Volume 2, Issue 5, May 2012 ISSN: 2277 128X
International Journal of Advanced Research in Computer Science and Software Engineering
Research Paper
Available online at: www.ijarcsse.com

AR Modeling for Automatic Cardiac Arrhythmia Diagnosis using QDF Based Algorithm

A.OUELLI *
Laboratoire de Physique de la Matière et Nanotechnologie University Sultan Moulay Slimane Beni Mellal Morocco
ouelli2000@yahoo.fr

B.ELHADADI
Laboratoire de Physique de la Matière et Nanotechnologie University Sultan Moulay Slimane Beni Mellal Morocco

H.AISSAOUI
Laboratoire de Physique de la Matière et Nanotechnologie University Sultan Moulay Slimane Beni Mellal Morocco

B.BOUIKHALENE
Laboratoire de Physique de la Matière et Nanotechnologie University Sultan Moulay Slimane Beni Mellal Morocco

Abstract— **In this paper, we proposed a three stage technique for cardiac arrhythmia classification. This method is done in three steps: de-noising, features extraction and classification: in the first step, we used a least squares filter for noise reduction of the electrocardiogram (ECG) signals. In the second step, we explore the ability of autoregressive model (AR) to extract relevant features from one-lead electrocardiogram signals in order to classify certain cardiac arrhythmias. In the third step, a Quadratic Discriminant function (QDF) based algorithm was used for classification. The data in the analysis including normal sinus rhythm, atrial premature contraction, premature ventricular contraction, ventricular tachycardia, ventricular fibrillation and supra-ventricular tachycardia is obtained from the MIT-BIH database. The results show the AR coefficients produce good features for ECG classification and diagnosis. The classification accuracies of the six types of arrhythmia were 96.7% to 100% which is a significant improvement.**

Keywords— **Autoregressive model, cardiac arrhythmia, ECG features, ECG classification, QDF, MIT-BIH database.**

I. INTRODUCTION

The cardiovascular disease becomes one of main diseases that threaten the human especially in developing countries. 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. Therefore, we need many techniques that analyze the ECG signal to detect the heart diseases.

Arrhythmia can be classified into following categories: Normal sinus rhythm (NSR), Pre-ventricular contraction (PVC), Ventricular tachycardia (VT), Ventricular fibrillation (VF), Supra ventricular tachycardia (SVT), Atrial premature contraction (APC) etc. Among those threatening arrhythmias, VT and VF are most dangerous because they produce the hemodynamic deterioration. Sudden death accounts for approximately half of all deaths from cardiovascular disease and is generally caused by VT and VF. PVC arrhythmia, result from irritated ectopic foci in the ventricular area of the heart. These foci cause premature contractions of the ventricles that are independent of the pace set by the sinoatrial node. Many studies have shown that PVCs, when associated with myocardial infraction, can be linked to mortality. So is the case with other arrhythmias. Consequently, their immediate detection and treatment is essential for patients with heart disease.

Continuous ECG monitoring permits observation of cardiac variations over an extended period of time, either at the bedside or when patients are ambulatory, providing more information to physicians. Thus, continuous monitoring increases the understanding of patients' circumstances and allows more reliable diagnosis of cardiac abnormalities. Due to the large number of patients in intensive care units and the huge amounts of ECG data, several methods for automated arrhythmia classification have been developed in the past few decades to simplify the monitoring task. In the literature, several methods have been proposed for the automatic classification of ECG signals. Among the most recently published works are those presented in [1–6]. In [1], the authors used a MLP neural network classifier and achieved an accuracy of 88.3% in their testing set. In [2], the authors used morphological information as the features and a neural network classifier for differentiating the ECG beats. In [3], the author used wavelet-transform and adaptive neuro-fuzzy inference system for cardiac arrhythmias classification. In [4, 5], the authors used the power spectral density (PSD) of ECG signals. In [6], the authors have modeled the ECG signals using the MME (modified mixture of experts) network structure with diverse features. Generally, the performance of ECG pattern classification strongly depends on the characterization power of the features extracted from the ECG data and the design of the classifier.

