Evaluation of conventional and new risk factors for the prediction of sudden cardiac arrest using ML Tools

[1]V.V.R.L.S.Gangadhar, [2]Dr G Rajesh Chandra
[1]Professor, Department of CSE, SLC'S Institute of engineering and technology, Rangareddy Dist, Telangana
[2]Professor, Department of CSE, KKR and KSR INSTITUTE of technology and sciences, Guntur, Andhra Pradesh

ABSTRACT— One of the categories of natural fatalities from cardiac causes is sudden cardiac death. Nearly an hour or few minutes before to the occurrence, the signs of death start to manifest. It's known as sudden cardiac arrest when symptoms first appear. It may manifest as a result of a previous heart condition that was harmful or occasionally without any known cardiac aetiology. Researchers and physiologists throughout the world are working on problems and publishing every answer they can think of to anticipate sudden cardiac death at an early stage. Heart rate variability (HRV) and all of its potential factors have recently been employed extensively for the prediction of SCD. Examples of these classifiers are k-nearest neighbour (k-NN), support vector machine (SVM), multi-layer perceptron (MLP), etc. However, practical implications of such techniques are still debatable because they incorrectly identify QRS peaks in SCD patients' ECG data. This study uses heart rate variability analysis in an effort to predict sudden cardiac death at an early stage, or one hour before it occurs. The ECG signals for healthy subjects with normal sinus rhythm and sudden cardiac death are acquired from an internet database (SCD subject). The classification of several derived HRV values using a k-NN classifier greatly supports the prediction of SCD at this advanced stage. The impact of inaccurate QRS peak detection on heart rate (measured in beats per minute, or BPM) is heavily taken into consideration for clinical applications of such technologies.

KEYWORDS: Early-stage forecasting, SCD, BPM, temporal domain, and frequency domain, heart rate variability.

I. INTRODUCTION

The sudden and natural death of a patient owing to cardiac reasons is known as sudden cardiac death (SCD). Patient passes away within an hour after the onset of acute symptoms due to sudden loss of consciousness (very serious cardiac event termed as sudden cardiac arrest). Numerous cardiac reasons have been linked to SCD, which has irregular/interrupted heart action as its source [1–5]. According to estimates, SCD claims more than 7 million lives annually, including over 300,000 in just the United States [6]. The annual incidence of SCD in India is estimated to be around 7 lakhs based on data from the study conducted by Rao et al. in 2012 and extrapolation of these data to national mortality estimates [7]. According to several studies, the annual incidence of SCD ranges from 36 to 128 per 100,000 people [8]. These numbers indicate that SCD causes between one and two deaths per 1000 people in the general population [8]. Numerous heart disorders are the primary causes of sudden cardiac death (SCD), according to epidemiological studies on the subject. These include congenital cardiac anomalies, valvular heart disease, cardiomyopathies, electrophysiologic abnormalities, coronary artery disease (CAD), etc. [9]. In addition to this, investigations may take into account gender, nonspecific cardiovascular disease, lifestyle, and psychosocial factors [9]. Since there are no unique regional, gender, or socioeconomic causes for heart illness, it is crucial to identify cardiac abnormalities early and administer prompt, effective treatment. Heart rate variability

(HRV) is widely used in the literature as a reliable indicator of cardiac health [10–14]. It is known as beat-to-beat variability, or HRV, and is described as a plot of inter beat intervals (IBI).

The interaction of many branches of the human central nervous system (CNS) and respiratory cardiovascular system (RCV) system results in this beat-to-beat variability in heart rate. It is derived from an electrocardiogram (ECG) and serves as the primary gauge of someone's heart health. To the best of the authors' knowledge, six to one minutes of warning before the occurrence of SCD have not yet been reported in the literature [13]. However, there are certain restrictions in this research regarding the data selection and obtaining the heart rate variability signal. According to the established requirements of HRV [15], the ECG signal should be at least 2 minutes long in order to extract the LF and HF powers in the frequency domain of HRV. Similar to this, the chosen ECG signal's duration should be at least 20 minutes in order to extract the VLF component of the PSD. Additionally, the computation must take into account both the LF (n.u) and HF (n.u) in order to properly understand the results using LF and HF powers in ms2 units. However, these recommendations were not taken into account in the studies [11], [13] that claimed a prediction of SCD had an accuracy of more than 91%. Just 4 minutes prior to the incidence of sudden cardiac death, Elias Ebrahimzadeh et al. [11] performed prediction of the event. K-NN and MLP classifiers are used to extract and categorise time-frequency and nonlinear information.

