TELKOMNIKA, Vol.14, No.3, September 2016, pp. 999~1008
ISSN: 1693-6930, accredited A by DIKTI, Decree No: 58/DIKTI/Kep/2013
DOI: 10.12928/TELKOMNIKA.v14i3.3665

Ventricular Tachyarrhythmia Onset Prediction Based on HRV and Genetic Algorithm

K. H. Boon*[1], MB Malarvili[2], M. Khalil-Hani[3]
Faculty of Electrical Engineering, Universiti Tekonologi Malaysia,
81310, Skudai, Johor, Malaysia.
*Corresponding author, e-mail: boon.khang.hua@gmail.com[1], malarvili@biomedical.utm.my[2], khalil@fke.utm.my[3]

Abstract

Predicting onset of ventricular tachyarrhythmia provides opportunities to reduce casualties due to sudden cardiac death. However, the prediction accuracy still needs improvement. Therefore, we aim to propose a method that can predict the onset of tachyarrhythmia events with improved accuracy based on heart rate variability and Support Vector Machine classifier. Fifty percent of sample data from standard database was used to train the classifier, and the remainder was used to verify the performance. Five minutes RR intervals immediately prior to tachyarrhythmia event from each sample data was cropped for ectopic beat correction and then converted to heart rate. Extraction of time domain, spectral, non-linear and bispectrum features were performed subsequently. Furthermore, genetic algorithm was used to simultaneously optimize the feature subset and classifier parameters. With the optimization, prediction accuracy of our proposed method able to outperform previous works with 77.94%, 80.88% and 79.41 % for sensitivity, specificity and accuracy respectively.

Keywords: *Heart Rate Variability, Arrhythmia Prediction, Ventricular Tachyarrhythmia (VTA), Genetic Algorithm, Bispectrum features.*

1. Introduction

Ventricular tachyarrhythmia (VTA) is a type of the arrhythmia (abnormal heart rhythms) that arises from improper electrical activity in the bottom chambers of the heart called ventricles. Occurrence of VTA is harmful to patient's health because it causes fainting, palpitations, asystole, and even triggers approximately 80% of sudden cardiac death (SCD) cases [1]. SCD accounts for one out of every two deaths from cardiovascular diseases [2]. Therefore, development of the reliable predictor of VTA onset, namely the ventricular tachycardia (VT) and ventricular fibrillation (VF), is clinical important because the prediction provides opportunity to timely prevention of the negative consequences brought by VTA through early termination by implantable cardioverter defibrillator (ICD) [3].

RR interval is the time interval between two consecutive heart beats while heart rate is the reciprocal of RR interval. Heart rate variability (HRV) signal is variation of heart rate that can be used to diagnose cardiovascular diseases. Generally, HRV signal can be obtained through measuring the consecutive R peaks of electrocardiogram (ECG) signal [4, 5]. Another more convenient method to obtain HRV is using the time difference between two consecutive pulses in the photoplethysmograph (PPG) signal [6]. HRV signal is usually analyzed with various HRV analysis based feature extraction techniques. These techniques have been widely applied in various medical related researches such as classification of cardia arrhythmia, diagnosis of neonatal sepsis, discrimination of sleep stage and etc., [7].

HRV analysis is also one of the popular methods that is actively being researched for application in VTA onset prediction. Initially, previous works have focused their study on statistical difference of the HRV features (features extracted based on different HRV analysis techniques) that extracted from the heart rate prior to VTA onset and the control data (HRV signal without VTA events) respectively. Many reports have found that there are statistical significant changes in HRV features values prior to VTAs [8-10]. Inspired by these findings, various VTA onset prediction methods based on HRV have been developed.

Received April 2, 2016; Revised June 10, 2016; Accepted June 29, 2016

However, the main issue is that the prediction results of the previous works are still unsatisfactory for predicting such life-threatening arrhythmia. Among the previous works, the highest achievable prediction accuracy was 75.6% [11]. Other previous works could not achieve that accuracy level. By using heart rate pattern, Thong and Raitt [12] achieved prediction performance with 53% of sensitivity and 91% of specificity. With decision rule based system based on multipole analysis, Rozen, et al., [13] achieved 50% of sensitivity and 91.6% of specificity. Wollman, et al., [14] proposed a method, which performed the VTA onset prediction based on time domain features and regression tree classifiers, obtained 70.9% of accuracy. Joo, et al., [11] proposed a prediction method that attained the most balanced and highest prediction performance among previous works with 77.3% of sensitivity, 73.8% of specificity and 75.6% of accuracy. Furthermore, their method used significantly shorter HRV signal (5 minutes only) to achieve higher prediction accuracy. In contrast, other previous works [12-14] utilized more than 10 minutes of HRV signal in length prior to VTA event for prediction. The main reasons for such higher prediction performance were utilization of more comprehensive types of HRV features than other previous works and employment of more advanced supervised classifier - artificial neural network (ANN). In their work, multiple categories of HRV features based on time domain analysis, spectral analysis based on fast Fourier transform (FFT), and poincare plot were employed to train the model of the ANN.

