Journal of Multidisciplinary Engineering Science and Technology (JMEST)
ISSN: 3159-0040
Vol. 2 Issue 3, March - 2015

Electrocardiogram Features Extraction and Classification for Arrhythmia Detection

Abdelhaq Ouelli, Benachir Elhadadi, Hicham Aissaoui, and Belaid Bouikhalene
Laboratoire de Développement Durable
Sultan Moulay Slimane University
Beni Mellal, Morocco.
ouelli2000@yahoo.fr

***Abstract*— This paper present a new automated detection method for cardiac arrhythmia. The detection system is implemented with integration of feature extraction and classification parts. In feature extraction phase of proposed method, the feature values for each arrhythmia are extracted using autoregressive (AR) and multivariate autoregressive (MVAR) modeling of one-lead and two-lead electrocardiogram signals. Obtained features are used as input to the classifier. The classification is performed using a quadratic discriminant function (QDF) and a multilayer perceptron (MLP). The results show that the MVAR coefficients produce the best accuracy rate.**

Keywords—arrhythmia classification; feature extraction; MIT-BIH database; multilayer perceptron; multivariate autoregressive modeling.

I. INTRODUCTION

The development of accurate and quick methods for automatic electrocardiogram (ECG) classification is vital for clinical diagnosis of heart disease. Arrhythmia can be defined as either an irregular single heartbeat or as an irregular group of heartbeats. The ECG is the most important biosignal used by cardiologists for diagnostic purposes. The ECG signal provides key information about the electrical activity of the heart. The early detection of the cardiac arrhythmias can prolong life and enhance the quality of living through appreciates treatment. In the past decades, many automatic ECG arrhythmia classification systems have been developed using computational intelligence techniques [1]-[3]. A successful ECG arrhythmia classification usually involves three important procedures: signal preprocessing, feature extraction, and classifier construction. Feature extraction is the important procedure that usually influences the classification performance of any ECG arrhythmia classification system. Therefore, the extraction of relevant features to achieve optimal classification results has become primary tasks for the ECG arrhythmia classification problems.

From a review of literature, we found that various ECG feature extraction methods have been successfully applied for arrhythmia classification. The feature extraction methods include time-domain methods [4]; frequency-domain methods [5]; time-frequency domain analysis [6]; wavelet transform [7]; statistical representation [2]; Lyapunov exponents [8]; and Hermite coefficients [9]. To name a few, Yu and Chen [10] extracted a total of 200 points before and after the R points to form the ECG samples. In addition, R–R time intervals were also extracted as the characteristic features for arrhythmia detection. The authors in [11] combined the Fourier transform to observe the changes in QRS complex, and used a neural network to discriminate three kinds of rhythms. Afonso and Tompkins [6] used the time–frequency distribution (TFD) of normal sinus rhythm, ventricular tachycardia, ventricular flutter, and ventricular fibrillation signals to detect ventricular fibrillation. Korurek and Nizam [7] extracted six time-domain features and PCA compressed wavelet coefficients from ECG signals for the ant colony optimization (ACO) classifier to discriminate six kinds of arrhythmias.

The purpose of the present work is to explore the feasibility of autoregressive (AR) and multivariate autoregressive (MVAR) modeling to extract relevant features from one-lead and two-lead ECG signals in order to classify more types of cardiac arrhythmias with higher accuracy. In this study, MAR and AR modeling were performed on the ECG data including normal sinus rhythm (NSR), atrial premature contraction (APC), premature ventricular contraction (PVC), ventricular tachycardia (VT), Ventricular fibrillation (VF) and supraventricular tachycardia (SVT) obtained from the MIT-BIH database. The classification was performed using quadratic discriminant function (QDF) based classifier and multilayer perceptron (MLP) neural networks. In this work, we have used a total of 10 records marked as: 100, 101, 105, 200, 209, 119, 800, cu01, cu02 and cu13; so that a total of 18 582 samples. For each trial, three hundred sample patterns each from the six classes were selected for analysis. A training data set consisted of 150 sample patterns each from the six classes, and the remaining data was used for testing. The results showed that the MVAR modeling is a useful classification and diagnosis tool for the cardiac arrhythmias.