However, the ECG signals chosen for simulation are just one minute long, which does not adhere to the rules that have been made public. In a similar vein, Houshyarifar Vahid et al. carried out the same simulation to forecast SCD 6 minutes in advance. However, the ECG signal that was chosen for bispectrum and time-domain feature extraction does not adhere to the criteria. Additionally, these studies are constrained by the use of conventional techniques [16] for the identification of QRS complexes in SCD patients. While eliminating the shortcomings of earlier studies, an attempt has been made in the present work to predict SCD an hour before its actual incidence. It also explains the practical difficulties in detecting QRS complexes using conventional methods.

There are five sections to this essay. The problem's introduction is presented in Section I. The research work's materials and methodology are covered in Section II. Results from subjects with normal and SCD are reported in Section III. Section IV discusses the limitations of the earlier investigations as well as a discussion of QRS complexes. The task is concluded in Section V.

II. RELATED WORK

When class distributions are severely skewed, data in datasets are frequently unbalanced in the medical profession, and this negatively impacts performance [12]. To put it more specifically, the chosen feature has an impact on how accurate the training model's output is. A class imbalance problem arises when a dataset is dominated by large classes or classes that have a disproportionately high number of instances compared to other uncommon or minority classes. Without sacrificing generality, one can infer that in the two-class scenario, the minority or rare class is the positive class and the majority class is the negative class. The minority class is frequently extremely rare, making up less than 1% of the dataset. The dataset's most common classifiers are likely to predict everything as negative if used (the majority class). People are often more interested in knowing about uncommon classes, though. Applications

like cancer diagnostic prediction, for instance, where it is typical to have fewer cancer patients than healthy people, are rare but significant diseases like cancer.

Some strategies can be used at the pre-processing step to address the imbalance issue. Data sampling, which enables classifiers to perform better by producing a more or less balanced class distribution by modifying the training samples. Traditional data sampling methods include under sampling, which reduces the original dataset to a smaller number of instances, oversampling, which increases the original dataset's number of instances while balancing the skewed class ratio by replicating some instances or generating new instances from existing ones, and resampling, which combines under sampling and oversampling. The common techniques are known as Synthetic Minority Oversampling Technique (SMOTE) [13], Borderline-SMOTE [14], Adaptive Synthetic Sampling Approach (ADASYN) [15], Random Sampling, and stratified sampling. These methods can be used to oversample, resample, or under-sample a dataset used in the typical classification problem. One of the most popular resampling methods is called SMOTE, which involves taking each sample from the minority class and adding synthetic examples along the line segments connecting any or all of the k minority class nearest neighbours. This method oversamples the minority class. Additionally, for a binary class problem, the ratio of the sample size of the minor class to that of the major class can be used to indicate the degree of a class distribution imbalance [16]. The ratio can be as extreme as 1:100, 1:1000, or even higher in real-world situations [17]. Additionally, one study examines the connection between the classification abilities of decision trees and the class distribution of a training dataset. Accordingly, the aforementioned research show that a generally balanced distribution typically yields a better outcome and that some methods may become inadequate for creating a good model at a ratio as low as 1:35 [16].
We use under sampling to address the issue of an unbalanced dataset, shifting windows with and without overlap, AUROC curves, and the F3 score to evaluate performance in our concurrent but earlier published research [13]. If recall is a concern, then using the F3 score as a measurement of the application is functional or necessary.

