Prediction of Ventricular Fibrillation Using Support Vector Machine

Wei Wei Heng[1], Eileen Su Lee Ming[1], Ahmad Nizar B Jamaluddin[2], Fauzan Khairi Che Harun[3], Nurul Ashikin Abdul-Kadir[1] and Che Fai Yeong[4]
[1,3,4] Faculty of Engineering, Universiti Teknologi Malaysia, Johor, MALAYSIA.
[2] Department of Cardiology, Ara Damansara Medical Centre, Selangor, MALAYSIA.

E-mail: vivihengww@gmail.com

Abstract. Sudden cardiac death (SCD) remains one of the top causes of high mortality rate. Early prediction of ventricular fibrillation (VF), and hence SCD, can improve the survival chance of a patient by enabling earlier treatment. Heart rate variability analysis (HRV) has been widely adopted by the researchers in VF prediction. Different combinations of features from multiple domains were explored but the spectral analysis was performed without the required preprocessing or on a shorter segment as opposed to the standards of The European and North American Task force on HRV. Thus, our study aimed to develop a robust prediction algorithm by including only time domain and nonlinear features while maintaining the prediction resolution of one minute. Nine time domain features and seven nonlinear features were extracted and classified using support vector machine (SVM) of different kernels. High accuracy of 94.7% and sensitivity of 100% were achieved using extraction of only two HRV features and Gaussian kernel SVM without complicated preprocessing of HRV signals. This algorithm with high accuracy and low computational burden is beneficial for embedded system and real-time application which could help alert the individuals sooner and hence improving patient survival chance.

1. Introduction

According to World Health Organization, cardiovascular disease causes 17.7 million deaths annually [1]. Sudden cardiac death (SCD) accounts for half of the death caused by cardiovascular disease. Ventricular fibrillation (VF) is postulated to be the common underlying condition that leads to SCD. Prediction of VF, and hence SCD, can provide physicians more time for treatment preparation or alert the patient to take immediate action, for example calling the paramedic.

Numerous researchers have worked on this prediction problem using electrocardiogram (ECG) signal and heart rate variability (HRV) analysis. Previous works combined time-domain, frequency-domain, time-frequency-domain and nonlinear features in different ways [2]–[11]. The time domain methods, which were used by recent research [6] to explore prediction time up to three hours, were claimed to provide inadequate representation of data [12]–[14]. The spectral analysis required ectopic removal or replacement as an essential HRV signal preprocessing [15] and proposed to be used on at least 2-5 minutes of HRV signals [16], of which some work [8], [10], [11] fulfilled but some did not [3], [5], [9]. The nonlinear features were expected to provide more intrinsic information due to the non-stationary characteristic of HRV signal and were extensively explored, such as Poincare plot [3], [8], [9], [17], detrended fractal analysis [3], [4], [8], [9], entropy measure [7], [8], [18], recurrence plot [17], [18] and so on.

Historically, the VF prediction problem was transformed into classification problem at a certain time, with most works ([4], [7], [9], [19]) evaluated algorithms on signals at the fourth minute before VF onset (Figure 1). Additionally, considering the questionable validity of spectral analysis on short term recordings, this work focused on extracting time domain and nonlinear HRV features, to maintain the prediction resolution of one minute and validity of the algorithm. The features were then classified using support vector machine (SVM) of different kernels for VF prediction at the fourth minute prior to its onset. Our study aims to develop a robust automated VF prediction algorithm with high accuracy and low computational complexity.

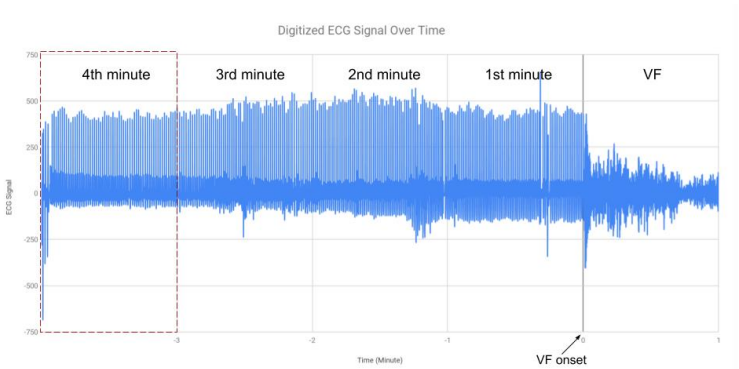

Figure 1. Most previous works analysed signals at the fourth minute before VF onset.

