Cardiac Arrhythmia Classification using Beat-by-Beat Autoregressive Modeling

Chusak Thanawattano and Thaweesak Yingthawornsuk

***Abstract*—This paper presents the classification of the cardiac arrhythmia whose features are extracted by applying the autoregressive signal modeling. The different coefficients of the 4th to 9th-order AR models were tested with the support vector machine as one-against-the-rest approach for each arrhythmia in order to investigate their performance. The results lead to the proposed method where the first step is to apply APC-against-non-APC SVM on the extracted AR features. Secondly, the PVC-against-Normal SVM classifier is applied to the non-APC feature classified from the first step. The mean value of the best overall accuracy is 94.40% which is obtained when the 6th-order AR model is applied to modeling. The APC-against-non-APC classification alone provides the high accuracy of 97.11%.**

***Index Terms*—Electrocardiography, support vector machine, classification, autoregressive estimation.**

I. INTRODUCTION

The electrocardiogram (ECG) provides significant clinical information of patients who have the abnormal activity of heart. By using the ECG record physicians can clinically categorize the abnormality into class which the disorder belongs. However, the ECG is normally monitored and recorded in a long time period. This is the time-consuming and inconvenience for the physician since he/she needs to be in alert at all time. More importantly, arrhythmia event can be missed by human detection due to the fatigue of long hours of work. The ECG record can be interpreted using a special machine that has an artificial intelligence inside. The detected arrhythmia, a pattern of ECG signal considered as abnormal heart functioning, event can be used as information for a pre-screening procedure so that the physician can pay attention to the suitable remedy.

There are several approaches for classifying the ECG arrhythmia record [1]-[8]. Support vector machine (SVM) is a classification tool that outperforms several classification methods. Many researches including our previous work [9] demonstrated the procedure of applying the SVM to the classification of arrhythmia. As the two-class classification method, we investigate the performance of SVM whose input feature dimensions, at each classification step, are fixed and freely-selected.

Manuscript received September 12, 2012; revised October 11, 2012.

C. Thanawattano is with National Electronic and Computer Technology Center, Pathumthani, Thailand (e-mail: chusak.tha@nectec.or.th).

T. Yingthawornsuk is with Department of Media Technology, King Mongkut's University of Technology Thonburi – Bang Khunthian Campus, Bangkok, Thailand (e-mail: thaweesak.yin@kmutt.ac.th).

II. MATERIALS AND METHODS

A. Database Selection

In this work, the ECG signal is obtained from MIT-BIH Arrhythmia database [10]. There are three types of the arrhythmia: normal beat (NORMAL), premature ventricular contraction (PVC) and arterial premature contraction (APC). Six files named 100, 109, 107, 118, 208 and 232 are chosen from the arrhythmia database, which contain enough beats of normal, PVC and APC arrhythmia for this experiment. The provided ECG signal is sampled with 360Hz. The classification of aforementioned types is investigated since they are more likely to be incorrectly interpreted by computation machine compared to other waveforms. Fig. 1 shows the general waveform of three different arrhythmia beats.

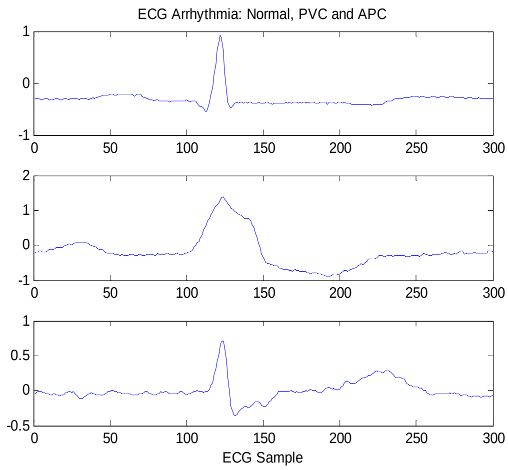

Fig. 1. The ECG waveforms of normal beat (N), premature ventricular contraction (PVC) and arterial premature contraction (APC) where the vertical scales are in mV.