Autoregressive modeling (AR) has been used in various applications, including classification of physiological signals like ECG, EEG, heart rate etc. The advantage of AR modeling is its simplicity and is suitable for real-time classification at the ICU or ambulatory monitoring. AR models are popular due to the linear form of the system, simultaneous equations involving the unknown AR model parameters and the availability of efficient algorithms for computing the solution. AR modeling has been used extensively to model heart rate variability (HRV) and for power spectrum estimation of ECG and HRV signals [7–11].

Several researchers have previously used the AR Modeling coefficients in different ways as parametric features rather than the original signal time series and achieved good classification performance [12, 13, 20 and 22]. Accordingly, in the current work we exploit the capability of AR modeling parameters to classify six types of cardiac arrhythmias namely NSR, APC, PVC, SVT, VT and VF. In the classification stage, we have used the QDF based algorithm. Three hundred sample patterns each from the six classes were selected for classification. In this paper, we propose an automated method for cardiac arrhythmia classification. In the preprocessing module, a FIR least squares filter is used to provide an informative representation that is both robust to noise and tuned to the morphological characteristics of the waveform features. For the feature extraction module we have used a suitable set of features that consist of six coefficients resulting from AR modeling of the three components of the ECG signal (P, QRS, T waves). Then we used a QDF based algorithm for classification. The paper is organized as follows. Section 2 describes the preprocessing module. Section 3 explains the feature extraction. Section 4 describes the database and performance metrics. Section 5 shows some simulation results. Section 6 discusses the results and finally Section 7 concludes the paper.

II. ECG DATA PROCESSING

A. *Preparing Data for AR Modeling*

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 bandpass [0 40] Hz to remove the DC drift and the other types of noise (Fig. 1).

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). Finally the following types of ECG data were ready for AR modeling and feature extraction.

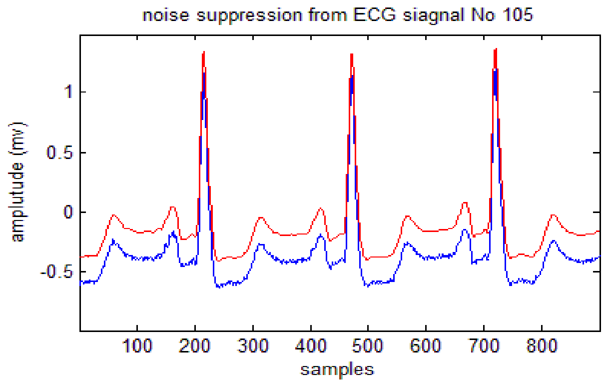

Fig. 1 Suppression of noise from ECG record No105

1) Normal sinus rhythms: All P waves upright, rounded and similar in size and shape. A P wave exists for every QRS complex. Each P wave is the same distance from the QRS complex less than 0.20 seconds. All QRS complexes have the same size and shape and point in the same direction. Each QRS has the same distance from the T waves and the QRS duration is 0.10 seconds or less. Heart rate is varying 60-100 beats/minute and is rhythmic. Refer Fig.2.

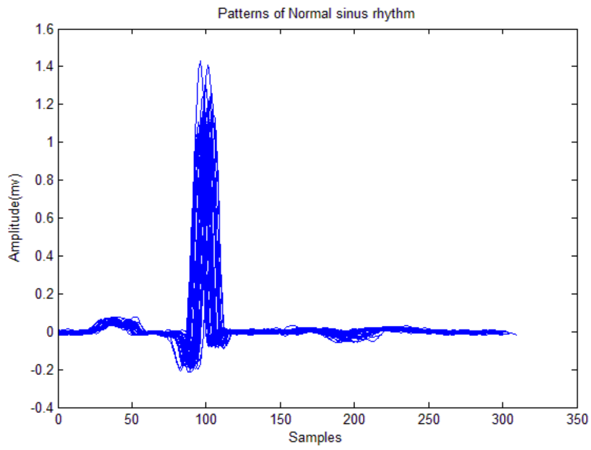

Fig.2 Patterns of NSR (record 100 of MIT-BIH database)

2) Ventricular tachycardia: VT is defined as three or more consecutive beats of ventricular origin at a rate greater than 100 beats/min. There are widened QRS complexes .The rhythm is usually regular, but on occasion it may be modestly irregular. Refer Fig.3.

