extraction is incorporated using Hilbert-Huang transforms, or empirical mode decomposition (EMD)-intrinsic mode functions (IMF1, IMF2, and IMF3) and wavelet transform, or multilevel 1-D wavelet decomposition (DWT)-approximation coefficients vectors (cAs) and detail coefficients vector (cDs) [37], [38]; iv) method developed for ranking of extracted features and selection using various statistical methods such as analysis of variance (ANOVA), correlation analysis (dCor), and ReliefF to find the features with the most deviations; and v) finally, we utilized artificial classifier such as k-nearest neighbor (KNN), discriminant analysis (Disc), naive bayes (NB), support vector machine (SVM), decision tree (DT), and ensembles of classifiers to identify regular and SCD risk areas and use these characteristics (Ens).

Contents of rest of the paper are: section 2 discusses about the datasets used in the work obtained from various databases of both normal and SCD patients. Section 3 describes the schematic diagram of the proposed methodology for SCD prediction along with algorithms developed. While section 4 discusses the important performance measures, results, and comparison with state of the art. Finally, the conclusion and future works are mentioned in section 5.

2. INPUT DATASET

For the SCD prediction process, at first, the ECG data are collected and pre-processed. The Massachusetts Institute of Technology-Beth Israel Hospital (MIT-BIH) is the largest publicly available

database that provides the ECG signals for this work. Here, the two databases are considered from the MIT-BIH namely, sudden cardiac death Holter (SCDH) and normal sinus rhythm (NSR) [39]. For the people who are at the SCD risk stage, the SCDH database is examined and for the normal people, the NSR database is examined. 18 recordings are involved in the SCD groups which are collected from the SCDH database. The collection of onset ventricular fibrillation (VF) is done before 30 mins of the first lead signal of ECG. This collection is taken at every recording of ECG. In this work, the VF is absent so that the SCDH having the No. 40, 42, and 49 recordings were not used. Because the amplitude R wave is low, the No. 38 and 41 recordings were not incorporated due to unknown lead. Normal sinus rhythm database (NSRDB) database built the recordings with the normal group of 18 half-hour. From Table 1, it is observed that the age of 61.1±18.7 years belongs to the SCD group as they range from 30-89 years and the age of 34.3±8.4 years belongs to the normal group and this group ranges from 20-50 years. There are 9 males and 8 females from the SCD group and 5 males and 13 females from the normal group but 1 SCD patient is omitted who is considered as an unknown gender. In Table 2, the descriptions are elaborately provided with record number, name, arrhythmia, and heart diseases group, and finally before the length of the VF. From the SCD group, it is observed that the majority of the patients are affected by malignant ventricular arrhythmias those who are undergoing heart disease and the abnormality of the cardiac substrate. The ECG signals with two examples are shown in Figure 1. Figure 1(a) belongs to the SCD group that shows around the onset of VF and the Figure 1(b) belongs to the normal group and this show around the signal's middle part.

Table 1. The SCD and regular groups' genders and ages

Group	Total	Gender			Age	
		Male	Female	Unknown	Range	Mean±SD
SCD	18	9	8	1	30-89	61.1±18.7
Normal	18	5	13	0	20-50	34.3±8.4

Table 2. ECG data of SCD patients out from MIT-BIH dataset

Database	Heart diseases	No of records	Record name	Length before VF onset	Arrhythmia categories
NSRDB	Not available (NA)	18	Whole database	NA	unknown
Sudden	Cardiac surgery	4	32, 35, 36, 50	30 min	Ventricular
Cardiac	Coronary artery	1	43		tachycardia,
Death	Unknown	9	30, 33, 34, 37, 44, 46, 47, 48, 51		Ventricular
Holter	Heart failure	2	31, 52		fibrillation,
Database	Ventricular ectopy	1	45		Ventricular flutter
(SDDB)	Acute myelogenous leukemia	1	39		