To address the aforementioned issue, the main objective of this paper is to propose a VTA onset prediction method with improved prediction accuracy. Contributions of our work can be summarized as follows. Inspired by Joo, et al., [11], this paper also employed multiple types of HRV features for prediction. However, additional type of HRV features such as Triangular Interpolation of RR interval histogram, sample entropy and higher order spectral analysis, which have not been used in [11], are also used to train Support vector machine (SVM) classifier for prediction in our work. Furthermore, the genetic algorithm (GA) based feature selection process proposed by Huang and Wang [15] is adopted to simultaneously optimize the HRV feature subset and SVM classifier parameters. With the additional types of HRV features and GA based optimization process, prediction performance of our method outperforms all previous works with 77.94%, 80.88% and 79.41 % for sensitivity, specificity and accuracy respectively even though we use stricter approach to evaluate our method. Prediction sensitivity and specificity of our method are more balanced when compared to previous works [12-14]. Furthermore, in contrast to most of previous works [12-14] that used more than 10 minutes of HRV signal, our method only uses 5 minutes HRV signal, which end immediately prior to VTA onset, for feature extraction. Finally, optimal HRV feature subset selected by GA for prediction is also reported for reference in future works.

The layout of the paper is as follows. Section 2 presents the database and proposed method. Section 3 presents the result and analysis. Section 4 presents the conclusion.

2. Proposed VTA onset Prediction Method

Block diagram in Figure 1 shows the overview of the proposed method. It comprises of pre-processing stage, HRV feature extraction stage, HRV feature selection stage and support vector machine (SVM) based classification stage.

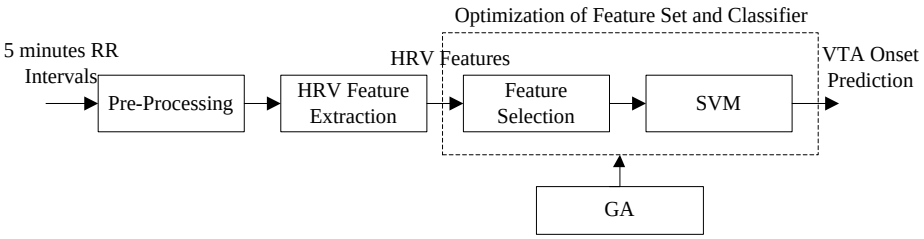

Figure 1. Overview of Proposed Method

Firstly, 5 minutes RR intervals that end immediately prior to VTA onset for 106 pre-VT/VF data and 5 minutes RR intervals of control data from the database are fed to the pre-processing

stage respectively. Pre-processing stage includes ectopic beat correction, resampling of RR interval and conversion to HRV sequences by computing the reciprocal of RR interval. In HRV feature extraction stage, 5 time-domain, 6 frequency-domain, 5 non-linear and 37 bispectrum features are then extracted from quantified HRV. During the feature selection stage, genetic algorithm (GA) is used to optimize the feature set. Finally, prediction performance of the proposed method is evaluated in term of sensitivity, specificity and accuracy. Detail of each block is described in the remaining sub-sections.

2.1. Database

RR intervals recording are obtained from the Spontaneous Ventricular Tachyarrhythmia Database Version 1.0 from Medtronic, Inc. from the Physionet [16]. The database was collected from records of 78 patients with ICDs (63 males and 15 females, aged from 20.7 to 75.3) and consisted of the following RR intervals: 106 pre-VT records, 29 pre-VF records, and 135 control data sets. Each data included 1024 RR intervals (corresponding to around 15 minutes). Short-term HRV analysis is performed on the 5 minutes RR intervals prior to each VTA(VT/VF) event.

2.2. Preprocessing

RR intervals are obtained from the database. HRV is then formed by computing the reciprocal of the intervals between successive R peaks. After that, the signal is resampled to 4 Hz signal by using the cubic spline interpolation as it has been reported that this technique is better than linear interpolation [17].