3) Supra ventricular tachycardia: SVT is any tachyarrhythmia that requires only atrial and/or atrioventricular (AV) nodal tissue for its initiation and maintenance. It is usually a narrow-complex tachycardia that has a regular, rapid rhythm; exceptions include atrial fibrillation and multifocal atrial tachycardia. Aberrant conduction during SVT results in a wide complex tachycardia. SVT occurs in persons of all age groups. This arrhythmia has such a fast rate that the P waves are frequently buried in preceding T waves and difficult to see. Refer Fig.4.

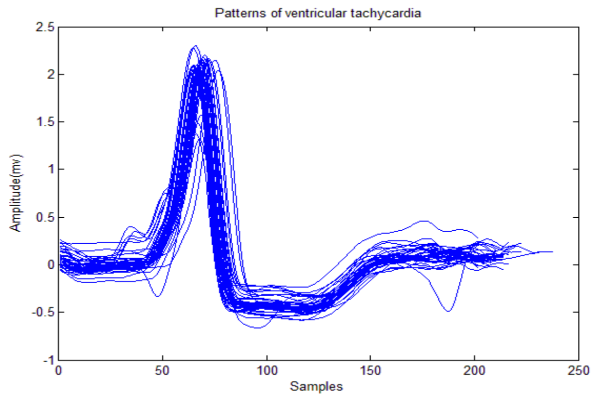

Fig.3 Patterns of VT (record cu13 of CUDB database)

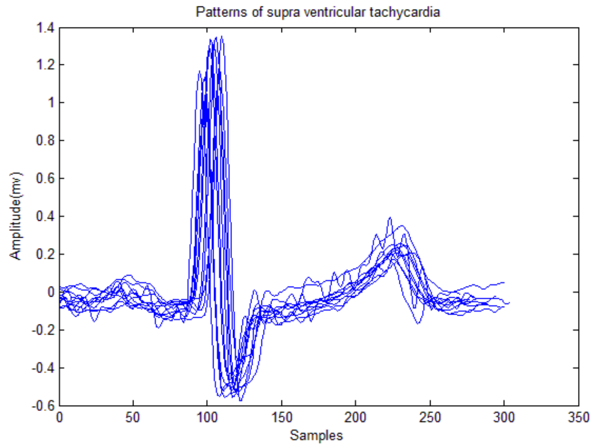

Fig. 4 Patterns of SVT (record 800 of MIT-BIH database)

4) Premature ventricular contraction: PVCs result from an irritable ventricular focus. It may be uniform (same form) or multiform (different forms).Usually PVC is followed by a full compensatory pause because the sinus node timing is not interrupted. Normally the sinus rate produces the next sinus impulse on time. In contrast, PVC may be followed by a non compensatory pause if the PVC enters the sinus node and resets its timing; this enables the following sinus P wave to appear earlier than expected. P waves are not associated with PVC and QRS is wide (>0.1 Sec). Refer Fig. 5.

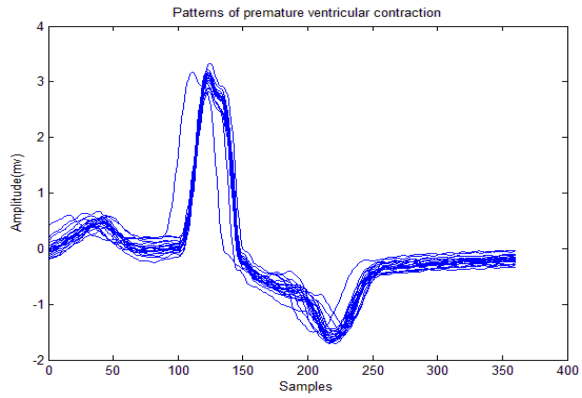

Fig. 5 Patterns of PVC (record 119 of MIT-BIH database)

5) Atrial premature contraction: A single contraction occurs earlier than the next expected sinus contraction. After the APC, sinus rhythm usually resumes. Rhythm is irregular whenever APC occurs. P Waves may have a different shape. PR Interval normally **v**aries otherwise have a normal range (0.12–0.20 sec). QRS interval is Normal (0.06–0.10 sec). An ECG of patient having APC is shown in Fig. 6.