III. METHODS

ECG signals for healthy patients (normal sinus rhythm database) and victims of sudden cardiac death are downloaded from database [17]. (Sudden cardiac death holter database). The datasets are made up of two ECG signals with a duration of around an hour for each patient. As a result, 82 signals total—46 signals for SCD patients (23 subjects with 2 ECG signals each) and 36 signals for normal participants (also 18 subjects with 2 ECG signals each)—have been analysed. For healthy patients, the sample frequency is 128 Hz, but for SCD, it is 250 Hz. For the purpose of predicting the occurrence of sudden cardiac death, the initial 4 min of the SCD dataset's ECG signal were obtained. Fast Fourier Transform is used to remove baseline wandering noise from the signals, while Notch Filter is used to remove power line interference. The heart rate variability signal is derived in order to obtain its nonlinear, time-domain, and frequency-domain properties. The flowchart in Fig. 1 depicts the stages in detail. The following is an explanation of the parameters taken from the HRV signal:

1. Mean RR: $1/N \Sigma RR(i)$
2. SDNN: $1/N \Sigma (RR(i) - Mean\ RR)^2$
3. RMSSD = $1/N \Sigma (RR(i+1) - RR(i))^2$

4. LF (ms^2) = Distribution of power in low frequency region i.e. 0.04–0.15 Hz of HRV
5. HF (ms^2) = Distribution of power in high frequency region i.e. 0.15–0.4 Hz of HRV
6. LF/HF Ratio = LF (n.u.)/HF (n.u.) where LF (n.u.) = LF (ms^2)/Total Power-VLF
7. HF (n.u.) = HF (ms^2)/Total Power-VLF
8. SD1 and SD2 = Extracted from Poincare plots.

Following parameter extraction, the parameters are then categorised using a k-NN classifier. Based on its training sets, the classifier divides the signals into normal and SCD subjects.

IV. RESULTS OBTAINED FOR PREDICTION

As seen in Table I, the derived HRV parameters for SCD participants differ considerably from those for normal subjects in the time domain (Mean RR, SDNN, and RMSSD), frequency domain (LF (ms2), HF (ms2), and parameters (SD1 and SD2) extracted from Poincare plot. Therefore, the values of the various indicators confirm the risk factor of SCD one hour before the actual incidence.

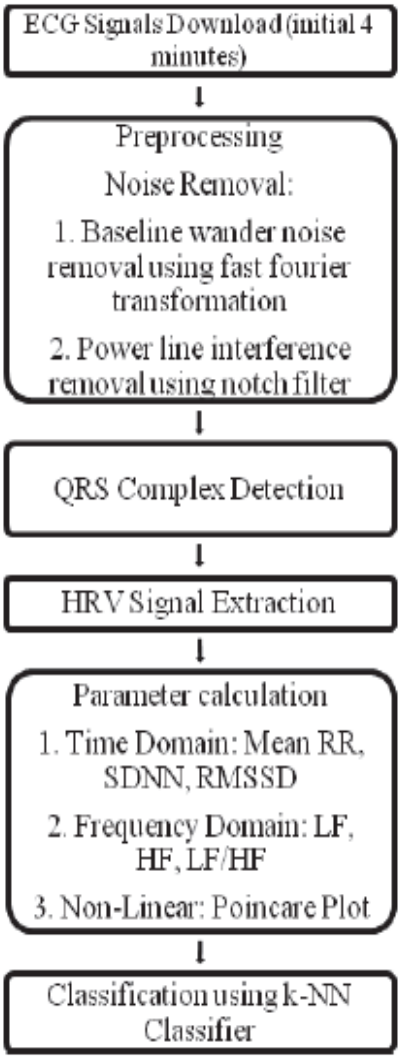

Fig. 1. Flow-chart for processing of ECG signal for prediction of SCD.

TABLE I: DIFFERENT PARAMETERS OF HRV FOR NORMAL SUBJECTS AND SCD PATIENTS.

Parameter	Normal subjects	SCD patients
Mean RR	780.89	740.46
SDNN	44.75	12.87
RMSSD	28.26	12.39
LF (ms^2)	1047.3	52.5
HF (ms^2)	324	51.1
Ratio LF/HF	4.9237	1.8609
SD1	20.19	8.79
SD2	59.4	15.36