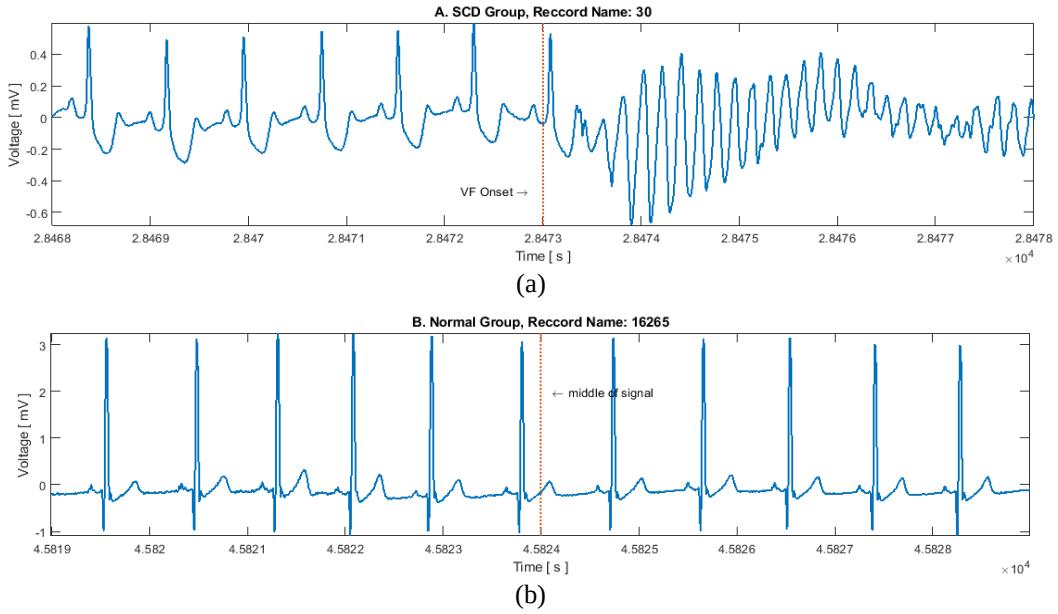

Figure 1. Two examples of ECG signals (a) 10 sec. of record 30 of SCD group around VF onset and (b) 10 sec. of record 16265 of normal group around the middle of the signal

3. ALGORITHM DEVELOPMENT

We developed an algorithm to predict SCD at early stage. Also, we have developed models using ECG signals in which different combinations of the main components of models based on machine learning are used. The overall stages of the proposed work are shown in Figure 2. In the following sections we briefly describe the aspects of the components of the models.

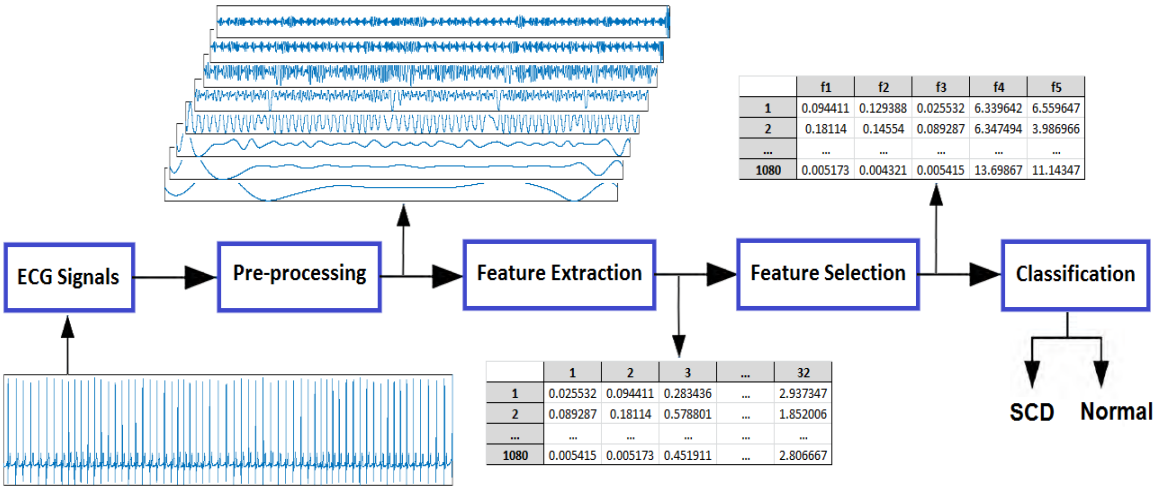

Figure 2. Proposed methodology for SCD prediction

3.1. Pre-processing of datasets

We have applied data pre-processing mechanisms mainly for two reasons: i) enable the algorithms to work on the datasets and ii) improve the quality of the datasets. Data pre-processing includes data cleaning, data reduction, feature selection, and data transformation. Data cleaning on the datasets deals with the missing values and/or reduce the noise.

In order to increase the precision and efficiency of the prediction algorithms, we designed a filter to reduce the noise of the ECG signals. For this, a Butterworth filter of order 6 is designed to pass signals with frequency higher than 0.5 Hz to remove the baseline wander and consider signals with frequency of higher than 30 Hz as noise and filter them out. Normal and SCD are transformed from categories to numerical indices. Without this transformation, the algorithms may not be able to work.

3.2. Construct a labeled database

The dataset's (ECG data) records are collected and pre-processed initially. Continuous one-minute signals image is segmented on each recording, and 30 one-minute ECG segments are generated, either classified as SCD or regular. For SCD group, the 30-minutes data are selected before VF onset and for normal group the 30-minutes data are selected from the middle of the records. At the end of this stage, we have a labeled database consisted of $30 \times (18 + 18) = 1080$ one-minute fragments which are saved and will be used for further study, i.e., to train and test our automated strategy of SCD prediction. This process is illustrated in Figure 3.

3.3. Feature extraction

In this study, two types of transforms are used to find the intrinsic attribute curves of the ECG signals: i) Hilbert-Huang transforms, or empirical mode decomposition (EMD) to find the intrinsic mode functions (IMFs); ii) wavelet transform, or multilevel 1-D wavelet decomposition (DWT) to find the approximate and detail coefficients (cAs and cDs). Figures 4 and 5 show some intrinsic curves found by aforesaid transforms for a small duration.

