Automatic Arrhythmia Classification Method Using Simple Statistical Features

Khudhur A. Alfarhan[#1], MohdYusoffMashor[#2], Abdul Rahman MohdSaad[*3]

[#]*Electronic & Biomedical Intelligent Systems (EBItS) Research Group, School of Mechatronic Engineering, Universiti Malaysia Perlis (UniMAP), Pauh Putra Campus, 02600 Arau, Perlis, Malaysia*

[*]*Electronic Department, Faculty of Engineering Technology, Universiti Malaysia Perlis (UniMAP), Unicity, Malaysia*

Abstract—*Arrhythmia detection and classification methods are normally involve complex procedures and it is quite hard to achieve good accuracy. Current study presents an arrhythmia detection and classification method using simple statistical features extraction methods and library support vector machine (LIBSVM) classifier. The electrocardiogram (ECG) signals for three classes, namely normal sinus rhythm (NSR), premature atrial contraction (PAC) and premature ventricle contraction (PVC) were obtained from MIT arrhythmia database. The obtained ECG signals were denoised using bandpass filter. The denoised signals were segmented into 3.34 seconds. The segmented ECG signals were normalized to prepare them for features extraction stage. The R-R intervals were calculated to extract the features from them. The features extraction methods that were used in this study are the mean absolute value (MAV) of ECG segment, root mean square (RMS) of ECG segment, the median of R-R intervals, and standard deviation (SD) of R-R intervals. The extracted features were classified using LIBSVM classifier. The achieved accuracy from this method is 98.54%. The results showed that the proposed method for arrhythmia detection and classification is accurate, simple and reasonable method comparing to the other studies.*

Keywords — *Electrocardiogram (ECG), Arrhythmia, LIBSVM, Statistical Features,*

I. INTRODUCTION

Arrhythmia detection and classification has been widely studied recently. There are many people have arrhythmia around the world and in Malaysia [1]. It is one of the common heart's diseases and it has many types. Usually, the ECG commonly used for diagnosing the arrhythmia. The physician can detect the arrhythmia from the plot of the ECG because most arrhythmia types effect heart's rhythm.

The recent studies of arrhythmia detection and classification used complicated procedures for extracting the features from the ECG signal and for classifying the arrhythmia. Fei [2] developed a detection and classification method for detecting and classifying the arrhythmia from ECG signal by using LIBSVM and ECG waves durations. The ECG signals were segmented into one pulse window to prepare it for the feature extraction process. The method used the intervals of R-R, P wave, QRS wave, T wave and the P-Q length. These intervals were extracted by using wavelet biorthogonal spline type. These intervals normalized and classified using LIBSVM classifier. The classification accuracy result is 95.652%. Bulusu et al. [3] proposed method for classifying arrhythmia using 12 morphological features and discrete wavelet transform (DWT). Three features extracted from the S-T segment, namely the ST-deviation, the slope of the ST-segment and the correlation coefficients. There are many morphological features extraction methods applied on the ECG signal like Q-R width, R-R interval, autocorrelation Value, the area under Q-R, etc. The extracted features were classified using support vector machine (SVM). The overall accuracy that was obtained from this method was 96.35%. Sarkaleh and Shahbahrami [4] proposed an arrhythmia classification method using DWT and artificial neural network (ANN). The ECG signal was filtered to remove the noise and ECG baseline wander. The Daubechies DWT applied to the signal to extract the features. The features vectors classified using Multi-Layer Perceptron (MLP) neural network. The obtained accuracy from this method is 96.5%. Zhang et al. [5] proposed a classification method for arrhythmia using five morphological features and SVM classifier. The morphological features that were extracted from ECG signal are inter-beat intervals, intra-beat intervals, morphological amplitudes, morphological areas and morphological distance as well as there are many other sub-features like R-R intervals, P-P intervals, QRS distance, the area under P-wave etc. The extracted features were classified using SVM classifier. The overall accuracy of this method is 86.91%. Pooyan1 and Akhoondi [6] proposed an arrhythmia detection and classification method using 20 morphological features and SVM. The ECG baseline wander and ECG noise removed using bandpass filter. The morphological features extracted from the denoised ECG signal. The extracted morphological features were 20 like R-R intervals, T-T intervals, Q-S intervals, falling and rising slope of QRS, the autocorrelation of T-T intervals, amplitudes of Q, R, and S, mean and standard deviation (SD) of T-wave duration and area under T-wave, etc. The extracted features classified using SVM classifier. The overall accuracy obtained from this method was 94.6%.