V. DISCUSSION ON DETECTION OF QRS PEAKS IN SCD SIGNALS

The predominant peaks in the ECG signal are QRS peaks. Finding this peak serves as the foundation for all study. Pan-Tompkins created a common algorithm for QRS peak identification in 1985. This approach is used in the majority of the literature-based studies on ECG signals. All reported efforts that predict SCD via HRV analysis use the Pan-Tompkins algorithm to find QRS peaks. The next step in this method is to identify the signal's highest peak, which is then amplified to become a threshold value. The current work demonstrates the identical process of locating the greatest peak, amplification with a weight of 0.6, and application to all samples being examined. The initial one-minute samples of the SCD data were acquired for this purpose. The initial one-minute samples of the SCD data were acquired for this purpose. The values collected for each record are displayed in Table II. The mean (e1) error value is the difference between the original signal that was downloaded and the baseline filtered signal. The difference between the baseline filtered signal and the power-line interference filtered signal yields the mean (e2) value of error. The amount of noise removed from the signal is indicated by these two parameters. BPM is calculated by dividing the highest number of R points found in the signal by the signal's minutes-long duration. Table II clearly shows that the significance of the threshold value used in this manner for detection. The record numbers 33, 39, 43, and 47's bolded entries in the table reveal the peaks that were detected above a threshold value of about 0.1 or 0.2 mV, which is too low to ever be a R peak. The same is true for record number 50, where the bolded entry indicates a threshold value of roughly 15.5 mV and the algorithm discovers all peaks above this value. Again, the amplitude of a R peak cannot be greater than 15.5 mV. These are the limits of the studies that were done to predict sudden cardiac death using an ECG signal that was recorded for one minute. Any study carrying out this kind of research must cite the technique for locating R peaks in the algorithm and must, therefore, adhere to the criteria and prescriptions that have been published.

TABLE II
EFFECT OF QRS DETECTION ON BEATS PER MINUTE (BPM).

Record No. (single lead ECG signal)	Mean (e1)	Mean (e2)	Threshold	BPM
30	0.0995	0.0103	0.5511	42
31	0.0415	0.0059	0.3116	49
32	0.0322	0.0084	0.3914	87
33	**0.0359**	**0.0084**	**0.2087**	**42**
34	0.2778	0.0145	0.9848	77
35	0.0563	0.0196	0.8837	84
36	0.0698	0.0094	0.4275	13
37	0.0763	0.0052	0.5977	34
38	0.1027	0.0190	0.5043	05
39	**0.0380**	**0.0038**	**0.1293**	**57**
40	0.0506	0.0184	0.3890	85
41	0.0438	0.0082	0.3173	03
42	0.2624	0.0916	2.1033	331
43	**0.0280**	**0.0033**	**0.1531**	**01**
44	0.0435	0.0152	1.1056	65
45	0.1718	0.0143	0.6680	41
46	0.0332	0.0070	0.3875	79
47	**0.0532**	**0.0103**	**0.2028**	**67**
48	0.1549	0.0160	0.8118	76
49	0.1937	0.0110	0.8555	55
50	**0.8241**	**0.0416**	**15.5251**	**05**
51	0.1377	0.0437	1.5476	17
52	0.0627	0.0193	1.4266	96

VI. CONCLUSION

The current study predicted sudden cardiac death at a very early stage, one hour before it occurred, which may help to save the lives of patients. To the best of our knowledge, no other research team has yet predicted abrupt cardiac death at such a young stage as one hour prior. The study's selection of ECG signals for simulation, the materials and methods used, and the extraction of various heart rate variability aspects are all in accordance with the Task Force of European Societies for Heart Rate Variability's established standards for heart rate variability. The data observed for SCD participants differ significantly from those of normal subjects, confirming the risk factor for a SCD occurrence. By using one minute of ECG signals for SCD prediction, the limitations of other studies are also discussed. One minute ECG signals from SCD participants exhibit varying QRS peaks. As a result, we have convincingly demonstrated the impact of falsely detecting QRS peaks on BPM for all SCD database records.

REFERENCES

[1] Ciprian Rezu¸s, Victor Dan Moga, Anca Ouatu, and Mariana Floria, “QT interval variations and mortality risk: Is there any relationship?,”Anatol. J. Cardiol., vol. 15, pp. 255–258, 2015.
[2] H. J .J. Wellens, Peter J. Schwartz, Fred W. Lindemans, Alfred E. Buxton, Jaffrey J. Goldberger, Stefan H. Hohnloser, Heikki V. Huikari, Stefan Kaab, Maria Teresa, La Rovere, Marek Malik, Robert J. Myerburg, Maarten L. Simoons, Karl Swedberg, Jan Tijssen, Adrrian A. Voors, and Arthur A. Wilde, “Risk stratification for sudden cardiac death: current status and challenges for the future,” European Heart Journal, vol. 35, pp. 1642–1651, 2014.
[3] Sabine M. J. M. Straus, Jan A. Kors, Marie L. De. Bruin, Cornelis S. van der Hooft, Albert Hofman, Jan Heeringa, Jaap W. Deckers, J. Herrre Kingma, Miriam C. J. M. Sturkenboom, Bruno H. Ch. Stricker, and Jacquiline C. M. Witteman, “Prolonged QTc interval and risk of