Then four statistical measures (mean, variance, skewness, kurtosis) are applied on the intrinsic curves found using Hilbert-Huang and wavelet transforms to specify a numerical value to each of these curves, which we use them as features. Since we initially do not know how many of features distinct enough are required to cover the whole aspects of ECG signal, we extracted quite a big number of them, 32 features. However, it is not needed to use all features.

Segmentation of the datasets
Input: ECG signals $\{(\mathbf{t}_i, \mathbf{r}_i, l_i, t_i^{VFonset}); i = 1, \dots, N, l_i \in \{Normal, SCD\}$
Output: Segmented database $\{(\mathbf{s}_i, y_i); i = 1, \dots, 30N, y_i \in \{Normal, SCD\}$
$c \leftarrow 0$
for i=1 to N **do**
  if $l_i = Normal$
    for k=1 to 30
$c \leftarrow c + 1$
$y_c \leftarrow l_i$
$\mathbf{j} \leftarrow 0.5t_i(end) + (-15 + k - 1) \times 60, \dots, 0.5t_i(end) + (-15 + k) \times 60$
$\mathbf{s}_c \leftarrow \mathbf{r}_i(\mathbf{j})$
  else
    for k=1 to 30
$c \leftarrow c + 1$
$y_c \leftarrow l_i$
$\mathbf{j} \leftarrow t_i^{VFonset} + (-30 + k - 1) \times 60, \dots, t_i^{VFonset} + (-30 + k) \times 60$
$\mathbf{s}_c \leftarrow \mathbf{r}_i(\mathbf{j})$

Figure 3. The process of segmentation of ECG signals to prepare the database

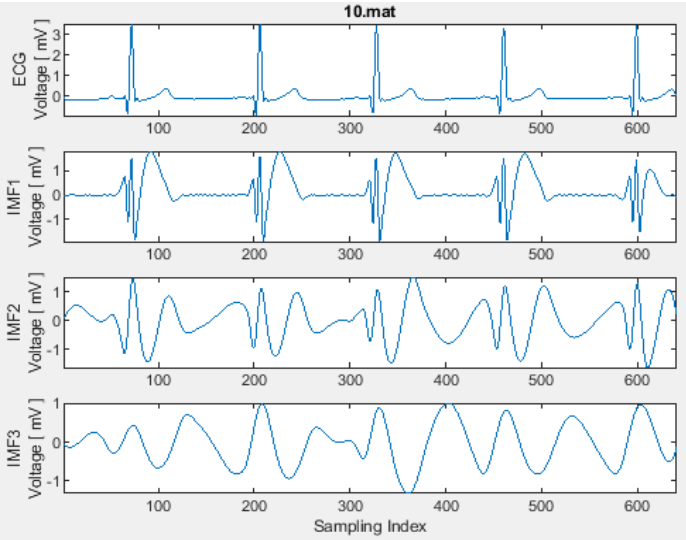

Figure 4. Some IMFs of an ECG signal (record 16265 of NSRDB), extracted using EMD

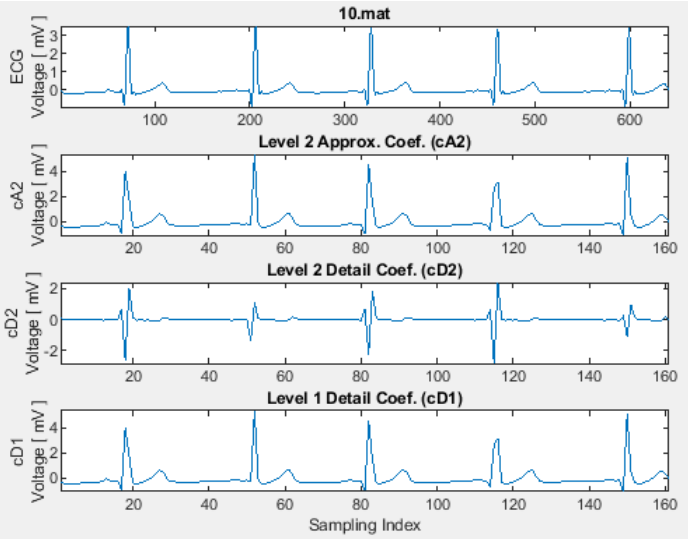

Figure 5. Some cA and cDs of an ECG signal (record 16265 of NSRDB), extracted using DWT

3.4. Feature selection

After feature extraction, we have applied three methods to rank the features: i) ANOVA (analysis of variance) and ii) ReliefF algorithm (rank importance of predictors), and iii) dCor (correlation analysis). The ANOVA using F-test technique is used to rank characteristics in terms of how significant they are in the categorization [40]. The strengths of predictors are determined using ReliefF. Predictors that offer various scores to neighbors in the very same class are penalized, while forecasters that provide different preferences to neighbors in separate classes are awarded. To rank the characteristics, dCor utilizes a measurement of interaction between matched vectors. Tables 3 and 4 demonstrate several of the features that were used in this investigation.