Before the HRV is resampled, it is evaluated and corrected based on McNames's algorithm [18]. This algorithm detects the abnormal heart rate signal in the series which may be caused by ectopic beat, artifact noise or miss peak detection. The algorithm evaluates heart rate using test statistic, D(n) calculated with following equation:

$$D(n) = \frac{|HR(n) - HR_m|}{1.483 med\{|HR(n) - HR_m|\}} \quad (1)$$

Where HR_m is the average heart rate of the series, HR(n) is instantaneous heart rate and med{ } is median filter. If D(n) value excess certain threshold, τ, the instantaneous heart rate is considered as abnormal heart rate. The heart rate is corrected with following equation:

$$\widehat{HR}(n) = med\left\{x(n+m): |m| < \frac{w_m - 1}{2}\right\} \quad (2)$$

Where w_m is the window length of the medium filter. In this work, τ and w_m are set to 4 and 11 respectively according to the literature review [18].

Figure 2 shows 5 minutes RR intervals of a patient prior to VTA event. RR intervals before and after correction by McNames's algorithm are shown.

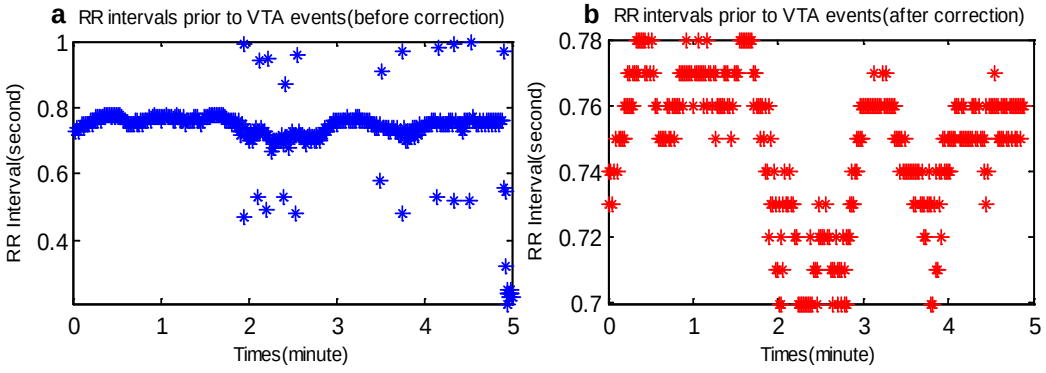

Figure 2. Five minutes RR intervals prior to VTA event before and after ectopic beat correction

2.3. HRV feature extraction

In this study, a number of 53 HRV features are extracted from HRV using time-domain, frequency domain and non-linear analysis. Each of features and its abbreviations are explained in the following sub-sections. All of the mentioned features are well known and have been used in other related HRV studies [19].

2.3.1. Time Domain Features

Six time-domain HRV features are computed by using statistical analysis. They are the mean of HRV (MeanRR), the standard deviation of HRV (SDRR), root mean square of successive difference intervals (RMSSD), number of adjacent RR intervals differing by more than 50 ms (NN50), and sum of NN50 divided by the total number of all RR intervals (pNN50). Besides, HRV triangular index (HRVTri) [20] was also employed to extract geometrical HRV feature. HRVTri is the total number NN intervals divided by number of RR intervals that fall to modal bin.

2.3.2 Spectral Features

For spectral analysis, power spectral density (PSD) was estimated from HRV signal. It is generally accepted that spectral power in low frequency (LF) band (0.04-0.15) and high frequency (HF) band (0.15-0.4 Hz) reflect the the sympathetic and parasympathetic activities of the Autonomic Nervous System (ANS) respectively [20]**.** In this study, fast Fourier transforms (FFT) [21] and auto-regressive (AR) model [22] were used to estimate the power spectrum from HF and LF bands. Ratio of LF to HF (LF/HF) band is also calculated. Coefficients of AR model were estimated with burg method [23] and the order was set to 16 which is optimal for HRV analysis [22].

Both FFT and AR are popular tools for spectral analysis. However, each of them has their strengths and weaknesses. The advantage of FFT is it is a non-parametric tool that does not assume the data is uniformly distributed with certain variance while AR model assumes the data is uniformly distributed with fixed variance value. Therefore, FFT does not suffer from poor performance when the property of data does not fit assumption. However, FFT suffer from spectral leakage effect when compared to AR model. Besides, AR model can provide better frequency resolution in power spectrum and perform better when it is applied to short time series data [24]. Therefore, this study employed both techniques for spectral analysis to take advantage of their strengths.

2.3.3. Bispectrum Features

PSD of spectral analysis does not provide the phase relations between frequency components. However, Higher Order Spectra (HOS) [25] can be used to analyze the non-linear signal which may involve the cross phase relations [26]. Since the HRV signal is non-linear and non-Gaussian in nature, bispectrum also can reveal the information that is not contained in power spectrum]. Besides, these features also can be employed to detect quadratic phase coupled harmonics arising from nonlinearities of the HRV signal. The bispectrum $\mathrm{B}(f_1, f_2)$ of a non-Gaussian signal, $x(t)$, is a two-dimensional Fourier transform of the third order cumulant $\mathrm{C}(\mathrm{m}, \mathrm{n})$ defined as:

$$\mathrm{C}(\mathrm{m}, \mathrm{n}) = E[x(k)x(k+m)x(k+n)] \quad (3)$$

$$\mathrm{B}(f_1, f_2) = E[X(f_1)X^*(f_2)X(f_1+f_2)] \quad (4)$$

Where E is expectation function, $X(f)$ is Fourier transform of $x(t)$ and $X^*(f)$ is complex conjugate.