2. Materials and Methods

The workflow of our current study is shown as block diagram in Figure 2.

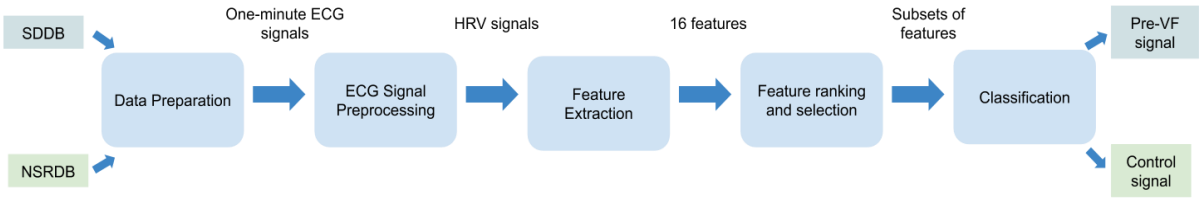

Figure 2. Workflow of our current study.

This study acquired pre-VF signals from Sudden Cardiac Death Holter Database (SDDB) [20], [21] and control signals from MIT-BIH Normal Sinus Rhythm Database (NSRDB) [21]. SDDB comprised 23 24-hours ECG records from subjects who experienced sustained ventricular arrhythmia and our study excluded three SDDB records without annotation of VF onset time (‘40’, ‘42’, ‘49’). NSRDB comprised 18 24-hours ECG records from subjects without significant arrhythmia and all NSRDB records were included in analysis. From both SDDB and NSRDB, only channel 1 of ECG signal per subject was utilized in our study to ensure the independence of signal samples and avoid the reporting of over-optimistic performance. Since HRV signal instead of the original ECG signal was analysed, any channel of data which provides accurate R peaks could be chosen. The first ten seconds of the signals from ‘nsrdb/16272’ and ‘sddb/31’ are juxtaposed in Figure 3, as visual inspection. VF onset was annotated in SDDB and two-hour mark was consistently used as a time mark for NSRDB instead of randomly chosen signal segment that inhibited test result reproducibility. One-minute ECG signal at the fourth minute before VF onset or before the 2-hour mark from normal sinus rhythm was extracted. In total, 38 one-minute ECG signals were prepared and subjected to the ECG preprocessing stage.

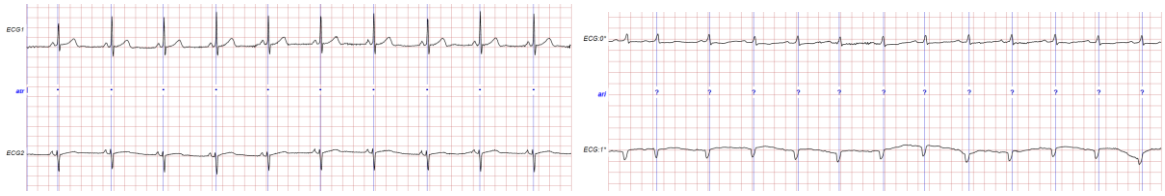

Figure 3. (a) 10-seconds signal of 'nsrdb/16272' on the left. (b) 10-seconds signal of 'sddb/31' on the right.

The ECG signals were filtered using fourth-order Butterworth filter with passband ranging from 1 Hz to 30 Hz to remove high-frequency measurement noise and baseline wandering. Hamilton-Tompkins algorithm [22], which was an enhanced version of Pan-Tompkins algorithm, was used to detect R peaks in the filtered signals and to extract HRV signal. The time-domain and nonlinear features, as in Table 1, were extracted from the extracted HRV signal. All the features were implemented in Python with the help of library [23] and [24].

Table 1. The extracted time domain and nonlinear features.