Table 3. Sample features extracted using EMD and ranked with correlation analysis

	Feature, X					Label			
Feature No.	2	6	1	4	8				
Feature formula	variance(lmf1)	variance(lmfz)	mean(lmf1)	kurtosis(lmf1)	kurtosis(lmfz)		row of x	record name	DB
	0.094411319	0.129388168	0.025532239	6.33964188	6.559645558	Normal	1	16265	NSRDB
	0.181140341	0.14554032	0.089286928	6.347494434	3.986966356	Normal	2	16265	NSRDB
	…	…	…	…	…	…	…	…	…
	0.01723577	0.017637224	0.000622635	9.354896841	5.765803261	Normal	540	19830	NSRDB
	0.00820838	0.024405427	0.005172282	8.178362444	7.110102235	SCD	541	30	SDDB
	…	…	…	…	…	…	…	…	…
	0.005173318	0.004320744	0.005415132	13.69867346	11.14347478	SCD	1080	52	SDDB

Table 4. Sample features extracted using DWT and ranked with ANOVA

	Feature, X					Label			
Feature No.	2	6	1	4	8				
Feature formula	kurtosis(cD7)	kurtosis(cD6)	kurtosis(cD5)	variance(cD1)	variance(cA7)		row of x	record name	DB
	16.43466415	14.68811494	7.884046875	0.345274906	0.345274906	Normal	1	16265	NSRDB
	20.41299529	18.90730632	9.593829463	0.306361446	0.306361446	Normal	2	16265	NSRDB
	…	…	…	…	…	…	…	…	…
	25.92484865	14.38789522	5.56937062	0.077736297	0.077736297	Normal	540	19830	NSRDB
	31.7408592	15.51188951	5.659636272	0.134887891	0.134887891	SCD	541	30	SDDB
	…	…	…	…	…	…	…	…	…
	38.5807801	25.3291934	22.97580617	1.289219856	1.289219856	SCD	1080	52	SDDB

3.5. Classification

To classify ECG data into the regular and SCD groups, we utilized 6 major classifiers. The best classification with the maximum accuracy was then chosen. These classifiers are: KNN, SVM, NB, DT, Disc., and Ens. One such algorithm that is used in DT is shown in Figure 6.

Algorithm AdaBoostM1

Input: Dataset $S = \{\mathbf{x}_i, y_i\}; i = 1, \dots, N; y_i \in \{-1, +1\}$,
T: Number of learners,
W: Algorithm of weak learner
Output: Boosted classifier $f(\mathbf{x}) = \sum_{t=1}^{T} a_t h_t(\mathbf{x})$
$d_i^{(1)} \leftarrow \frac{1}{N}; i = 1, \dots, N$
for *t*=1 to *T* **do**
$h_t \leftarrow W(S, d^{(t)})$
$k_i^{(t)} \leftarrow 0; i = 1, \dots, N$
if $y_i \neq h_t(\mathbf{x}_i); i = 1, \dots, N$
$k_i^{(t)} \leftarrow 1$
$e_t = \sum_{i=1}^{N} d_i^{(t)} k_i^{(t)}$
if $e_t > 0.5$ **then**
$T \leftarrow t - 1$
return
$a_t \leftarrow 0.5 \ln(e_t^{-1} - 1)$
$d_i^{(t+1)} \leftarrow d_i^{(t)} \exp(-a_t h_t(\mathbf{x}_i) y_i); i = 1, \dots, N$
$d_i^{(t+1)} \leftarrow \frac{d_i^{(t+1)}}{\sum_{i=1}^{N} d_i^{(t+1)}}; i = 1, \dots, N$

Figure 6. Algorithm of AdaBoostM1

3.6. 5-fold cross-validation

The data are partitioned for classification. In other words, we have employed 5-fold cross-validation. It means, we partition the instances to 5 parts and use 4 parts for training and 1 part for testing. Therefore, 80% of the instances are used for training and 20% of them are used for testing. This scheme is repeated for all parts, i.e., 5 times.

4. RESULTS AND DISCUSSION

4.1. Performance measures

As mentioned before, we employed 5-fold cross validation method to build and evaluate the performance of the classifiers used. After finding the model some metrics are used to evaluate the performance of models found using the classifiers. For this purpose, widely used classification performance measures, sensitivity (or recall), specificity, accuracy, precision, and F1-score are computed by (1)-(5). They are used to evaluate the performance of the proposed methods for prediction of SCD.

$$Sensitivity = \frac{TP}{TP+FN} \quad (1)$$

$$Specificity = \frac{TN}{TN+FP} \quad (2)$$

$$Accuracy = \frac{TN+TP}{TN+TP+FP+FN} \quad (3)$$

$$Pr\,e\,cision = \frac{TP}{TP+FP} \quad (4)$$

$$F1-score = \frac{2.Pr\,ecision.Sensitivity}{Pr\,ecision+Sensitivity} \quad (5)$$

In which the true positive (TP), true negative (TN), false positive (FP), and false negative (FN) are the components of the confusion matrix and their meanings are summarized in Table 5.