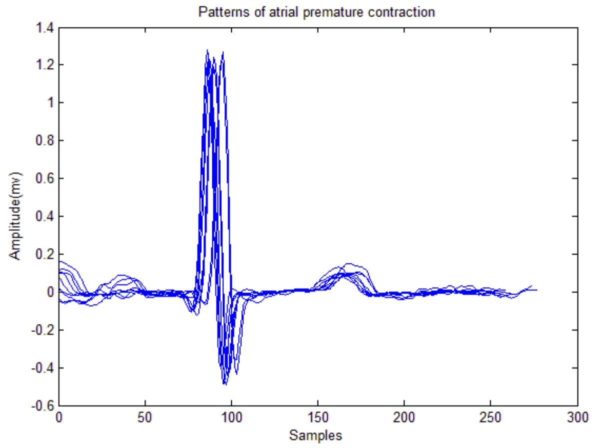

Fig. 6 Patterns of APC (record 209 of MIT-BIH database)

6) Ventricular fibrillation: Chaotic electrical activity occurs with no ventricular depolarization or contraction. The amplitude and frequency of the fibrillatory activity can be used to define the type of fibrillation as coarse, medium, or fine. Small baseline undulations are considered fine; large ones are coarse. No P waves, no QRS interval and no PR interval. Refer Fig. 7.

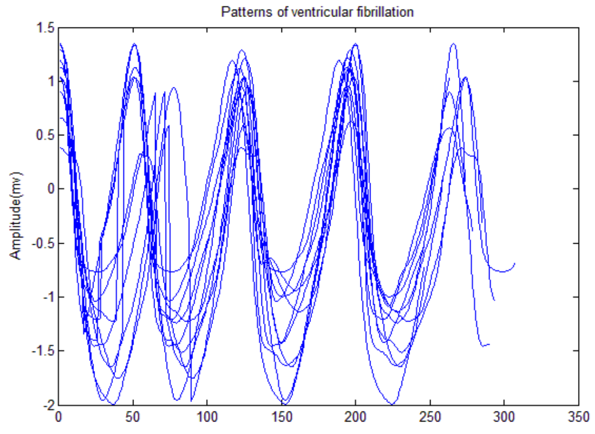

Fig. 7 Patterns of VF (record cu01 of CUDB database)

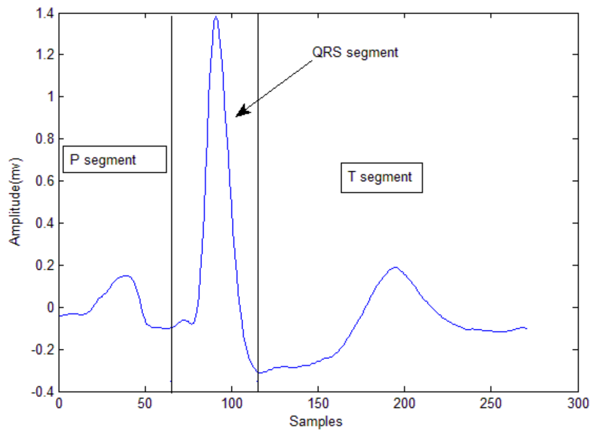

Fig. 8 Extraction of the three part of a particular cardiac cycle

B. Autoregressive Modeling

The six classes of ECG signals with a sampling frequency of 360Hz were utilized for AR modeling after filtering. A general AR model of order P can be expressed as

$$y(n) = \sum_{i}^{P} a_i(n)\, y(n-i) + e(n)$$

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

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, an order of 2 appears sufficient to model ECG signal for the purpose of classification. Burg's algorithm was used to compute the AR coefficients.

III. ECG Feature Extraction

Physiological signals are frequently characterized by a non-stationary time behavior; the ECG signal is highly non-stationary within each beat. The AR model coefficients $a_i(n)$ are time-varying. In this section a new method of feature extraction for purpose of cardiac arrhythmia classification is presented. The classifier based on the QDF algorithm uses 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 segment contains RRi samples (number of samples between two successive R peaks), RRi/3 before Ri peak and 2/3*RRi after Ri peak, the QRS complex were presented by a segment of RRi/6 samples centered on Ri peak as shown in Fig. 8.