In this study, Bispectrum was estimated based on the direct method described in [25]. The 4 Hz cubic spline interpolated HRV signal was divided into several segments with each segment is consisted of 512 data points. Then bispectrum was then computed from Fourier transform of each segment. After that, bispectrum features were extracted from different regions of the two-dimensional bipsectrum.

Bispectrum of HRV signal can be divided into 3 subband regions inside region of interest (ROI) [27]. They are LF–LF (LL), LF–HF (LH), and HF–HF (HH) region which cover

different ranges of frequencies as shown in Figure 3. Formulas in [27, 28] were employed to compute bispectrum features from each subband region and the ROI. These features include mean magnitude (M_{ave}), normalized bispectral entropy (P1), normalized bispectral squared entropy (P2), sum of logarithmic amplitudes of the bispectrum (H1), sum of logarithmic amplitudes of diagonal elements in bispectrum (H2), first-order spectral moment of the amplitudes of diagonal elements in the bipsectrum (H3), Second-order spectral moment of the amplitudes of diagonal elements in the bispectrum (H4), weighted center of the bipsectrum, WCOB (f_{1m}, f_{2m}). For LH region, H2, H3 and H4 are excluded because the diagonal elements are not existed.

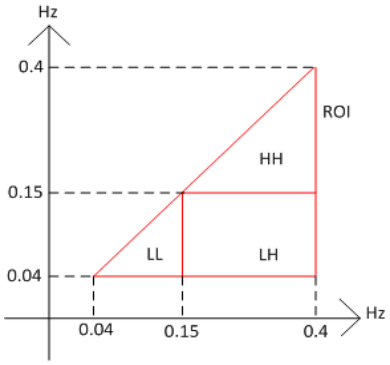

Figure 3. Sub-band Regions (LL, LH and HH) and Region of Interest (ROI)

2.3.4. Nonlinear dynamics features

Generally, non-linear analysis is recognized to be able to describe the biological process in more effective way. Review in [19] has shown that various non-linear techniques have been extended to study various cardiac arrhythmias. In this study, Poincare plot and sample entropy (SampEn) are employed to extract features from the HRV.

Poincare plot is drawn by plotting each RR interval against next RR interval. Each RR interval is the timing difference between successive R peaks of HRV signal. An ellipse is then fitted to the shape of Poincare plot. The computation of the width (SD1) and length (SD2) of the eclipse can be simplified by employing formulas below. Standard deviation denoted SD1 is related to fast beat-to-eat variability in data while SD2 describes the long term variability of RR. The ratio of SD1/SD2 is also computed to describe the relation between two components.

$$SD1 == \frac{1}{2} * SDSD^2 \quad (5)$$

$$SD2 = 2 * SDRR^2 - \frac{1}{2} * SDSD^2 \quad (6)$$

Sample entropy (SampEn) is a statistic measure that quantifies the regularity of times series data. The method proposed in [29] was used to compute the SampEn of the HRV signal. Parameters of sample entropy are set as follows: Embedding dimension, m is set to 2 and tolerance distance, r is set to 20% of the standard deviation of HRV sequences according to recommendation by Pincus and Goldberger [29]. Sample entropy is defined as:

$$SampEn(m,r) = \lim_{N\to\infty}(-\frac{\frac{1}{N-m}\sum_{i=1}^{N-m} A_i^m}{\frac{1}{N-m}\sum_{i=1}^{N-m} B_i^m}) \quad (7)$$

Where, $\theta(.)$ is Heaviside function and,

$$A_i^m = \frac{1}{N-m}\sum_{j=1,j\neq i}^{N-m}\theta(r - ||x_{m+1}(i) - x_{m+1}(j)||) \quad (8)$$

$$B_i^m = \frac{1}{N-m}\sum_{j=1,j\neq i}^{N-m}\theta(r - ||x_m(i) - x_m(j)||) \quad (9)$$

2.4. Classification: Support vector machine (SVM)

In this study, SVM is used as supervised classifier to classify the HRV sequences to either "normal episode" (control data) or "abnormal episode" (HRV sequences prior to VTA event). Input of the SVM is the HRV features that are extracted during the feature extraction stage. SVM is chosen because similar HRV based related HRV analysis based works [30] as reported good classification performance with this classifier.