Table 5. Meanings of the components of the confusion matrix

Term	Meaning	Meaning in this work
TP	True Positive	The number of SCD which are recognized as SCD
TN	True Negative	The number of not SCD which are recognized as not SCD
FP	False Positive	The number of not SCD which are recognized as SCD
FN	False Negative	The number of SCD which are recognized as not SCD

All the 1080 one-minute fragments are used to train and test our automated strategy of the SCD prediction. These fragments are passed through a Butterworth filter to reduce the noise. Then they are supplied to the transforms and found the intrinsic curves. The intrinsic curves are supplied to 4 statistical measures and the results are used as extracted features. The extracted features are ranked and selected for classification. Then the classifiers are employed to find and test the models. The details of these stages are explained in the previous sections. Table 6 summarize the average performances of the various models we used by different combinations of the feature extractions, feature selections, and classifiers. As it can be seen, the best performance we got is 100% accuracy.

Detailed statistics of performance measures are presented in Table 6. With DWT as feature extraction technique, ANOVA as feature ranking method and KNN classifier shows the best performance measures (100%) for 6 features. Whereas, for 5 features-(DWT, dCor, Ens.), (DWT ReliefF, KNN) and (DWT, ReliefF, Ens) show best performance measures of almost 100% and at the same time, other classifiers were exhibiting poor performance. It is seen that DWT as a feature extraction technique has performed the best when compared with EMD. KNN and ensemble of classifiers have shown the best performance among other classifiers, where as in feature ranking, all the three have performed well when ranked with a smaller number of features i.e., 5 and 6.

4.2. Comparative study

Table 7 (in appendix) compares the performance of numerous studies that was using ECG or HRV data to identify SCD. We found 7 papers from 2015 to 2019 that can provide specific information on the research in three areas: data material, technique, and results. We specified the sorts of signals used the

databases from which the data was acquired, and the length of signals (in minutes) used in each study in terms of materials. In terms of methodologies, we mention the techniques applied by each work separately, and the number of markers and the classification. We offer performance data in terms of accuracy, specificity, and sensitivity.

Table 6. The statistical measures and the average performance of methods we used

Method	Feature ranking	No. of features	TP	TN	FP	FN	Sensitivity	Specificity	Accuracy	Precision	F1-score
EMD+KNN	ANOVA	10	444	499	41	96	82.2	92.4	87.3	91.5	86.6
EMD+SVM	ANOVA	10	442	471	69	98	81.9	87.2	84.5	86.5	84.1
EMD+Ensemble	ANOVA	10	534	533	7	6	98.9	98.7	98.8	98.7	98.8
EMD+NB	ANOVA	10	447	466	74	93	82.8	86.3	84.5	85.8	84.3
EMD+DT	ANOVA	10	433	490	50	107	80.2	90.7	85.5	89.6	84.7
EMD+Discrement	ANOVA	10	357	496	44	183	66.1	91.9	79.0	89.0	75.9
DWT+KNN	ANOVA	6	540	540	0	0	100.0	100.0	100.0	100.0	100.0
DWT+SVM	ANOVA	6	501	498	42	39	92.8	92.2	92.5	92.3	92.5
DWT+Ensemble	ANOVA	6	534	531	9	6	98.9	98.3	98.6	98.3	98.6
DWT+NB	ANOVA	6	511	524	16	29	94.6	97.0	95.8	97.0	95.8
DWT+DT	ANOVA	6	533	522	18	7	98.7	96.7	97.7	96.7	97.7
DWT+Discrement	ANOVA	6	455	530	10	85	84.3	98.1	91.2	97.8	90.5
EMD+KNN	Correlation analysis	5	526	518	22	14	97.4	95.9	96.7	96.0	96.7
EMD+SVM	Correlation analysis	5	496	505	35	44	91.9	93.5	92.7	93.4	92.6
EMD+Ensemble	Correlation analysis	5	536	538	2	4	99.3	99.6	99.4	99.6	99.4
EMD+NB	Correlation analysis	5	523	447	93	17	96.9	82.8	89.8	84.9	90.5
EMD+DT	Correlation analysis	5	510	523	17	30	94.4	96.9	95.6	96.8	95.6
EMD+Discrement	Correlation analysis	5	511	461	79	29	94.6	85.4	90.0	86.6	90.4
DWT+KNN	Correlation analysis	5	534	537	3	6	98.9	99.4	99.2	99.4	99.2
DWT+SVM	Correlation analysis	5	535	537	3	5	99.1	99.4	99.3	99.4	99.3
DWT+Ensemble	Correlation analysis	5	540	540	0	0	100.0	100.0	100.0	100.0	100.0
DWT+NB	Correlation analysis	5	522	540	0	18	96.7	100.0	98.3	100.0	98.3
DWT+DT	Correlation analysis	5	537	536	4	3	99.4	99.3	99.4	99.3	99.4
DWT+Discrement	Correlation analysis	5	521	536	4	19	96.5	99.3	97.9	99.2	97.8
EMD+KNN	ReliefF	5	510	518	22	30	94.4	95.9	95.2	95.9	95.1
EMD+SVM	ReliefF	5	506	515	25	34	93.7	95.4	94.5	95.3	94.5
EMD+Ensemble	ReliefF	5	519	517	23	21	96.1	95.7	95.9	95.8	95.9
EMD+NB	ReliefF	5	382	524	16	158	70.7	97.0	83.9	96.0	81.4
EMD+DT	ReliefF	5	507	523	17	33	93.9	96.9	95.4	96.8	95.3
EMD+Discrement	ReliefF	5	499	480	60	41	92.4	88.9	90.6	89.3	90.8
DWT+KNN	ReliefF	5	540	540	0	0	100.0	100.0	100.0	100.0	100.0
DWT+SVM	ReliefF	5	539	540	0	1	99.8	100.0	99.9	100.0	99.9
DWT+Ensemble	ReliefF	5	540	540	0	0	100.0	100.0	100.0	100.0	100.0
DWT+NB	ReliefF	5	525	539	1	15	97.2	99.8	98.5	99.8	98.5
DWT+DT	ReliefF	5	533	537	3	7	98.7	99.4	99.1	99.4	99.1
DWT+Discrement	ReliefF	5	537	538	2	3	99.4	99.6	99.5	99.6	99.5