II. METHODS

A. *Preprocessing*

The data in the analysis was obtained from the MIT-BIH Arrhythmia Database (MITDB), Creighton University Ventricular Tachyarrhythmia Database

(CUDB), and MITBIH Supraventricular Arrhythmia Database (SVDB). The NSR, PVC and APC were sampled at 360Hz, the VT and VF were sampled at 250Hz, and the SVT was sampled at 128Hz. The data including VT, VF, and SVT were resampled in order that all the ECG signals in the analysis had a sampling frequency of 360Hz.

ECG recordings are very often contaminated by residual power-line (PL) interference, base-line drift, artifacts and EMG disturbances due to involuntary muscle contractions (tremor) of the patient. The base-line drift resulting from electrochemical processes at the electrode-to-skin barrier is a typical low-frequency noise that distorts the susceptible ST segment. All ECG data have been filtered by a FIR least squares filter with a pass band [0 40] Hz to remove the DC drift and the other types of noise.

The QRS complexes used in this context were extracted from the filtered signals based on the arrhythmia database annotations. A normal ECG refers to the usual case in the health adults where the heart rate is 60~100 beats per minute, RR intervals in APC are shorter than NSR, the RR intervals in VF and VT are much shorter than normal. In the current study, the sample size of the various segments was cycle varying, it is dynamically estimated according to the cardiac rhythm; 34% of RR interval samples before R peak and 66% of RR interval samples after R peak were picked for modeling. It is adequate to capture most of the information from a particular cardiac cycle (P, QRS, and T).

B. *AR and MVAR Modeling*

A general AR model of order p can be expressed as

$$y(n) = -\sum_{i=1}^{p} a_i(n)\, y(n-i) + w(n) \qquad (1)$$

Where $\mathrm{y(n)}$ represents ECG signal, $\mathrm{w(n)}$ represents unknown, zero mean white noise, which is called modeling error, $\mathrm{a_i(n)}$ represents the AR model coefficients.

A general MVAR model of order p can be expressed as

$$Y(n) = -\sum_{i=1}^{p} A_i(n)\, Y(n-i) + W(n) \qquad (2)$$

Where $Y(\mathrm{n})$ represents two-lead ECG signal, $W(n)$ is the vector of multivariate zero mean uncorrelated white noise process, which is called modeling error, $\mathrm{A}_i(\mathrm{n})$ are the coefficients matrices of the MVAR model. The vector $Y(\mathrm{n})$ consists of sampling signal by two channels at n time.

These two equations show that the ECG signals at n time can be estimated by their values at past time and the white noise. The model order p means that p past data samples are needed to predict the present value of the data. The model was estimated from the points of data from each cardiac cycle of the six types of ECG signals. The model order selection was performed on the six types of the ECG signals, various model orders were preselected to choose the more adequate for best classification accuracy, an order of 2 appears sufficient to model ECG signal for the purpose of classification. The Burg algorithm and the algorithm presented in [12] were used to estimate the AR and the MVAR coefficients matrices respectively.

C. *Feature Extraction*

Physiological signals are inherently characterized by a nonstationary time behavior; the ECG signal is highly nonstationary within each beat. The AR and MVAR model coefficients $\mathrm{a}_i(\mathrm{n})$ and $\mathrm{A}_i(\mathrm{n})$ are time-varying. In this section a new method of features extraction for purpose of cardiac arrhythmia classification is presented. The two classifiers based on the quadratic discriminant function and on the MLP neural networks, respectively, use new features extracted from the morphology of each part of the ECG data. The ECG data segments were extracted according to the arrhythmia database annotations. Each pattern contains RR_i samples (number of samples between two successive R peaks), $\frac{1}{3} * \mathrm{RR}_i$ before R_i peak and $\frac{2}{3} * \mathrm{RR}_i$ after R_i peak, the QRS complex were presented by a segment of $\mathrm{RR}_i/6$ samples centered on R_i peak as shown in *Fig. 1* and *Fig. 2*.