In this study, four statistical features will be extracted from the ECG signal. These features are MAV of the ECG signal, RMS of the ECG signal, the median of the R-R intervals, and SD of the R-R intervals. The extracted features will be classified using LIBSVM classifier. LIBSVM will classify two types of arrhythmia, namely PVC and PAC as well as the NSR.

II. Methodology

ECG signals for three classes were obtained from Massachusetts institute of technology (MIT) arrhythmia database. The collected ECG signals will be preprocessed (filtration, segmentation, and normalization). After that, the R-R intervals will be calculated and four statistical features will be extracted. Finally, the LIBSVM classifier will be used to classify three classes, namely NSR, PVC and PAC.

A. Filtration

ECG signal affected by different types of noise, like power line noise (50Hz or 60Hz), body muscles noise, and breathing noise. Generally, the noise in ECG signal is divided into two bands: high-frequency (as a power line noise) and low-frequency noise which is a breathing noise. Breathing noise known as ECG baseline wandering, it makes the ECG baseline wavy. A bandpass zero-phase FIR filter was used in this study, to filter the ECG signal. This filter was designed with 400th order and with a cutoff frequency of 0.5 Hz and 40 Hz. This filter was used to remove the ECG baseline wandering, power line noise, and other noise [7] [8].

B. Segmentation

Filtered ECG signals are ready for segmentation step. This step determines the length of the ECG signals that will be used in features extraction step. In this research work, the segment length of each ECG signal is 668 data samples that equal to 3.34 seconds. The 3.34 seconds segment length can frame four ECG pulses if the heartbeat rate (HBR) was 72 beat per minute (BPM). The first sample of the long window was chosen arbitrarily in the segmentation process. In other words, the required ECG pulses were segmented without any reference point such as the position of the R-peaks. Fig 1 shows four examples for ECG signal segmentation. Finally, the resulting segments of the classification group are three classes: NSR, PAC, and PVC. The classification group comprises of 288 segments, (96 for each class). Each class dataset is divided into two sets. Where 75% of the dataset is used to train the classifier and 25% of the dataset is used to test the classification.

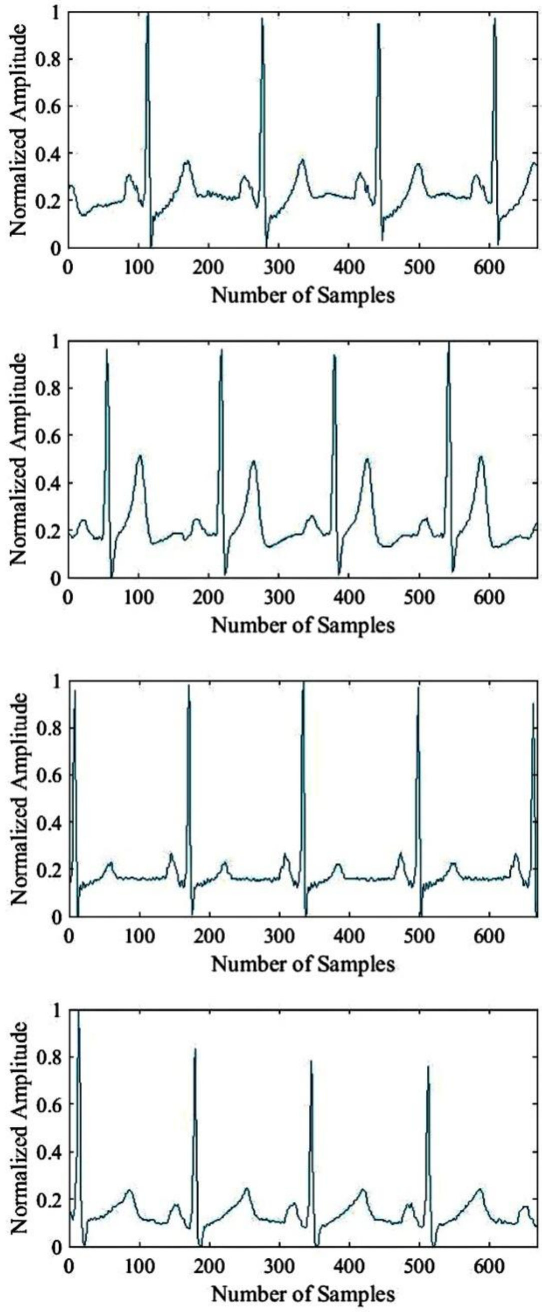

Fig. 1The segmented and normalized ECG signals

C. Normalization

The normalization is a very important step, it normalizes the data to be in specific range. In this study, the normalization is applied to the segmented ECG signals to make the amplitude of the ECG signals range from zero to one. Equation (1) was used to normalize the segmented ECG signal.