5. CONCLUSION AND FUTURE WORK

We have proposed an automated method for predicting SCD utilizing statistical measures in this paper. Different classifiers are used in this work. With an accuracy rate of 100.0% (KNN), 98.5% (NB), 100.0% (Disc.), 99.9% (SVM), 99.4% (DT), 98.5% (NB), and the system can predict SCD very quickly, that is 30 minutes before it occurs (Ens.). From the studies, it is observed that prediction of SCD is always a challenging task. The current work is carried based on the ECG databases available publicly, i.e., NSRDB and are of relatively small databases. More research is required to acquire a large amount of clinical data to educate the suggested classifications and to assess their validity. Additionally, it should be noted that the methodology used in this study must be tested in clinical conditions. Furthermore, such projections might be more realistic and useful if they have been made in real time on mobile devices in hospitals or at home.

APPENDIX

Table 7. A comparison of our research with some other previous ECG or HRV signal-based approaches

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) [31]	ECG	SDDB NSRDB	4 minutes before SCD	Nonlinear features (18) and SCDI	DT, SVM	92.11%	92.50%	91.67%
Fujita (2016) [27]	HRV	SDDB NSRDB	4 minutes before SCD	Nonlinear features (4), nonlinear heart rate variability analysis	SVM, KNN	94.70%	95.00%	94.40%
Sanchez (2018) [27]	ECG	SDDB NSRDB	20 minutes before SCD	Nonlinear methods HI, Wave packet transform	EPNN	95.80%	unknown	unknown
Khazaei (2018) [21]	HRV	SDDB NSRDB	6 minutes before SCD	Wave packet transform RQA (13) and increment entropy (2 out of 14) Nonlinear method	DT, KNN, SVM, NB	95.00%	95.00%	95.00%
Ebrahimzadeh (2018) [19]	HRV	SDDB NSRDB	12 minutes before SCD	HRV features (23) Time local subset feature selection	MLP	88.29%	unknown	unknown
Ebrahimzadeh (2019) [20]	HRV	SDDB NSRDB	13 minutes before SCD	HRV features (23) time local subset feature selection	MLP	90.18%	unknown	unknown
Lai (2019) [35]	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%
Present work	ECG	SDDB NSRDB	30 minutes before SCD	Nonlinear (5) (EMD and DWT)	KNN, SVM, NB, DT, Dis, Ens	100%	100%	100%

REFERENCES

[1] A. G. Yow, V. Rajasurya, and S. Sharma, "Sudden cardiac death," *StatPearls*. 2021, Accessed: Mar. 10, 2022. [Online]. Available: https://www.ncbi.nlm.nih.gov/books/NBK507854/

[2] D. P. Zipes and H. J. J. Wellens, "Sudden cardiac death," *Circulation*, vol. 98, no. 21, pp. 2334–2351, Nov. 1998, doi: 10.1161/01.CIR.98.21.2334.

[3] R. J. Myerburg and A. Castellanos, "Cardiac arrest and sudden cardiac death," in *Braunwald's Heart Disease: A Textbook of Cardiovascular Medicine*, Elsevier, 2012, pp. 845–884.

[4] R. M. John *et al.*, "Ventricular arrhythmias and sudden cardiac death," *The Lancet*, vol. 380, no. 9852, pp. 1520–1529, Oct. 2012, doi: 10.1016/S0140-6736(12)61413-5.

[5] T.-W. Shen, H.-P. Shen, C.-H. Lin, and Y.-L. Ou, "Detection and prediction of sudden cardiac death (SCD) for personal healthcare," in *2007 29th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, Aug. 2007, pp. 2575–2578, doi: 10.1109/IEMBS.2007.4352855.

[6] 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, Feb. 2013, doi: 10.1161/CIRCULATIONAHA.112.128413.

[7] G. Finocchiaro, M. Papadakis, S. Sharma, and M. Sheppard, "Sudden cardiac death," *European Heart Journal*, vol. 38, no. 17, pp. 1280–1282, May 2017, doi: 10.1093/eurheartj/ehx194.