SVM is supervised classifier based on statistical learning theory [31]. SVM maps the training samples from the input space to higher-dimensional features space via a kernel function. In this study, the radial basis function (RBF) is used as the kernel function. Parameters of kernel-kernel width γ and regularization constant C – are set according to feature selection method in section 2.5.

2.5. Genetic Algorithm Based Feature Selection

The purpose of feature selection is to select optimal subset of features from original feature set without transforms original feature. Using all the extracted features does not always give the best classification performance [32]. Feature selection process can offer two benefits: enhance classification performance and reduce number of features required for classification model. Furthermore, it also can help us to understand which features are important for classification.

In our work, feature selection process based on Genetic Algorithm (GA) proposed by Huang and Wang [15] is adopted to simultaneously optimize the HRV feature subset and SVM parameters (C and γ). Initially, GA produces an initial population with size of N chromosomes. Each chromosome is represented by fixed length binary string. Binary string can be divided into 3 segments. First segment is 53 bits binary string that represents a feature subset, such that "1" represents selection while ''0'' represents the deletion of the specific feature from the feature set. Second and third segment are 20 bits binary string that represents the encoded value of parameter C and parameter γ for SVM respectively. Binary string of second and third segment is decoded back to real value with equation 9:

$$p = min_p + \frac{max_p - min_p}{2^l - 1} \times d \quad (10)$$

Where p is real value of the binary string P, d is decimal of bit string P, max_p is maximum value of parameter, min_p is minimum value of parameter, l is length of bit string. In this paper, both $min_p$and max_p are set to 0.1 and 1000 respectively for both SVM parameters.

Fitness function in equation (11) is used to evaluate fitness of a chromosome.

$$fitness = W_A \times SVM_{Accuracy} + W_F\left(\sum_{i=1}^{n_f} C_i \times F_i\right)^{-1} \quad (11)$$

Where W_A is weight for SVM prediction accuracy, W_F is weight for selected features, $SVM_{Accuracy}$ is prediction accuracy of SVM, C_i is cost of ith feature, F_i represents the whether ith feature is selected or not. In this paper, W_A and W_F are set to 0.8 and 0.2 respectively according to experiment setting in Huang and Wang [15]. Besides, all C_i are set to "1" since the interest of feature selection in our work is to minimize feature count during the optimization (not to reduce feature computation cost or dollar cost).

Roulette Wheel Selection method is used as a selection strategy. Double point crossover operator and bit flip mutation are employed as genetic operators to explore the search space. Other parameters of GA are set as follows: Probability of crossover (Pc) = 0.7, Probability of mutation (Pm) =0.01, Stopping Generation (GN) = 5000 and population size (N) =60.

2.6. Performance Evaluation

In this study, SVM classifier is used to evaluate the prediction performance of selected HRV features in the feature selection process. Performance metrics such as sensitivity (SEN), specificity (SPE), and accuracy (ACC), which have been used in [33], are employed to measure the prediction performance of algorithm. Positive prediction means the algorithm classifies the RR intervals recording as prior to PAF event correctly while negative prediction means the

algorithm classifies the RR intervals recording from control data correctly. The sensitivity (SEN) is defined as the ratio of the number correct positive prediction to the total number of positive prediction. Specificity (SPE) is the ratio of the number of correct negative prediction to total number of negatives prediction. Accuracy (ACC) is the ratio of total number of correct prediction.

In current works, 50% of sample data are randomly selected as training set and remaining data are used as testing set. Furthermore, sample data for both training set and testing are subjected independent after they are randomly partitioned. Therefore, RR interval recordings from both training set and testing set are definitely come from different patients.

3. Result and Discussions

Table 1 shows the benchmarking results of our proposed method against previous works. The benchmarking aspects include the required HRV signal length for prediction, type of HRV feature extraction method, performance evaluation method and prediction performance.

In Table 1, the results have shown that the prediction performance of our method outperforms all previous works. With 5 minutes of HRV signal length prior to VTA onset, our method achieves 79.41% of accuracy, which is higher than the accuracy reported by Joo, et al., [11] (best previous work) and Wollman, et al., [14] respectively. Although the accuracy levels were not reported by Thong & Raitt [12] and Rozen, et al., [13], their prediction sensitivity and specificity were not balanced with low sensitivity rate (53 and 50%). It shows that their prediction methods achieved poor performance in predicting the VTA onset successfully. In contrast, our method can achieve acceptable and balanced prediction sensitivity and specificity with 77.94% and 80.88% respectively. Similar balanced prediction sensitivity and specificity are only reported by Joo, et al., [11].