$$NormalizedECG = \frac{x_n - x_{min}}{x_{max} - x_{min}} \quad (1)$$

Where (x_n) are the ECG signal samples, (x_{max}, x_{min}) are the maximum and minimum amplitudes in the ECG segment. The normalized ECG signal was shown in Fig 1.

D. Features Extraction

Statistical features extraction methods that were used in this study are MAV, RMS, the median of R-R intervals, and the standard deviation (SD) of R-R intervals. R-R intervals are extracted by finding the R-peaks in each segment and calculating the periods between them. To find the R-peaks of each normalized ECG segment, the ECG segment is squared and the threshold value is applied to it. After that, finding the highest points in each peak above the threshold line. The highest points that were found are the R-peaks [9], [10]. Fig 2 shows the threshold line was applied on the squared ECG signal.

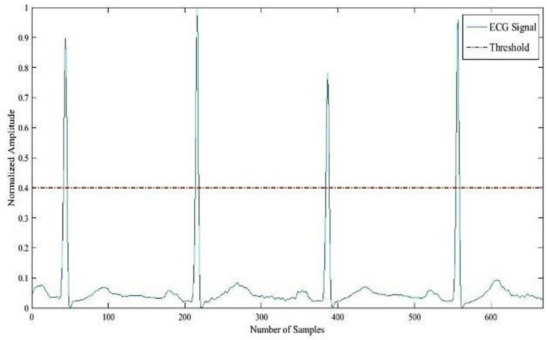

Fig. 2 Threshold line was applied on the squared ECG signal

MAV of the ECG signal is calculated by using the following equation:

$$MAV = \frac{1}{N}\sum_{n=0}^{N-1}|x(n)| \quad (2)$$

where N is the signal length, n is the number of the sample, and $x(n)$ is the ECG signal sample [11]. RMS of the ECG signal is calculated by using the following formula:

$$RMS = \sqrt{\frac{1}{N}\sum_{n=1}^{N} x^2\,(n)} \quad (3)$$

where N is the signal length, n is the number of the sample, and $x(n)$ is the ECG signal sample [12]. The median of the even number of the R-R intervals is calculated by using the following formula:

$$Median = \frac{\frac{R_n}{2}+\left(\frac{R_n}{2}+1\right)}{2} \quad (4)$$

and median of an odd number of the R-R intervals is calculated by using the following formula:

$$Median = \frac{R_n+1}{2} \quad (5)$$

where (R_n) is the number of the R-R intervals in the ECG signal segment [12]. SD of the R-R intervals is calculated by using the following equation:

$$SD = \sqrt{\frac{1}{N-1}\sum_{n=1}^{N}(x_n - M)^2} \quad (6)$$

where N is the number of R-R intervals, n is the number of R-R interval sample, x_n is the value of R-R interval sample, and M is the mean of the R-R intervals [13]. The results of all these approaches are used to create the features vector of the ECG signal.

E. Classification

The multi-class LIBSVM classifier with Gaussian radial basis function (RBF) kernel was used in this study to classify three types of the ECG signals. These types of ECG signals are divided into three classes, namely NSR, PAC, and PVC. LIBSVM is one of the most commonly used classifiers for arrhythmia classification. The training dataset is used to train the LIBSVM classifier where the decision function of the LIBSVM [14] is as follow:

$$sign(w^T\Phi(x)+b) = sign\left(\sum_{i=1}^{L} y_i\,\alpha_i K(x_i,x)+b\right) \quad (7)$$

Where $K(x_i,x)$ is the kernel function, x are the training vectors, i=1,....,L, $y$is the vector of the training labels, α contains the support values, w is the separating hyperplane in features space, and b is a bias term [14]. $y_i\alpha_i\forall_i$,b, support vectors, kernel parameters and label names are stored in the model object to be used in prediction step. After completing the training, the testing dataset is used to test the classification. The LIBSVM uses the following formula for prediction:

$$f(z) = sign\left(\sum_{i=1}^{totalSV} y_i\,\alpha_i K(x_i,z)-b\right) sign(\langle w,\Phi(z)\rangle - b) \quad (8)$$

Where the *totalSV* is the total support values.