Time Domain [23]	Nonlinear
• Number, mean, standard deviation, coefficient of variation of RR intervals • Standard deviation, root mean square of successive differences (RMSSD) of RR intervals • Percentage of adjacent RR intervals that differ by more than 50 ms (pNN50) • HRV triangular index • Width of triangular interpolation of RR intervals histogram	• Poincare plot - SD1, SD2, SD ratio [23] • Detrended fluctuation analysis - short (alpha1) and long (alpha2) term scaling exponents [23] • Sample entropy [23] • Hurst exponent [24]

The statistically significant features were selected using two-sample T-test and ranked by t-value. The highest-ranked features (with the highest t-values), in time domain and nonlinear category, were fed into SVM for classification, and forward feature selection, one feature at a time, were adopted to investigate the performance difference and to search for the minimum number of features needed to acquire the highest performance. In the forward feature selection process, only statistically significant features with p-value less than 0.05 were selected to be combined with the previously selected features. MATLAB Classification Learner tool was first used to experiment different SVM kernels, including linear, quadratic, cubic and gaussian (radial basis function). The gamma parameter of Gaussian kernel SVM was determined based on the number of features being used while the box constraint parameter was held as 1. Then, the performance of the selected features and classifier parameters were evaluated using 10-fold cross validation scheme. The reported accuracy, sensitivity and selectivity were averaged over 10-fold validations with 100 iterations for each fold.

3. Main Results

At the feature selection stage, alpha2 and sample entropy were discarded because they contained unknown or infinite value while the number of RR intervals, mean of RR intervals and triangular interpolation width were excluded by two-sample T-test because the p-value exceeded 0.05 as alpha value. The highest-rank features were Hurst exponent (t = 6.28) in nonlinear category and pnn50 (t = 3.81) in time domain. Combination of Hurst exponent with Gaussian SVM yielded 89.8%, 95.0% and 84.1% of accuracy, sensitivity and specificity respectively, while pnn50 with linear SVM yielded 72.2%, 55% and 91.28%. Then the features were appended one at each step to investigate the performance. Only the best-performing combination of feature subset and SVM model parameters in each feature category is presented in Table 2. The highest performance of 94.7% accuracy, 100% sensitivity and 88.9% specificity was achieved using only two nonlinear features (SD1 and Hurst exponent) and Gaussian SVM (gamma = 1.4). This feature subset attained the best performance among all two-features classifications and hence integrated with more features to explore further performance improvement but none occurred. As shown by both feature ranking and performance evaluation, combination of different nonlinear features were proven to be superior to time domain features for the 4th-minute prediction, confirming the previous studies findings [4], [7].

Table 2. The performance metrics with their standard deviation (sd) of different feature subsets and SVM kernel parameters averaged over 100 iterations of 10-fold cross validation.

Feature Category	Features	SVM Kernel Type (Gamma)	Accuracy (sd), %	Sensitivity (sd), %	Specificity (sd), %
1 Nonlinear Feature	Hurst exponent	Gaussian (0.25)	89.8 (1.2)	95.0 (0.0)	84.1 (2.6)
1 Time Feature	pNN50	Linear	72.2 (1.5)	55.0 (0.0)	91.3 (3.2)
2 Nonlinear Features	Hurst exponent, SD2	Gaussian (1.4)	94.7 (0.0)	100.0 (0.0)	88.9 (0.0)
2 Time Features	pNN50, RMSSD	Gaussian (1.4)	84.3 (0.8)	75.1 (1.4)	94.4 (0.0)
1 Nonlinear + 1 Time Features	Hurst exponent, pNN50	Gaussian (1.4)	92.7 (2.0)	99.0 (2.1)	85.6 (3.5)
3 Features	Hurst exponent, SD2, pNN50	Gaussian (1.7)	94.66 (0.45)	99.85 (0.86)	88.89 (0.00)
4 Features	Hurst exponent, SD2, pNN50, RMSSD	Gaussian (0.5)	94.37 (1.40)	100.00 (0.00)	88.11 (2.96)

Our prediction result was benchmarked against previous works on 4th-minute prediction (Table 3). Our study achieved similar prediction performance as the other works while using all annotated records instead of eliminating more records as done by [3], [4] and using only single channel data from each patient instead of both [5]. We achieved the same performance as [7] using two instead of three features and our method did not require digital wavelet transform before extraction of nonlinear features. Besides, our study achieved 100% sensitivity, which is zero false negative rate which was more important than the high specificity and was critical to prevent the potential SCD.