Finally AR modeling was performed on every segment of the cardiac cycle and coefficients were estimated by Burg's algorithm, a feature vector of six elements was produced for each data patterns. Table 1 show a sample of these feature vectors for the six considered cardiac arrhythmia.

Table 1 Samples of features vectors of the six ECG types.

APC	-1.8873	0.9735	-1.9244	0.9263	-1.9263	0.9333
NSR	-1.9060	0.9595	-1.9499	0.9504	-1.9398	0.9399
SVT	-1.9485	0.9841	-1.8979	0.9126	-1.9459	0.9541
VF	-1.9584	0.9792	-1.9748	0.9920	-1.9761	0.9925
VT	-1.9733	0.9871	-1.9520	0.9689	-1.9778	0.9808
PVC	-1.9768	0.9872	-1.9778	0.9781	-1.9844	0.9846

IV. Database And Performance Metrics

A. Performance Metrics

One of the significant issues in ECG beat classification is how to appropriately evaluate the performance of a proposed method. In this study, we have considered three statistical indices:

$$Accuracy\ \% = \frac{Number\ \text{of total tested beats} - FP - FN}{Number\ \text{of total tested beats}} \times 100$$

$$\text{Specificity}\ \% = \frac{Number\ \text{of total tested beats} - FP}{Number\ \text{of total tested beats}} \times 100$$

$$Sensitivity\ \% = \frac{Number\ \text{of total tested beats} - FN}{Number\ \text{of total tested beats}} \times 100$$

where FP stands for the words false positive and FN stands for the words false negative. Sensitivity measures how successfully a classifier recognizes beats of a certain class without missing them, whereas specificity measures how exclusively it classifies beats of a certain type.

B. ECG Database

In this study, the MIT-BIH database was used for training and performance evaluation of the proposed ECG classifier.

We used the MIT-BIH Arrhythmia Database (MITDB) [14], Creighton University Ventricular Tachyarrhythmia Database (CUDB) [15], and the MIT-BIH Supraventricular Arrhythmia Database (SVDB) [16] to evaluate our algorithm. The MITDB contains 48 files, 2 channels per file, each channel 1805 seconds long. The CUDB contains 35 files, 1 channel per file, each channel 508 seconds long. The SVDB database contains 78 files, 2 channels per file, each channel 1800 seconds long. In this work, we used a total of 10 records marked as: 100, 101, 105, 200, 209, 119, 800, cu01, cu02 and cu13.

V. RESULTS

The ECG features were extracted by applying AR modeling of order 2 to the ECG signals. This resulted in six AR coefficients to represent an ECG segment. Each beat is featured only with six parameters, two from each beat segment (P, QRS, and T). VTs are characterized by widened QRSs which are less widened in SVTs and P waves are often buried in preceding T waves, then information extracted from the three segments may discriminates them effectively. Fig. 9 shows that there is a very good class separation between VT and SVT. In Fig 10 this discrepancy is less obvious but the class separation will be better when the four other features will be taken in consideration. Then we can conclude that the six features give a good power to classify the different classes of ECG signals.

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 results of the classification are shown in Tables 2 and 3. The results for a sample training set are shown in Table 2 and the classification sensitivity, specificity and accuracy for various classes are shown in Table 3. The accuracy of detecting NSR, APC, PVC, SVT, VT and VF were 97.3%, 96.7%, 98%, 100%, 98%, and 99.3% respectively, the overall accuracy is 98.56%.