RBF Kernel parameter gamma (γ) and LIBSVM parameter cost (C) must be tuned to find the highest accuracy. The parameter C is a trade-off between the error of training and the flatness of the solution. Then the best values of the parameters (C, γ) used to train all training dataset and create the final LIBSVM model [15]. The LIBSVM is a simple classification tool and widely used to classify the practical problems [16] including ECG beats classification. Hence, it was used in this research to classify three classes, namely NSR, PAC, and PVC.

III. Experimental Results and Discussion

The LIBSVM classifier was used to classify three classes of ECG signals. Where these signals were segmented into 3.34 seconds, and their features were extracted by using statistical features extraction methods. The accuracy of the LIBSVM was calculated by using the following equation:

$$Accuracy = \frac{correctlypredicteddata}{totaltestingdata} \times 100\% \quad (9)$$

All of the C and γ values were tuned to find the optimum accuracy. Classification accuracy results are shown in Table 1 with the values of C and γ.

The highest accuracy is noticeable when γ values range from 0.1 to 200. Whereas, the lowest and highest γ values show low classification accuracy. In Table 1 the highest accuracy results were highlighted to be clear. Fig 3 shows the classification accuracy related to the values of γ and C. The noticed from Fig 3 that the highest accuracy results are confined between γ=0.1 to γ=200 as shown clearly in the figure.

TABLE 1
CLASSIFICATION ACCURACY WITH VALUES OF C AND γ

Cost C / Gama γ	C = 3	C = 5	C = 6	C = 10	C = 10000
γ = 0.05	87	97.08	97.81	96.71	97.81
γ = 0.1	95.98	97.81	98.17	98.54	97.81
γ = 0.3	98.54	98.17	98.54	98.17	98.17
γ = 1	98.54	98.54	98.54	98.54	97.81
γ = 200	97.81	98.17	98.17	97.44	97.81
γ = 1000	96.35	96.17	96.35	96.71	96.35

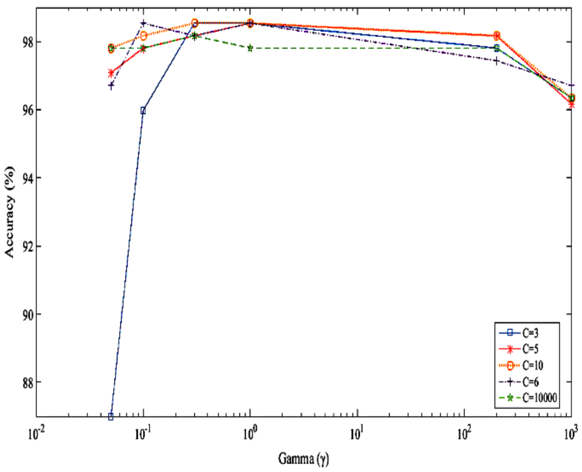

Fig. 3 Classification accuracy related to the values of γ and C

Table 2 shows the comparison between proposed method of arrhythmia classification and previous studies.

TABLE 2
COMPARISON BETWEEN PREVIOUS STUDIES AND PROPOSED METHOD

Reference	Number of features	Classifier	Accuracy (%)
Fei [2]	6	LIBSVM	95.652
Bulusu et al. [3]	13	SVM	96.35
Sarkaleh and Shahbahrami [4]	1	ANN	96.5
Zhang et al. [5]	5	SVM	86.91
Pooyan1 and Akhoondi [6]	20	SVM	94.6
Proposed method	5	LIBSVM	98.54

It is noticeable from Table 2 that the proposed method has slightly highest accuracy compared to the other studies with a reasonable number of features.

IV. CONCLUSIONS

In this study, two types of arrhythmia, namely PVC and PAC as well as NSR from MIT arrhythmia database were examined. Five features extraction methods, namely R-R intervals, MAV, RMS, median, and SD were applied to extract the features from the ECG signal. The LIBSVM classifier was used to classify the three classes. The highest accuracy results are confined between γ=0.1 to γ=200 and between C=3 to C=10000. It can be concluded that the proposed method of arrhythmia detection and classification is highly accurate, simple and reasonable comparing to the previous studies and it has potential to be used in heart monitoring systems or for heart diseases diagnosis.