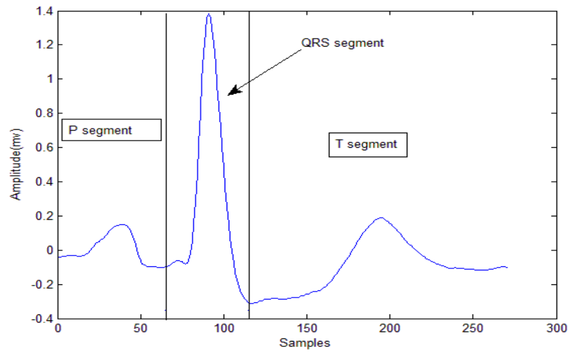

Fig. 1. Extraction of the three part of a particular cardiac cycle from one-lead ECG signal.

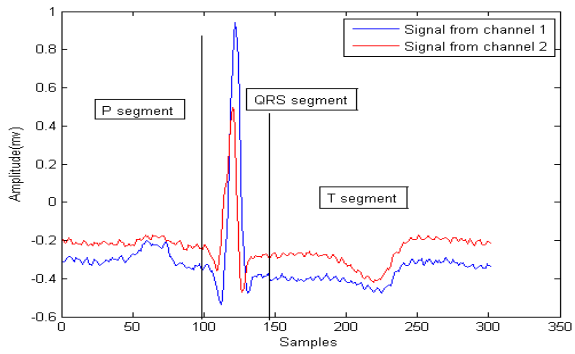

Fig. 2. Extraction of the three parts from a PVC two-lead ECG signal.

Finally AR and MVAR modeling were performed on every segment of the cardiac cycle and coefficients matrices were estimated, respectively, by Burg and ARFIT algorithms presented in [12], a feature vector of 6 elements was produced for each data patterns; 2 elements per each segment of the cardiac cycle using

AR modeling and a feature vector of 24 elements was produced for each data patterns; 8 elements per each segment of the cardiac cycle using MVAR modeling.

III. RESULTS

In this research, six types of ECG signals namely, NSR, APC, PVC, VF, VT and SVT were considered for classification which was performed using QDF based algorithm. One hundred and fifty cases each from the six classes were selected at random in training phase, and the remaining cases were used for testing in testing phase. The sensitivity, specificity and accuracy values were computed for all the ECG classes. The classification accuracy was from 93.3% to 100% as shown in Table 1, the total accuracy is 98.6%.

TABLE I. PERFORMANCE OF THE CLASSIFICATION BASED ON AR COEFFICIENTS AND QDF BASED CLASSIFIER

Classes	NSR	PVC	APC	VF	VT	SVT
Sensitivity	97.3%	98.7%	96.7%	99.3%	99.3%	100%
Specificity	97.3%	99.3%	96.7%	99.3%	98.7%	100%
Accuracy	94.7%	98%	93.3%	98.7%	98%	100%

In the second stage, we designed a number of neural network multilayer perceptron (MLP) by varying the number of layers, the activation functions and the learning algorithms. Performances in terms of convergence speed and classification accuracy are evaluated for different types of ECG data. Among the different learning algorithms, the algorithm Resilient back propagation (RP) shows the best convergence rate and the algorithm of Levenberg-Marquardt (LM) achieved the best overall accuracy rate. The results show that the classification accuracy rate of the six types of arrhythmias was from 97.3% to 100% as shown in Table 2, the total accuracy is 99.6%.

TABLE II. PERFORMANCE OF THE CLASSIFICATION BASED ON AR COEFFICIENTS AND MLP BASED CLASSIFIER

Classes	NSR	PVC	APC	VF	VT	SVT
Sensitivity	99.3%	100%	98.7%	100%	100%	99.3%
Specificity	99.3%	100%	98.7%	99.3%	100%	100%
Accuracy	98.7%	100%	97.3%	99.3%	100%	99.3%

In the last stage, we integrate the MVAR features extraction with the appropriate MLP. The ECG features were extracted by applying multivariate autoregressive modeling of order 2 to the two-lead ECG signals. This resulted in 24 MVAR coefficients to represent an ECG pattern. In this research, five types of ECG signals namely, NSR, APC, PVC, VT and SVT were considered for classification which was performed using multilayer perceptron neural networks. The algorithm of Levenberg-Marquardt (LM) achieved the best overall accuracy rate. Classification accuracy rate of the five types of arrhythmias were 98.7% to 100%, the overall accuracy is 99.7%, which shows a significant improvement as shown in Table 3.