The ECG signal from Modified Lead II (ML-II) is chosen for processing. The database is prepared with the R-peak locations and arrhythmia annotations so that we can skip the QRS detection and perform the feature extraction as described in later section. In our experiment, we use the R-peak of QRS complex to be the significant sampling point. By selecting from 120 samples prior to the R-peak to 179 samples after the R-peak, we have a pre-classified ECG record of 300 samples. This sequence occupies 333 ms prior to R-peak sample and 497 ms after the R-peak sample. In each simulation, ECG beats are randomly selected from those six files so that the dataset is composed of 1,000 beats for each type.

B. Signal Preprocessing

The 300-sample sequence is then decomposed using the

Daubechies 4 (DB4) wavelet algorithm. To extract the feature of the ECG signal, we select the lower-frequency component for our consideration, since the higher-frequency component is usually a noise component. In this paper, with the effect of down-sampling, we take two levels of DB4 so that the sequence of 360-sample ECG is reduced to 77-sample wavelet coefficients of lower-frequency components. Fig. 2 shows the ECG signal of NORMAL arrhythmia and its wavelet coefficients. The training and validating datasets are randomly selected from the whole dataset so that there are 300 (100 for each type) and 2,700 (900 for each type) wavelet coefficient sequences for training and validating sets of data, respectively.

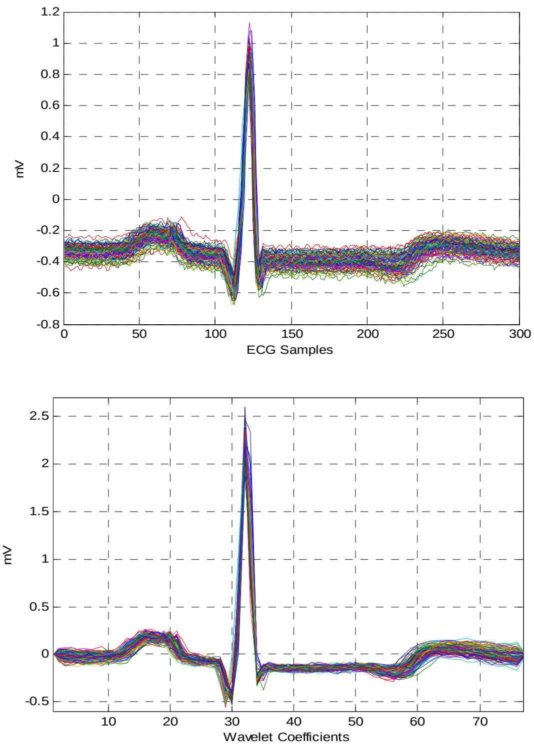

Fig. 2. The ECG signal of normal arrhythmia (top) and its two-pass wavelet coefficients (bottom).

C. Feature Extraction

The autoregressive or AR signal modeling method models the arbitrary signal as the output of a linear system driven by whitenoise with zero mean and unknown variance [11]-[12]. AR models have been found to provide sufficiently accurate representation for many different types of signal in many different applications. The AR model is defined by following

$$v(k) = \sum_{i=2}^{P+1} a_k v(k-i+1) + n(k) \qquad (1)$$

where $v(k)$ is the arbitrary time series, $n(k)$ is zero-mean whitenoise, $a_i{}'$s are the AR coefficients and P is the AR order.

By selecting from the 4[th]- to the 9[th]-order AR models, the AR coefficients of each type are shown in Table I. Fig. 3 shows the distribution of the 6[th]-order AR coefficients where the solid, short-dash and long-dash lines connect the mean values of all coefficients a_1 to a_6 for the Normal, PVC and APC, respectively.