REFERENCES

[1] WHR, "Coronary Heart Disease in Malaysia," *World Health Rankings*, 2014. [Online]. Available: http://www.worldlifeexpectancy.com/malaysia-coronary-heart-disease.

[2] S. W. Fei, "Diagnostic study on arrhythmia cordis based on particle swarm optimization-based support vector machine," *Expert Syst. Appl.*, vol. 37, no. 10, pp. 6748–6752, 2010.

[3] S. C. Bulusu, M. Faezipour, V. Ng, M. Nourani, L. S. Tamil, and S. Banerjee, "Transient ST-segment episode detection for ECG beat classification," in *Proceedings of the 2011 IEEE/NIH Life Science Systems and Applications Workshop, LiSSA 2011*, 2011, pp. 121–124.

[4] M. K. Sarkaleh and A. Shahbahrami, "Classification of ECG arrhythmias Using Discrete Wavelet Transform and Neural Networks," *Int. J. Comput. Sci. Eng. Appl.*, vol. 2, no. 1, pp. 1–13, 2012.

[5] Z. Zhang, J. Dong, X. Luo, K. S. Choi, and X. Wu, "Heartbeat classification using disease-specific feature selection," *Comput. Biol. Med.*, vol. 46, no. 1, pp. 79–89, 2014.

[6] M. Pooyan and F. Akhoondi, "Accurate Detection and Cassification of Ventricular Abnormamlitiess Using Morphological Features," *J. Fundam. Appl. Sci.*, vol. 8, no. 3S, pp. 1659–1670, 2016.

[7] F. Buendía-Fuentes, M. a. Arnau-Vives, a. Arnau-Vives, Y. Jiménez-Jiménez, J. Rueda-Soriano, E. Zorio-Grima, a. Osa-Sáez, L. V. Martínez-Dolz, L. Almenar-Bonet, and M. a. Palencia-Pérez, "High-Bandpass Filters in Electrocardiography: Source of Error in the Interpretation of the ST Segment," *ISRN Cardiol.*, vol. 2012, pp. 1–10, 2012.

[8] B. S. Raghavendra, D. Bera, A. S. Bopardikar, and R. Narayanan, "Cardiac arrhythmia detection using dynamic time warping of ECG beats in e-healthcare systems," in *2011 IEEE International Symposium on a World of Wireless, Mobile and Multimedia Networks*, 2011, pp. 1–6.

[9] E. Dolatabadi and S. Primak, "Ubiquitous WBAN-based electrocardiogram monitoring system," *2011 IEEE 13th Int. Conf. e-Health Networking, Appl. Serv. Heal. 2011*, pp. 110–113, 2011.

[10] S. Xue, X. Chen, Z. Fang, and S. Xia, "An ECG arrhythmia classification and heart rate variability analysis system based on android platform," *2015 2nd Int. Symp. Futur. Inf. Commun. Technol. Ubiquitous Healthc.*, pp. 1–5, 2015.

[11] E. M. A. Anas, S. Y. Lee, and K. H. Hasan, "Sequential algorithm for life threatening cardiac pathologies detection based on mean signal strength and EMD functions," *Biomed. Eng. Online*, pp. 1–22, 2010.

[12] J. F. Kenney and E. S. Keeping, *Mathematics of Statistics*, 3rd ed. Princeton, NJ: Van Nostrand, 1962.

[13] L. I. Smith, *A tutorial on Principal Components Analysis Introduction*, 1st ed. 2002.

[14] C. Chang and C. Lin, "LIBSVM : A Library for Support Vector Machines," *ACM Trans. Intell. Syst. Technol.*, vol. 2, pp. 1–39, 2013.

[15] L. Qiao, C. Rajagopalan, and G. D. Clifford, "A machine learning approach to multi-level ECG signal quality classification," *Comput. Methods Programs Biomed.*, vol. 117, no. 3, pp. 435–447, 2014.

[16] S. Chen, W. Hua, Z. Li, L. Jian, and X. Gao, "Heartbeat classification using projected and dynamic features of ECG signal," *Biomed. Signal Process. Control*, vol. 31, pp. 165–173, 2017.