Table 1. Benchmarking against previous works

Previous Work	HRV Signal Length (Minutes)	Feature Extraction Method	Performance Evaluation Method	SEN (%)	SPE (%)	ACC (%)
Thong & Raitt, 2007. [12]	1.8 Hours	Decision rule based on HR pattern.	Used all 208 data as both training and testing set.	53.0	91.0	-
Rozen et al., 2013 [13]	10-60	Decision rule based on Multipole analysis.	Used all 124 data as both training and testing set. (64 pre-VT/VF, 60 control data)	50.0	91.6	-
Wollman et al., 2015 [14]	20-40	Time domain features. (CART)	Used all 155 data as both training and testing set. (68 pre-VT/VF, 72 control data)	94.4	50.6	70.9
Joo et al., 2012. [11]	5	Time, Welch based FFT, Poincare. ANN	Partitioned the database into 175 training data and 86 testing data.	77.3 (34/44)	73.8 (31/42)	75.6 (65/86)
Our proposed Method	5	Time, FFT, AR, Poincare, Higher Order Spectral. SVM.	Partitioned the database into 134 training data and 135 testing data.	77.94 (53/68)	80.88 (55/68)	79.41 (108/136)

In term of performance evaluation method, direct comparison can be made between our method and the best prediction method proposed by Joo, et al., [11] because Table 1 shows that both of our works use independent training and testing data set to train and evaluate the proposed method. Other previous works [12-14] did not perform such partitioning. Instead, their training and testing set shared the same sample data. In our work, the approach to evaluate the prediction performance of the proposed method is stricter than in [11]. Firstly, higher number of testing sample data is used to measure the prediction accuracy. We have used 50% of sample data as training set while remaining 50% of data are used as testing set. In contrast, Joo, et al., [11] used 67% of data for training and only 32.9% of data for testing. Therefore, the number of sample data to test the trained supervised SVM classifier in our work is increased by 62% (26 samples). Secondly, sample data for both training and testing set in our work are subject

independent after data are randomly partitioned. Consequently, the RR interval recordings from both training and testing set are definitely obtained from different patients. This approach was not reported in [11] and they only stated the data are randomly divided into training and testing set. With stricter performance evaluation approach, our proposed method achieves higher prediction accuracy with 79.41% than method in [11] (75.6% accuracy).

Higher prediction performance of our method can be attributed to employment of more types of comprehensive HRV features. Our proposed method uses the HRV features extracted from sample entropy, higher order spectral analysis (HOS) and Triangular Interpolation of NN interval histogram (TINN) that are not used in [11]. Furthermore, optimal feature subset is also optimized by genetic algorithm (GA) to reduce the feature count of optimal feature subset. Optimal feature subset selected by GA contains following features: meanRR, SDNN, NN50, HRVTri, sample entropy, SD2, ratio of SD1 to SD2, low frequency band energy of FFT, 2 bispectrum features from LL region (P2 H1), 3 bispectrum features from LH region (P1, P2 and WCOB (f_{2m}).), 2 bispectrum features of HH region (M_{ave} and H2), 5 bispectrum features from ROI region (P1, WCOB (f_{1m})., H2, H3 and H4).

Another advantage of our method is using shorter HRV signal length during the feature extraction. Table 1 shows that our method only uses 5 minutes of HRV signal length that end immediately prior to VTA onset to predict the VTA onset. In contrast, almost all previous works except [11] used more than 10 minutes of HRV signal for prediction. Rozen, et al., [13] performed multipole analysis on 10 to 60 minutes of HRV signal. Wollman, et al., [14] extracted their HRV time domain features from 20 to 40 minutes HRV signal. Finally, Thong & Raitt [12] identified the HR pattern in 1.8 hours of signal for prediction. Long duration of HRV signal length for VTA onset prediction causes several disadvantages when the prediction method is implemented in battery powered ICD device. Firstly, longer duration of input (HRV signal) introduces longer processing time in the feature extraction stage which may prove prohibitive in real-time prediction and termination of VTA onset. Furthermore, in recent years, many researches [34-36] have been performed to address the power consumption issue in ICD or similar devices that use HRV analysis for real time disease diagnosis. In the case of VTA onset prediction methods, the main concern is that long duration of signal and compute-intensive HRV analysis algorithms may burden the ICD battery life, and consequently shortening its operation time. This may cause higher frequency of body surgery processes to replace the ICD battery, which can affect the health of the patient [34]. (Generally, the ICD device is expected to operate for more than 5 years after it is implanted in the human body). Therefore, shorter HRV signal length in feature extraction stage can reduce both the time lag between input signal and output prediction, and the burden to the battery of electronic device.