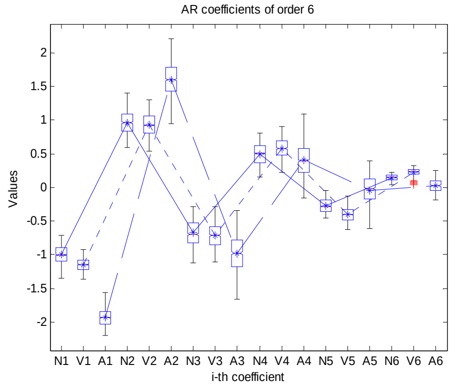

Fig. 3. The distribution of the 6[th]-order AR coefficients of arrhythmia types: Normal (N), PVC (V) and APC (A).

D. Signal Classification

The support vector machine or SVM [13] separates an input $x \in \mathrm{R}^d$ into two classes. A decision function of SVM separates two classes by $f(x) > 0$ or $f(x) < 0$. The training data set which is used in training phase is $\{x_i, y_i\}$ for $i = 1, \ldots l$ where $x_i \in \mathrm{R}^d$ is the input pattern for the i^{th} sample and $y_i \in \{-1, +1\}$ is the class label. Support Vector Classifiers map x_i into some new space of higher dimensionality which depends on a nonlinear function $\phi(x)$ and look for a hyperplane in that new space. The separating hyperplane is optimized by maximizing the margin. Therefore, SVM can be solved as the following quadratic programming problem,

$$\max_{\alpha_i} \left\{ \sum_{i=1}^{l} \alpha_i - \frac{1}{2} \sum \sum \alpha_i \alpha_j y_i y_j K(x_i, x_j) \right\} \qquad (2)$$

$$\text{Subject to } 0 \le \alpha_i \le \mathrm{C} \text{ and } \sum_{i=1}^{l} \alpha_i\, y_i = 0$$

where C is a parameter to be chosen by user, a larger C corresponding to a higher penalty assigned to errors, and $\alpha \ge 0$ are Lagrange multipliers. When the optimization problem has solved, system provides many which are the required Support vector.

Note that kernel function $K(x_i,\, y_i) = \phi^{\mathrm{T}}(x_i)\phi(x_i)$ where $\phi(\cdot)$ is a nonlinear operator mapping input vector $x \in \mathbf{R}^d$ to a higher dimensional space. In this work, we choose the radius basis function (RBF) as the kernel function since it provides best performance compared to other well-known kernels to our application.

III. Experimental Results

Classification consists of two steps: training and validation. In the training phase, SVM receives features as input. In this investigation, features extracted from heartbeat are represented by the AR coefficients of orders from 4 to 9. The system under investigation is configured as follows. The dataset is categorized into two classes as NORMAL beats and non-NORMAL beats where training and validating sets are randomly selected as described above. The classification is performed by using the coefficients of AR orders from 4 to 9 as features input to the SVM.

TABLE I: MEAN VALUES OF THE AR COEFFICIENTS OF ORDERS 4 TO 9 FOR NORMAL(N), PVC(V) AND APC(A)

	order	type	a_1	a_2	a_3	a_4	a_5	a_6	a_7	a_8	a_9	
	4	N	-1.075	0.935	-0.557	0.266						
	4	V	-1.073	0.784	-0.470	0.233						
	4	A	-1.923	1.585	-0.963	0.384						
	5	N	-1.110	1.009	-0.678	0.401	-0.122					
	5	V	-1.111	0.860	-0.593	0.398	-0.154					
	5	A	-1.924	1.586	-0.965	0.381	0.003					
	6	N	-1.133	1.084	-0.801	0.578	-0.313	0.165				
	6	V	-1.149	0.957	-0.738	0.603	-0.411	0.230				
	6	A	-1.926	1.599	-0.992	0.423	-0.046	0.026				
	7	N	-1.143	1.104	-0.835	0.623	-0.370	0.220	-0.047			
	7	V	-1.168	0.990	-0.786	0.662	-0.484	0.313	-0.070			
	7	A	-1.937	1.618	-1.057	0.560	-0.265	0.287	-0.134			
	8	N	-1.150	1.135	-0.887	0.706	-0.477	0.356	-0.183	0.114		
	8	V	-1.176	1.020	-0.832	0.725	-0.558	0.406	-0.176	0.088		
	8	A	-1.954	1.656	-1.092	0.611	-0.352	0.419	-0.288	0.077		
	9	N	-1.154	1.142	-0.898	0.721	-0.496	0.376	-0.204	0.132	-0.014	
	9	V	-1.184	1.036	-0.865	0.769	-0.614	0.470	-0.252	0.173	-0.070	
	9	A	-1.966	1.692	-1.139	0.646	-0.402	0.506	-0.416	0.223	-0.073	