Table 3. Performance benchmark of our study to previous works on 4th-minute prediction.

Author (Year)	Records	Features	Classifier	Accuracy ,%	Sensitivity ,%	Specificity ,%
Fujita et al (2016) [7]	SDDB (20), NSRDB (18)	Complexity, mobility and energy of first level digital wavelet transform coefficient	Gaussian SVM	94.7	95	94.4
Houshyar ifar et al (2016) [5]	SDDB (20), NSRDB (36)	1 LDA components derived from 6 bispectrum and 2 time-domain features	Gaussian SVM	94.5	-	-
Urda et al (2017) [4]	SDDB (16), NSRDB (18)	Multifractal h-fluctuation index, Hurst exponent	Sigmoid SVM	97.1	100	94.4
Ebrahimz adeh et al (2018) [3]	SDDB (35), NSRDB (35)	Features selected from 27 features	Multilayer Perceptron	95.2	93.8	94.6
Our Study	**SDDB (20), NSRDB (18)**	**SD1, Hurst exponent**	**Gaussian SVM**	**94.7**	**100**	**88.9**

4. Conclusion

Our prediction algorithm requires only extraction of two nonlinear HRV features and Gaussian SVM without complicated pre-processing of HRV signals while achieving high accuracy of 94.7% and sensitivity of 100%. Together with the high performance, this algorithm with lower computational burden is beneficial for embedded system and real-time application. High prediction accuracy and low false positive rate at a short time frame before VF occurrences, four minutes in this case, is important in hospital to alert the staff and physicians for VF treatment preparation and hence reducing in-hospital SCD. Besides, future research could also explore earlier prediction time frame that enables more patients to reach hospital in time before the VF onset and hence reducing out-of-hospital SCD.

Acknowledgments

This work was supported by CREST R&D grant T20C3-14 (4B244), UTM Matching Grant No. Q.J13000.3001.01M13 and UTM Grant No. Q.J130000.2523.16H47.