TABLE III. PERFORMANCE OF THE CLASSIFICATION BASED ON MVAR COEFFICIENTS AND MLP BASED CLASSIFIER

Classes	NSR	PVC	APC	VT	SVT
Sensitivity	100%	100%	99.3%	100%	99.3%
Specificity	100%	100%	99.3%	100%	99.3%
Accuracy	100%	100%	98.7%	100%	98.7%

IV. DISCUSSIONS

The main objective of this study was to explore the ability of multivariate autoregressive modeling to extract relevant features from two lead electrocardiogram signals in order to classify certain cardiac arrhythmias. The modeling results showed that an MVAR order of 2 was sufficient for modeling ECG signals for the purpose of the classification, the same order was considered in [13] where two AR coefficients and the mean-square value of QRS complex segment were utilized as features for classifying PVC and NSR using a fuzzy ARTMAP classifier, sensitivity of 97% and specificity of 99% were achieved.

In the current study a variable sample size based on R-R intervals was considered to extract the relevant information from a particular cardiac cycle then better features were extracted from modeling process, in [14] and [15] a fixed sample size of 1.2 seconds and 0.9 seconds respectively had been used, which lack of exactitude due to the heart variability rhythm (APC and VT samples are much shorter than SNR). The use of variable sample size enhances considerably the classification accuracy by producing better features. Furthermore, two lead ECG signals contain more information than one lead ECG signal, so we can produce more reliable features by using two lead ECG signals.

Physiological signals are frequently characterized by a nonstationary time behavior, the ECG signal is highly nonstationary within each beat; the MVAR model coefficients $A_i(n)$ are time-varying. In [14-15] these AR coefficients were considered constants over the whole pattern samples; the nonstationary nature of the ECG signal is omitted. In our study we divide a cardiac cycle into three part (P, QRS, and T segments), MVAR coefficients were estimated for each segment separately, which pick some information about the nonstationary nature of the ECG signal. Considering the nonstationary nature of ECG signals give us better features that enhance the classification performance. Moreover, this procedure simulates the cardiologist behavior in inspecting visually the morphology of these three parts.

The classification results show that MVAR modeling can be used to discriminate between different arrhythmias. The classification results achieved using MVAR modeling is comparable to the recently published results on the classification of cardiac arrhythmias. The current study classifies five types of ECG arrhythmias, the same set of arrhythmias was considered for classification in [14-15]. The accuracy of discrimination of NSR, APC, SVT, PVC, VT and VF were 93.2%%, 96.4%, 100%, 94.8%, 97.7% and 98.6%, respectively with autoregressive modeling and generalized linear model (GLM) based classifier [14]. In [15] the accuracy of detecting the same types of arrhythmia were 97.3%, 97.3%,98.6%, 98%, 99.3%, and 100% using a quadratic discriminant function (QDF) based classifier and autoregressive modeling; the results of our proposed method were 98.7% to 100%.

AR and MVAR modeling methods are used in [16, 17, 18], the extracted features produce good rates of classification accuracy using both QDF and MLP base classifier.

V. Conclusions

Multivariate autoregressive modeling of the three part of the cardiac cycle produces efficient features that characterize separately the three important part of the cardiac cycle; the most relevant information is stored in the QRS complexes.

It should be noted that in addition to selection of classifier type and architecture, feature set selection might also have a vital role in classification results. Further examination of other feature sets with more relevant information about the nonlinear and nonstationary nature of the ECG signal may be warranted in future studies. Other classifiers with higher overall recognition accuracies may also be studied. Reduction of features may improve the performance of the classifier.

References

[1] B.M. Asl, S.K. Setarehdan, and M. Mohebbi, “Support vector machine-based arrhythmia classification using reduced features of heart rate variability signal,” Artif. Intell. Med. 44, pp 51–64, 2008.