4. Conclusion

In this paper, a ventricular tachyarrhythmia (VTA) onset prediction method based on HRV analysis and GA is proposed. With additional type of HRV features and optimization by GA, we have shown that prediction accuracy of proposed VTA onset prediction method outperforms all previous works even with stricter performance evaluation approach. The prediction sensitivity and specificity of our method are also acceptable and balanced when compared to previous works [12-14]. Besides, the selected optimal feature set is also reported. Furthermore, our method only requires 5 minutes HRV signal in length, which is shorter than most of the previous works, for the accurate prediction.

As for limitation, prediction results of proposed prediction method are limited by small sample size (270 in total) of real data from patients although we have used highest number of sample data when compared to previous works. Therefore, results may suffer from a lack of statistical sampling for VTA patient.

Some future works can be done to extend current work. Firstly, the optimal value of parameter for some feature extraction techniques can be investigated to improve prediction performance of HRV features. These parameters include order of AR regressive, embedding dimension and tolerance of sample entropy, window function of FFT and etc. More complex supervised classifier such as back propagation neural network (BPNN) also can be employed to replace the SVM for development of better supervised prediction model.

References

[1] de Luna AB, Coumel P, Leclercq JF. Ambulatory sudden cardiac death: Mechanisms of production of fatal arrhythmia on the basis of data from 157 cases. *American Heart Journal*. 1989; 117(1): 151-159.

[2] Honkola J, Hookana E, Malinen S, Kaikkonen KS, Junttila MJ, Isohanni M, et al. Psychotropic medications and the risk of sudden cardiac death during an acute coronary event. *European heart journal*. 2012; 33(6): 745-751.

[3] Tereshchenko LG, Fetics BJ, Domitrovich PP, Lindsay BD, Berger RD. Prediction of ventricular tachyarrhythmias by intracardiac repolarization variability analysis. *Circulation Arrhythmia and electrophysiology*. 2009; 2(3): 276-284.

[4] Zokaee S, Islamic Azad University Qazvin B, Faez K, Amirkabir University of T. Human Identification Based on ECG and Palmprint. *International Journal of Electrical and Computer Engineering (IJECE)*. 2012; 2(2): 261-266.

[5] Busono P, Khasanah YN. *Development of ECG Feature Extraction Software*. Proceeding of the Electrical Engineering Computer Science and Informatics. 2015; 2(1).

[6] Chiang E, Vellaisamy M. Intelligent Pillow for Heart Rate Monitor. *International Journal of Electrical and Computer Engineering (IJECE)*. 2013; 3(6): 791-796.

[7] Rajendra Acharya U, Paul Joseph K, Kannathal N, Lim CM, Suri JS. Heart rate variability: a review. *Medical and Biological Engineering and Computing*. 2006; 44(12): 1031-1051.

[8] Reed MJ, Robertson CE, Addison PS. Heart rate variability measurements and the prediction of ventricular arrhythmias. *QJM : monthly journal of the Association of Physicians*. 2005; 98(2): 87-95.

[9] Watanabe MA. Heart rate turbulence slope reduction in imminent ventricular tachyarrhythmia and its implications. *Journal of cardiovascular electrophysiology*. 2006; 17(7): 735-740.

[10] Skinner JE, Pratt CM, Vybiral T. A reduction in the correlation dimension of heartbeat intervals precedes imminent ventricular fibrillation in human subjects. *American Heart Journal*. 1993; 125(3): 731-743.

[11] Joo S, Choi K-J, Huh S-J. Prediction of spontaneous ventricular tachyarrhythmia by an artificial neural network using parameters gleaned from short-term heart rate variability. *Expert Systems with Applications*. 2012; 39(3): 3862-3866.

[12] Thong T, Raitt MH. Predicting Imminent Episodes of Ventricular Tachyarrhythmia Using Heart Rate. *Pacing and Clinical Electrophysiology*. 2007; 30(7): 874-884.

[13] Rozen G, Kobo R, Beinart R, Feldman S, Sapunar M, Luria D, et al. Multipole analysis of heart rate variability as a predictor of imminent ventricular arrhythmias in ICD patients. *Pacing and clinical electrophysiology: PACE*. 2013; 36(11): 1342-1347.

[14] Wollmann CG, Gradaus R, Böcker D, Fetsch T, Hintringer F, Hoh G, et al. Variations of heart rate variability parameters prior to the onset of ventricular tachyarrhythmia and sinus tachycardia in ICD patients. Results from the heart rate variability analysis with automated ICDs (HAWAI) registry. *Physiological Measurement*. 2015; 36(5): 1047.

[15] Huang CL, Wang CJ. A GA-based feature selection and parameters optimizationfor support vector machines. *Expert Systems with Applications*. 2006; 31(2): 231-240.