sudden cardiac death in a population of older adults," Journal of the American College of Cardiology, vol. 47, no. 2, pp. 362–367, 2006.
[4] T. Nishibe, K. Sato, K. Yoshino, R. Seki, K. Yana, and T. Ono, "RR-QT interval trend covariabilty for sudden cardiac death risk stratification," in Engineering in Medicine and Biology Society (EMBC), Annual International Conference of the IEEE San Diego, CA, 2012, pp. 4287–4290, Aug. 28, 2012–Sept. 1.
[5] Timo H. Makikallio, Heikki V. Huikuri, Anne Makikallio, Leif B. Sourander, Raul D. Mitrani, Agustin Castellanos, and Robert J. Myerburg, "Prediction of sudden cardiac death by fractal analysis of heart rate variability in elderly subjects," Journal of the American College of Cardiology, vol. 37, no. 5, 2001.
[6] http://emedicine.medscape.com/article/151907-overview, April, 29, 2014.
[7] Bianca S. Honnekeri, Disha Lokhandwala, Gopi Krishna Panicker, and Yash Lokhandwala, "Sudden cardiac death in India: A growing concern," Journal of the association of Physicians of India, vol. 62, pp. 36–40, Dec. 2014.
[8] R. J. Myerburg, M. J. Junttila, "Sudden cardiac death caused by coronary heart disease," Circulation, vol. 125, pp. 1043–1052, 2012.
[9] Yash Lokhandwala, Gopi Krishna Panicker, and Shantanu Deshpande, "Sudden cardiac death-an Indian perspective," CVD Prevention and Control, vol. 4, pp. 103–108, 2009.
[10] B. H. Rao, "Global burden of sudden cardiac death and insights from India," Indian Heart Journal, vol. 66, pp. S18-S23, 2014.
[11] E. Ebrahimzadeh, M. Pooyan, and A. Bijar, "A novel approach to predict sudden cardiac death (SCD) using nonlinear and time-frequency analysis from HRV signals," PLoS ONE, vol. 9, no. 2, 2014.
[12] H. Li, S. Kwong, L. Yang, D. Huang, and D. Xiao, "Hilbert-huang transform for analysis of heart rate variability in cardiac health," IEEE/ACM Transactions on Computational Biology and Bioinformatics, vol. 8, no. 6, pp. 1557–1567, 2011.
[13] Vahid Houshyarifar and Mehdi Chehel Amirani, "An approach to predict sudden cardiac death (SCD) using time domain and bispectrum features from HRV signal," Bio-Medical Materials and Engineering, vol. 27, no. 2–3, pp. 275–285, 2016.
[14] Veena N. Hegde, Ravishankar Deekshit, and P. S. Satyanarayan, "Heart rate variability feature extraction using continuous wavelet transform," Journal of Medical Imaging and Health Informatics, vol. 5, no. 5, pp. 898–906, 2015.
[15] Jose Eduardo Saubhia Natali, Paulo Nogueira Starzynsky, and Jose Guilherme Chaui-Berlinck, "Oscillatory patterns in heart rate variability and complexity: A meta-analysis," Biomedical Signal Processing and Control, vol. 33, pp. 66–71, 2017.
[16] "Task Force of the European Society of Cardiology and the North American Society of Pacing and Electrophysiology: Heart rate variability: Standards of measurement, physiological interpretation, and clinical use," European Heart Journal, vol. 17, pp. 354–381, 1996.
[17] Jiapu Pan and Willis J. Tompkins, "A real-time QRS detection algorithm," IEEE Transactions on Bio-medical Engineering, vol., no. 3, 1985.
[18] https://physionet.org.