[8] R. Passman, "Prevention of sudden cardiac death in dialysis patients: drugs, defibrillators or what else," *Blood Purification*, vol. 35, no. 1–3, pp. 49–54, 2013, doi: 10.1159/000345178.

[9] A. Cheng *et al.*, "Prospective observational study of implantable cardioverter-defibrillators in primary prevention of sudden cardiac death: study design and cohort description," *Journal of the American Heart Association*, vol. 2, no. 1, Jan. 2013, doi: 10.1161/JAHA.112.000083.

[10] S. S. Chugh, "Sudden cardiac death with apparently normal heart: Clinical implications of progress in pathophysiology," *Cardiac Electrophysiology Review*, vol. 5, no. 4, p. 394, 2001.

[11] G. I. Fishman *et al.*, "Sudden cardiac death prediction and prevention," *Circulation*, vol. 122, no. 22, pp. 2335–2348, Nov. 2010, doi: 10.1161/CIRCULATIONAHA.110.976092.

[12] Fei Zhang, Pengye Li, Fan Jiang, and Dakun Lai, "A shockable rhythm detection algorithm for automatic external defibrillators by combining a slope variability analyzer with a band-pass digital filter," in *2014 IEEE Workshop on Electronics, Computer and Applications*, May 2014, pp. 828–831, doi: 10.1109/IWECA.2014.6845749.

[13] R. Passman and J. J. Goldberger, "Predicting the future," *Circulation*, vol. 125, no. 24, pp. 3031–3037, Jun. 2012, doi: 10.1161/CIRCULATIONAHA.111.023879.

[14] E. F. Aziz, F. Javed, B. Pratap, and E. Herzog, "Strategies for the prevention and treatment of sudden cardiac death," *Open Access Emergency Medicine*, Dec. 2010, doi: 10.2147/OAEM.S6869.

[15] S. D. Solomon *et al.*, "Sudden death in patients with myocardial infarction and left ventricular dysfunction, heart failure, or both," *New England Journal of Medicine*, vol. 352, no. 25, pp. 2581–2588, Jun. 2005, doi: 10.1056/NEJMoa043938.

[16] R. Pacher *et al.*, "Prognostic impact of big endothelin-1 plasma concentrations compared with invasive hemodynamic evaluation in severe heart failure," *Journal of the American College of Cardiology*, vol. 27, no. 3, pp. 633–641, Mar. 1996, doi: 10.1016/0735-1097(95)00520-X.

[17] S. Mandala and T. C. Di, "ECG parameters for malignant ventricular arrhythmias: a comprehensive review," *Journal of Medical and Biological Engineering*, vol. 37, no. 4, pp. 441–453, Aug. 2017, doi: 10.1007/s40846-017-0281-x.

[18] Y. Castro-Torres, "Ventricular repolarization markers for predicting malignant arrhythmias in clinical practice," *World Journal of Clinical Cases*, vol. 3, no. 8, p. 705, 2015, doi: 10.12998/wjcc.v3.i8.705.

[19] 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," *Medical and Biological Engineering and Computing*, vol. 56, no. 7, pp. 1253–1270, Jul. 2018, doi: 10.1007/s11517-017-1764-1.

[20] E. Ebrahimzadeh *et al.*, "An optimal strategy for prediction of sudden cardiac death through a pioneering feature-selection approach from HRV signal," *Computer Methods and Programs in Biomedicine*, vol. 169, pp. 19–36, Feb. 2019, doi:

10.1016/j.cmpb.2018.12.001.
[21] M. Khazaei, K. Raeisi, A. Goshvarpour, and M. Ahmadzadeh, "Early detection of sudden cardiac death using nonlinear analysis of heart rate variability," *Biocybernetics and Biomedical Engineering*, vol. 38, no. 4, pp. 931–940, 2018, doi: 10.1016/j.bbe.2018.06.003.
[22] G. Tse and B. P. Yan, "Traditional and novel electrocardiographic conduction and repolarization markers of sudden cardiac death," *EP Europace*, vol. 19, no. 5, pp. 712–721, May 2017, doi: 10.1093/europace/euw280.
[23] D. P. Morin *et al.*, "QRS duration predicts sudden cardiac death in hypertensive patients undergoing intensive medical therapy: the LIFE study," *European Heart Journal*, vol. 30, no. 23, pp. 2908–2914, Dec. 2009, doi: 10.1093/eurheartj/ehp321.
[24] K. S. Spargias *et al.*, "QT dispersion as a predictor of long-term mortality in patients with acute myocardial infarction and clinical evidence of heart failure," *European Heart Journal*, vol. 20, no. 16, pp. 1158–1165, Aug. 1999, doi: 10.1053/euhj.1998.1445.
[25] R. L. Verrier and T. Ikeda, "Ambulatory ECG-based T-wave alternans monitoring for risk assessment and guiding medical therapy: mechanisms and clinical applications," *Progress in Cardiovascular Diseases*, vol. 56, no. 2, pp. 172–185, Sep. 2013, doi: 10.1016/j.pcad.2013.07.002.
[26] I. Cygankiewicz *et al.*, "Heart rate turbulence predicts all-cause mortality and sudden death in congestive heart failure patients," *Heart Rhythm*, vol. 5, no. 8, pp. 1095–1102, Aug. 2008, doi: 10.1016/j.hrthm.2008.04.017.
[27] H. Fujita *et al.*, "Sudden cardiac death (SCD) prediction based on nonlinear heart rate variability features and SCD index," *Applied Soft Computing*, vol. 43, pp. 510–519, Jun. 2016, doi: 10.1016/j.asoc.2016.02.049.
[28] M. T. La Rovere, J. T. Bigger, F. I. Marcus, A. Mortara, and P. J. Schwartz, "Baroreflex sensitivity and heart-rate variability in prediction of total cardiac mortality after myocardial infarction," *The Lancet*, vol. 351, no. 9101, pp. 478–484, Feb. 1998, doi: 10.1016/S0140-6736(97)11144-8.
[29] R. Liew, "Electrocardiogram-based predictors of sudden cardiac death in patients with coronary artery disease," *Clinical Cardiology*, vol. 34, no. 8, pp. 466–473, Aug. 2011, doi: 10.1002/clc.20924.
[30] H. Evrengul *et al.*, "The relationship between heart rate recovery and heart rate variability in coronary artery disease," *Annals of Noninvasive Electrocardiology*, vol. 11, no. 2, pp. 154–162, Apr. 2006, doi: 10.1111/j.1542-474X.2006.00097.x.
[31] U. R. Acharya *et al.*, "An integrated index for detection of sudden cardiac death using discrete wavelet transform and nonlinear features," *Knowledge-Based Systems*, vol. 83, pp. 149–158, Jul. 2015, doi: 10.1016/j.knosys.2015.03.015.
[32] J. P. 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," *Journal of Medical Systems*, vol. 42, no. 10, Oct. 2018, doi: 10.1007/s10916-018-1031-5.
[33] D. Ghista, *Applied biomedical engineering mechanics*. CRC Press, 2008.
[34] U. R. Acharya *et al.*, "An integrated diabetic index using heart rate variability signal features for diagnosis of diabetes," *Computer Methods in Biomechanics and Biomedical Engineering*, vol. 16, no. 2, pp. 222–234, 2013, doi: 10.1080/10255842.2011.616945.
[35] D. Lai, Y. Zhang, X. Zhang, Y. Su, and M. B. Bin Heyat, "An automated strategy for early risk identification of sudden cardiac death by using machine learning approach on measurable arrhythmic risk markers," *IEEE Access*, vol. 7, pp. 94701–94716, 2019, doi: 10.1109/ACCESS.2019.2925847.
[36] J. R. Mohammed, "Low complexity adaptive noise canceller for mobile phones based remote health monitoring," *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 4, no. 3, pp. 422–432, Jun. 2014, doi: 10.11591/ijece.v4i3.5534.
[37] N. E. Huang and N. O. Attoh-Okine, Eds., *The Hilbert-Huang transform in engineering*. CRC Press, 2005.
[38] T. L. T. da Silveira, A. J. Kozakevicius, and C. R. Rodrigues, "Single-channel EEG sleep stage classification based on a streamlined set of statistical features in wavelet domain," *Medical and Biological Engineering and Computing*, vol. 55, no. 2, pp. 343–352, Feb. 2017, doi: 10.1007/s11517-016-1519-4.
[39] A. L. Goldberger *et al.*, "PhysioBank, PhysioToolkit, and PhysioNet," *Circulation*, vol. 101, no. 23, Jun. 2000, doi: 10.1161/01.CIR.101.23.e215.
[40] A. Grünauer and M. Vincze, "Using dimension reduction to improve the classification of high-dimensional data," *arxiv.org/abs/1505.06907*, May 2015.

BIOGRAPHIES OF AUTHORS

Karna Viswavardhan Reddy received the Bachelor of Technology, degree in Electronics and Communication Engineering from JNTU-Hyderabad, A. P, India, in 2008, the M. S. degree in Telecommunication Systems from Blekinge Institute of Technology Karlskrona, Sweden, in 2011. Currently, he is a research assistant at Department of Electronics Engineering, Jain University, Bengaluru, India. His research interests include machine learning, artificial intelligence, wireless body area networks, wireless sensor networks, and software defined networks. He can be contacted at email: viswavardhanreddy@gmail.com, viswavardhank@rvce.edu.in.

Navin Kumar graduated from a joint PhD program from the university of Minho, Aveiro and Porto (MAP Tele) of Portugal in Europe from 2007-2011 and worked as research collaborator at Institution of telecommunication, Aveiro, Portugal until 2012. Currently, serves as head of the department at the Department of Electronics and Communication, School of Engineering, Amrita Vishwa Vidyapeetham, Bengaluru, India. He is Sr Member of IEEE, AIENG (HK), a Life Member of IETE and Fellow IE (India). His research areas include 5G (mmWave and massive MIMO), visible light communication, optical wireless communication, IoT health care and smart city and intelligent transportation systems. He is also the part of IEEE 5G and future network initiative. He can be contacted at email: navinkumar@ieee.org, navin_kum3@yahoo.com.