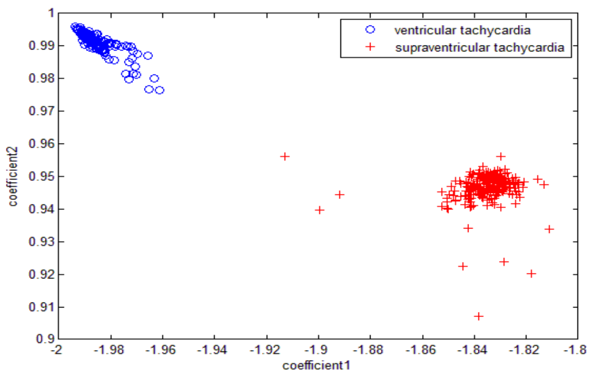

Fig.9 Discrepancy of VT and SVT

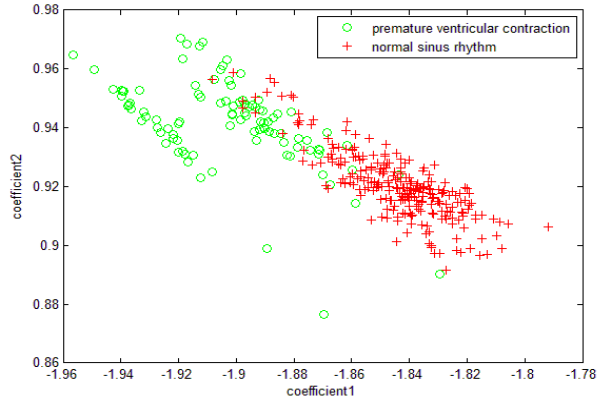

Fig.10 Discrepancy of PVC and NSR

Table 2
Classification results based on AR coefficients for sample training set

classes	NSR	PVC	APC	VF	VT	SVT
NSR	146	0	4	0	0	0
PVC	0	148	0	0	1	0
APC	4	0	145	1	0	0
VF	0	0	1	149	0	0
VT	0	2	0	0	149	0
SVT	0	0	0	0	0	150

Table 3
Performance of the classification based on AR coefficients

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	97.3%	98%	96.7%	99.3%	98%	100%

VI. DISCUSSION

The main objective of this study was to explore the ability of autoregressive modeling (AR) to extract relevant features from one-lead electrocardiogram signals in order to classify certain cardiac arrhythmias. The modeling results showed that the AR order of 2 was sufficient to model ECG signals for the purpose of the classification, the same order was considered in [17] 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 so better features were extracted from modeling process, in [13 and 22] 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 enhance considerably the classification accuracy by producing better features.

Physiological signals are frequently characterized by a non-stationary time behavior, the ECG signal is highly non-stationary within each beat; the AR model coefficients $a_i(n)$ are time-varying. In [12, 13, and 22] these AR coefficients were considered constants over the whole pattern samples; the non stationary nature of the ECG signal is omitted. In our study we divide a cardiac cycle into three part (P, QRS, and T segments), AR coefficients were estimated for each segment separately, which pick some information about the non stationary nature of the ECG signal. Considering the non stationary nature of ECG signals give us better features that enhance the classification performance.

The classification results show that AR modeling can be used to discriminate between different arrhythmias. The classification results achieved using AR modeling is comparable to the recently published results on the classification of cardiac arrhythmias. Normal and abnormal PVC conditions have been classified using the QRS power spectrum and self–organizing maps with sensitivity of 82.71% and specificity of 88.06% [18]. Accuracy of 93% and 96% has been reported for VT and VF respectively using a modified sequential probability ratio test algorithm [10]. An overall accuracy of 93% to 99% was achieved with decimated ECG data and artificial neural networks [19]. The total least squares-based Prony modeling technique produced an accuracy of 95.24%, 96% and 97.78% for SVT, VT and VF respectively [20].

The current study classifies six types of ECG arrhythmias, the same set of arrhythmias was considered for classification in [12 and 21]. The accuracy of discrimination of NSR, APC, SVT, PVC, VT and VF were 99.307%, 99.274%, 99.854%, 98.344%, 99.441% and 99.883%, respectively with a support vector machine (SVM) based classifier [21]. In [12] the accuracy of detecting the same types of arrhythmia were 93.2% to 100% using AR modeling and a generalized linear model based algorithm, the results of our proposed method were 96.7% to 100%.

VII. CONCLUSIONS

We have proposed an efficient method to accurately classify six types of cardiac arrhythmia. This method includes three modules: an efficient preprocessing module, a feature extraction module and a classifier. We used a FIR least squares filter for smoothing and normalisation in the preprocessing module. In the feature extraction module we extracted 6 coefficients from AR modeling as the effective features for differentiating normal beats, and other beats. Then classification was performed using QDF based algorithm. An overall classification accuracy of 98.56% for testing dataset was achieved over various ECG signals from the MIT–BIH database.