TABLE II: SVM CLASSIFICATION RESULTS OF ONE-AGAINST-THE-REST APPROACH FOR EACH TYPE OF ARRHYTHMIA AT DIFFERENT AR ORDERS

		One against the rest SVM results (Accuracy %)						
	Type	order						
		4	5	6	7	8	9	
	N	81.28	87.14	89.81	89.54	89.29	89.60	
	V	81.01	86.96	89.91	90.00	89.21	88.56	
	A	97.88	97.19	97.08	97.00	96.71	96.22	

TABLE III: CLASSIFICATION RESULTS OF STEP 1 (APC AGAINST NON-APC), STEP 2 (PVC AGAINST N) AND OVERALL PROCESS

		Proposed method results (Accuracy %)						
	Type	order						
		4	5	6	7	8	9	
	APC against non-APC	97.89	97.14	97.11	96.99	96.69	96.31	
	PVC against N	72.45	82.08	85.97	84.60	85.76	85.44	
	Overall	90.26	93.18	94.40	93.88	94.10	93.90	

The similar procedure is applied to dataset so that there are the classifications of PVC beats versus non-PVC beats and APC beats versus non-APC beats. Fig. 4 shows the performance of SVM classifications of one-against-the-rest trial where N_i, V_i and A_i represent the i^{th} order used in AR modeling of N-, PVC- and APC-against-the-rest approach, respectively. As one can see from Table II, the classification of APC-against-the-rest has the best performance.

Let us then configure the proposed method as shown in Fig. 5 at which the 1st step of feature classification, the APC-against-non-APC classification is applied to the validating features. The 2nd step is the classification of the non-APC group by the PVC-against-N classifier.

Fig. 6 shows the performance of SVM classifier where the long dash, short dash and solid lines connect the mean values of accuracy at each different AR order of step1 (APC-against-non-APC), step 2 (PVC-against-N) and overall classification, respectively. Table III also shows the same result in numbers.

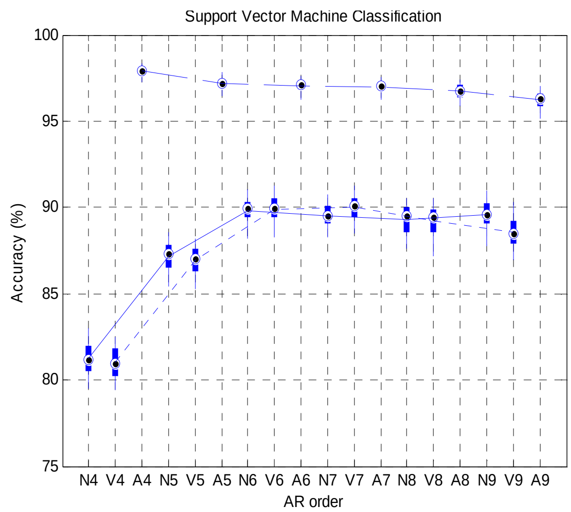

Fig. 4. The accuracies of SVM classification by one-against-the-rest trial using AR coefficients of orders from forth to ninth as features.

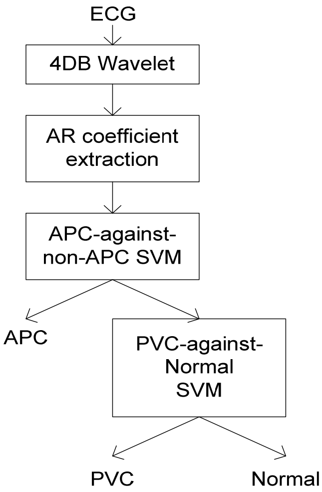

Fig. 5. The proposed classification workflow.