[16] Goldberger AL, Amaral LAN, Glass L, Hausdorff JM, Ivanov PC, Mark RG, et al. PhysioBank, PhysioToolkit, and PhysioNet : Components of a New Research Resource for Complex Physiologic Signals. *Circulation*. 2000; 101(23): 215-220.

[17] Clifford GD, Tarassenko L. Quantifying errors in spectral estimates of HRV due to beat replacement and resampling. *IEEE Transactions on Biomedical Engineering*. 2005; 52(4): 630-638.

[18] McNames J, Thong T, Aboy M. *Impulse rejection filter for artifact removal in spectral analysis of biomedical signals*. Engineering in Medicine and Biology Society, Annual International Conference of the IEEE. 2004.

[19] Rajendra Acharya U, Paul Joseph K, Kannathal N, Lim CM, Suri JS. Heart rate variability: a review. *Medical & biological engineering & computing*. 2006; 44(12): 1031-1051.

[20] Electrophysiology TFotESoCtNASoP. Heart Rate Variability: Standards of Measurement, Physiological Interpretation, and Clinical Use. *Circulation*. 1996; 93(5): 1043-1065.

[21] Hickey B, Heneghan C. *Screening for paroxysmal atrial fibrillation using atrial premature contractions and spectral measures*. Computers in Cardiology. 2002.

[22] Anita B, Fernando Soares S, Ana Paula R, Argentina L. A study on the optimum order of autoregressive models for heart rate variability. *Physiological Measurement*. 2002; 23(2): 325.

[23] Padmavathi K, Gokaraju Rangaraju Institute of E, Technology, Ramakrishna KS, Velagapudi Ramakrishna Siddhartha Engineering C. Classification of ECG signal during Atrial Fibrillation using Burg's method. *International Journal of Electrical and Computer Engineering (IJECE)*. 2015; 5(1): 64-70.

[24] Spyers-Ashby JM, Bain PG, Roberts SJ. A comparison of fast fourier transform (FFT) and autoregressive (AR) spectral estimation techniques for the analysis of tremor data. *Journal of Neuroscience Methods*. 1998; 83(1): 35-43.

[25] Nikias CL, Raghuveer MR. B*ispectrum estimation: A digital signal processing framework*. Proceedings of the IEEE. 1987; 75(7): 869-891.

[26] Pinhas I, Toledo E, Aravot D, Akselrod S. Bicoherence analysis of new cardiovascular spectral components observed in heart-transplant patients: statistical approach for bicoherence thresholding. *IEEE Transactions on Biomedical Engineering*. 2004; 51(10): 1774-1783.
[27] Yu SN, Lee MY. Bispectral analysis and genetic algorithm for congestive heart failure recognition based on heart rate variability. *Computers in biology and medicine*. 2012; 42(8): 816-825.
[28] Zhou SM, Gan JQ, Sepulveda F. Classifying mental tasks based on features of higher-order statistics from EEG signals in brain–computer interface. *Information Sciences*. 2008; 178(6): 1629-1640.
[29] Pincus SM, Goldberger AL. Physiological time-series analysis: what does regularity quantify? *American Journal of Physiology - Heart and Circulatory Physiology*. 1994; 266(4): 1643-1656.
[30] Mohebbi M, Ghassemian H. Prediction of paroxysmal atrial fibrillation based on non-linear analysis and spectrum and bispectrum features of the heart rate variability signal. *Computer methods and programs in biomedicine*. 2012; 105(1): 40-49.
[31] Cortes C, Vapnik V. Support-Vector Networks. *Machine Learning*. 20(3): 273-297.
[32] Narin A, Isler Y, Ozer M. Investigating the performance improvement of HRV Indices in CHF using feature selection methods based on backward elimination and statistical significance. *Computers in biology and medicine*. 2014; 45: 72-79.
[33] Yuwono T, Setiawan NA, Nugroho HA, Persada AG, Prasojo I, Dewi SK, et al. *Decision Support System for Heart Disease Diagnosing Using K-NN Algorithm*. Proceeding of the Electrical Engineering Computer Science and Informatics. 2015; 2(1).
[34] Kilhwan K, Unsun C, Yunho J, Jaeseok K, editors. *Design and implementation of biomedical SoC for implantable cardioverter defibrillators*. Solid-State Circuits Conference, ASSCC '07 IEEE Asian. 2007.
[35] Massagram W, Hafner N, Chen M, Macchiarulo L, Lubecke VM, Boric-Lubecke O. Digital Heart-Rate Variability Parameter Monitoring and Assessment ASIC. *IEEE Transactions on Biomedical Circuits and Systems*. 2010; 4(1): 19-26.
[36] Fang WC, Huang HC, Tseng SY. Design of heart rate variability processor for portable 3-lead ECG monitoring system-on-chip. *Expert Systems with Applications*. 2013; 40(5): 1491-1504.