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

REFERENCES

[1] H.G. Hosseini, K.J. Reynolds, D. Powers, A multi-stage neural network classifier for ECG events, in: Proceedings of the 23rd International Conference on IEEE EMBS, vol. 2, 2001, pp. 1672–1675.

[2] O.T. Inan, L. Giovangrandi, G.T.A. Kovacs, Robust neural-networkbased classification of premature ventricular contractions using wavelet transform and timing interval features, IEEE Trans. Biomed. Eng. 53 (2006) 2507–2515.

[3] P.D.Khandait,N.G.Bawane,S.S.limaye,S.P.Khandait: efficient ECG Abnormalities recognition using neuro-fuzzy Approach.IJARCSSE,vol2,issue4,402~407,April 2012.

[4] U.R. Acharya et al., Automatic identification of cardiac health using modeling techniques: a comparative study, Inform. Sci. 178 (2008) 4571–4582.

[5] C.H. Lin, Frequency-domain features for ECG beat discrimination using grey relational analysis-based classifier, Comput. Math. Appl. 55 (2008) 680–690.

[6] E.D. Ubeyli.Features for analysis of electrocardiographic changes in partial epileptic patients, Expert Syst. Appl. 36 (2009) 6780–6789.

[7] Mukhopadhyay S, Sircar P: Parametric modelling of ECG signal. Med Biol Eng Comp 1996, 34:171-173

[8] Pinna GD, Maestri R, Cesare AD: Application of time series spectral analysis theory: analysis of cardiovascular variability signals. Med Biol Eng Comp 1996, 34:142-148

[9] Bennett FM, Chrisstini DJ, Ahmed H, Lutchen K: Time series modeling of heart rate dynamics. Proc Comp Cardiol 1993, 273-276

[10] Arnold M, Miltner WHR, Witte H: Adaptive AR modeling of nonstationary time series by means of Kalman filtering. IEEE Trans Biomed Eng 1998, 45:553-562

[11] Mainardi LT, Bianchi AM, Baselli G, Cerutti S: Pole-tracking algorithms for the extraction of time-variant heart rate variability spectral parameters. IEEE Trans Biomed Eng 1995, 42:250-258

[12] Mehmet Engin : ECG beat classification using neuro-fuzzy network. Pattern Recognition Letters 25 (2004) 1715–1722

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

[14] Massachusetts Institute of Technology, MIT-BIH arrhythmia database. [http://www. physionet.org /physiobank/database/mitdb].

[15] Massachusetts Institute of Technology, CU database. [http://www. physionet.org / physiobank /database/cudb].

[16] Massachusetts Institute of Technology, MIT-BIH Supraventricular Arrhythmia database. [http://www.physionet.org / physiobank /database /svdb]

[17] Ham F M, Han S. Classification of cardiac arrhythmias using fuzzy ARTMAP. *IEEE Transactions on Biomedical Engineering*, 1996, 43(4): 425~430

[18] M.L.Talbi, A.Charaf. PVC discrimination using the QRS power spectrum and self–organizing maps. Computer methods and programs in biomedicine.2009,223~231

[19] Fukunaga K. *Introduction to Statistical Pattern Recognition*. New York: Academic Press, 1990, 153~154 and 400~409

[20] Ge Ding-Fei, Xia Shun-Ren. Application of AR model in telediagnosis of cardiac arrhythmias. *Chinese Journal of Biomedical Engineering*, 2004, 23(3): 222~229.

[21] Mi Hye Song, J Lee, S P Cho. Support vector machine based arrhythmia classification using reduced features. International journal of control, automation, and system, vol 3, no. 4, pp 571-579,December 2005.

[22] Dingfei Ge, HOU Bei-Ping and XIANG Xin-Jian: Study of Feature Extraction Based on Autoregressive Modeling in ECG Automatic Diagnosis. ACTA AUTOMATICA SINICA,Vol.33,No.5,462~466.May,2007