References

[1] WHO, "WHO | Noncommunicable diseases," *Who*, 2017. [Online]. Available: http://www.who.int/mediacentre/factsheets/fs355/en/. [Accessed: 25-Oct-2017].
[2] 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," *J. Med. Syst.*, vol. 42, no. 10, p. 176, Oct. 2018.
[3] 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, Jul. 2018.
[4] R. Urda Benitez and A. O. Duque, "Sudden cardiac death prediction based on a nonlinear estimation," in *IFMBE Proceedings*, vol. 60, Springer, Singapore, 2017, pp. 485–488.
[5] V. Houshyarifar and M. Chehel Amirani, "An approach to predict Sudden Cardiac Death (SCD) using time domain and bispectrum features from HRV signal," *Biomed. Mater. Eng.*, vol. 27, no. 2–3, pp. 275–285, Aug. 2016.
[6] A. Raka, G. Naik, and R. Chai, "Computational Algorithms Underlying the Time-Based Detection of Sudden Cardiac Arrest via Electrocardiographic Markers," *Appl. Sci.*, vol. 7, no. 9, p. 954, 2017.
[7] H. Fujita *et al.*, "Sudden cardiac death (SCD) prediction based on nonlinear heart rate variability features and SCD index," *Appl. Soft Comput. J.*, vol. 43, pp. 510–519, Jun. 2016.
[8] L. Murukesan, M. Murugappan, M. Iqbal, and K. Saravanan, "Machine Learning Approach for Sudden Cardiac Arrest Prediction Based on Optimal Heart Rate Variability Features," *J. Med. Imaging Heal. Informatics*, vol. 4, no. 4, pp. 521–532, 2014.
[9] E. Ebrahimzadeh, M. Pooyan, and A. Bijar, "A novel approach to predict sudden cardiac death (SCD) using nonlinear and time-frequency analyses from HRV signals," *PLoS One*, vol. 9, no. 2, p. e81896, Feb. 2014.
[10] E. Ebrahimzadeh and M. Pooyan, "Early detection of sudden cardiac death by using classical linear techniques and time-frequency methods on electrocardiogram signals," *J. Biomed. Sci. Eng.*, vol. 4, no. 11, pp. 699–706, 2011.
[11] K. H. Boon, M. Khalil-Hani, M. B. Malarvili, and C. W. Sia, "Paroxysmal atrial fibrillation prediction method with shorter HRV sequences," *Comput. Methods Programs Biomed.*, vol. 134, pp. 187–196, Oct. 2016.
[12] A. Voss *et al.*, "The application of methods of non-linear dynamics for the improved and predictive recognition of patients threatened by sudden cardiac death," *Cardiovasc. Res.*, vol. 31, no. 3, pp. 419–433, Mar. 1996.
[13] F. Lombardi, T. H. Mäkikallio, R. J. Myerburg, and H. V. Huikuri, "Sudden cardiac death: Role of heart rate variability to identify patients at risk," *Cardiovasc. Res.*, vol. 50, no. 2, pp. 210–217, May 2001.
[14] U. Meyerfeldt *et al.*, "Heart rate variability before the onset of ventricular tachycardia: Differences between slow and fast arrhythmias," *Int. J. Cardiol.*, vol. 84, no. 2–3, pp. 141–151, Aug. 2002.
[15] G. D. Clifford and L. Tarassenko, "Quantifying errors in spectral estimates of HRV due to beat replacement and resampling," *IEEE Trans. Biomed. Eng.*, vol. 52, no. 4, pp. 630–638, Apr. 2005.
[16] H. F. Jelinek, D. J. Cornforth, and A. H. Khandoker, "Heart rate variability: Standards of measurement, physiological interpretation, and clinical use," *Eur. Heart J.*, vol. 17, pp. 354–381, 1996.
[17] V. Houshyarifar and M. C. Amirani, "Early detection of sudden cardiac death using Poincaré plots and recurrence plot-based features from HRV signals," *Turkish J. Electr. Eng. Comput. Sci.*, vol. 25, no. 2, pp. 1541–1553, 2017.
[18] M. Khazaei, K. Raeisi, A. Goshvarpour, and M. Ahmadzadeh, "Early detection of sudden cardiac death using nonlinear analysis of heart rate variability," *Biocybern. Biomed. Eng.*,

vol. 38, no. 4, pp. 931–940, Jan. 2018.

[19] U. R. Acharya, H. Fujita, V. K. Sudarshan, D. N. Ghista, W. J. E. Lim, and J. E. Koh, "Automated Prediction of Sudden Cardiac Death Risk Using Kolmogorov Complexity and Recurrence Quantification Analysis Features Extracted from HRV Signals," in *Proceedings - 2015 IEEE International Conference on Systems, Man, and Cybernetics, SMC 2015*, 2016, pp. 1110–1115.

[20] S. D. Greenwald, "The Development and Analysis of a Ventricular Fibrillation Detector," pp. 1–241, 1986.

[21] A. L. Goldberger *et al.*, "PhysioBank, PhysioToolkit, and PhysioNet: components of a new research resource for complex physiologic signals.," *Circulation*, vol. 101, no. 23, pp. E215-20, 2000.

[22] P. S. Hamilton and W. J. Tompkins, "Quantitative Investigation of QRS Detection Rules Using the MIT/BIH Arrhythmia Database," *IEEE Trans. Biomed. Eng.*, vol. BME-33, no. 12, pp. 1157–1165, Dec. 1986.

[23] P. Gomes, H. Silva, and P. Margaritoff, "pyHRV - Open-Source Python Toolbox for Heart Rate Variability," 2018. [Online]. Available: https://github.com/PGomes92/pyhrv. [Accessed: 09-Oct-2019].

[24] Christopher Schölzel, "nolds · PyPI." [Online]. Available: https://pypi.org/project/nolds/. [Accessed: 08-Oct-2019].