[2] S. Osowski and T.H. Linh, “ECG beat recognition using fuzzy hybrid neural network,” IEEE Trans. Biomed. Eng. 48, pp 1265–1271. 2001.

[3] OY.C. Yeh, W.J. Wang, and C.W. Chiou, “Cardiac arrhythmia diagnosis method using linear discriminant analysis on ECG signals,” Measurement 42, pp 778–789, 2009.

[4] S.N. Yu and K.T. Chou, “Integration of independent component analysis and neural networks for ECG beat classification,” Expert Syst. Appl. 34, pp 2841–2846, 2008.

[5] K.I. Minami, H. Nakajima, and T. Toyoshima, “Real-time discrimination of ventricular tachyarrhythmia with Fourier-transform neural network,” IEEE Trans.Biomed. Eng. 46, pp 179–185, 1999.

[6] V.X. Afonso and W.J. Tompkins, “Detecting ventricular fibrillation: Selecting the appropriate time–frequency analysis tool for the application,” IEEE Eng. Med.Biol. 14, pp 152–159, 1995.

[7] M. Korurek and A. Nizam, “Clustering MIT-BIH arrhythmias with Ant Colony Optimization using time domain and PCA compressed wavelet coefficients,” Digit Signal Process. 20, pp1050–1060, 2010.

[8] M.I. Owis, A.H.A. Zied, A.B.M. Youssef, and Y.M. Kadah, “Study of features based on nonlinear dynamical modeling in ECG arrhythmia detection and classification,” IEEE Trans. Biomed. Eng. 49, pp 733–736, 2002.

[9] T.H. Linh, S. Osowski, and M. Stodolski, “On-line heart beat recognition using Hermite polynomials and neuro-fuzzy network,” IEEE Trans. Instrum. Meas.52, pp 1224–1231, 2003.

[10] S.N. Yu and Y.H. Chen, “Electrocardiogram beat classification based on wavelet transformation and probabilistic neural network,” Pattern Recognition Lett. 28, pp 1142–1150, 2007.

[11] K.I. Minami, H. Nakajima, and T. Toyoshima, “Real-time discrimination of ventricular tachyarrhythmia with Fourier-transform neural network,” IEEE Trans. Biomed. Eng. 46, pp 179–185, 1999.

[12] T. Schneider and A. Neumaier, ”Alghorith 808: ARFIT-A Matlab pakage for the estimation of parameters and eigenmodes of multivariate autoegressive models,”ACM Transactions on Mathematical Software, vol. 27, No. 1,pp 58-65, March 2001.

[13] Ham F M and Han S. “Classification of cardiac arrhythmias using fuzzy ARTMAP,”. IEEE Transactions on Biomedical Engineering, 43(4), pp 425-430, 1996.

[14] Dingfei Ge, Narayanan Srinivasan, and Shankar M Krishnan, ”Cardiac arrhythmia classification using autoregressive modeling,”BioMedical Engineering OnLine, pp 1:5, 2002.

[15] Dingfei Ge, HOU Bei-Ping, and XIANG Xin-Jian,”Study of Feature Extraction Based on Autoregressive Modeling in ECG automatic diagnosis,”Acta Automatica Sinika, Vol. 33, N. 5, pp 462-466.May 2007.

[16] A.Ouelli, B.Elhadadi, H.Aissaoui, and B.Bouikhalene, “AR modeling for automatic cardiac arrhythmia diagnosis using QDF based algorithm,” IJARCSSE, 2(5), pp 493-499, 2012.

[17] A.Ouelli, B.Elhadadi, H.Aissaoui, and B.Bouikhalene, “AR modeling for cardiac arrhythmia classification using MLP neural networks,” International Journal of Computer Applications, 47(24), pp 44-51, 2012.

[18] A.Ouelli, B.Elhadadi, and B.Bouikhalene, “Multivariate autoregressive modeling for cardiac arrhythmia classification using multilayer perceptron neural networks,” International Conference on Multimedia Computing and Systems (ICMCS), pp 402- 406, 2014.