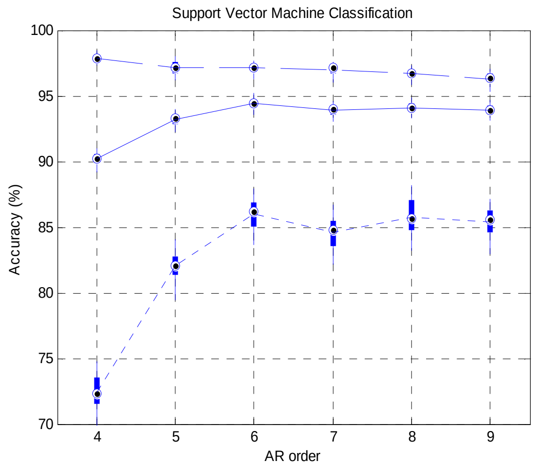

Fig. 6. The accuracies of SVM classification by proposed method using AR coefficients of orders from forth to ninth as features.

IV. Discussion and Conclusion

This paper presents the support vector machine (SVM) classification of three different arrhythmias including normal, PVC and APC whose features are extracted by applying autoregressive (AR) signal modeling. The classification performances were investigated at the different levels of AR orders from 4 to 9. The best performance of overall accuracy is obtained when the classification features are the 6th-order AR coefficients. Moreover, with the same order of AR model, the APC-against-non-APC classification (97.11%) outperforms the specific arrhythmia classification (96.4%) reported in the literature [14]. It clearly indicates that the lower AR order as if selected will provide the comparable performance since APC arrhythmia has the distinguish values of AR coefficient compared to those of normal and PVC coefficients. While the selected arrhythmias have most similarity of wave shape compared to other arrhythmias, the author is confident that this approach will provide comparable performance to other types of arrhythmia as well.

References

[1] P. Chazal and R. B. Reilly, "Automatic classification of ECG betas using waveform shape and heart beat interal features," *International Conference on Acoustics*, Speech and Signal Processing (ICASSP'03), vol. 2, pp. 269-272, 2003.

[2] O. T. Inan, L. Giovangrandi, and G. T. A. Kovacs, "Robust Neural-Network Based Classification of Premature Ventricular Contractions Using Wavelet Transform and Timing Interval Features," *IEEE Transactions on Biomedical Engineering*, Dec. 2006, vol. 53, pp. 2507-2515.

[3] M. H. Kadbi, J. Hashemi, H. R. Mohseni, and A. Maghsoudi, "Classification of ECG Arrhythmias Based on Statistical and Time-Frequency Features," *Advances in Medical, Signal and Information Processing*, 2006. MEDSIP 2006. IET 3rd Intr. Conf., July 2006, pp. 1-4.

[4] Q. Zhao and L. Zhang, "ECG Feature Extraction and Classification Using Wavelet," *International Conference on Neural Networks and Brain*, 2005, (ICNN&B '05), vol. 2, pp. 1089- 1092, 2005.

[5] D. Ge, N. Srinivasan, and S. M. Krishnan, "Cardiac Arrhythmia Classification Using Autoregressive Modeling," *Bio Medical Engineering*, 2002,

[6] Y. H. Hu, S. Palreddy, and W. J. Tompkins, "A Patient-Adaptable ECG Beat Classifier Using a Mixture of Experts Approach," *IEEE Transactions on Biomedical Engineering*, vol. 44, pp. 891-900, 1997.

[7] I. Atsushi, M. Hwa, A. Hassankhani, T. Liu, and S. M. Narayan, "Abnormal Heart Rate Turbulence Predicts the Initiative of Ventricular Arrhythmias," *Pacing Clinical Electrophysiology*, vol. 11, pp. 1189-97, Nov. 28, 2005.

[8] H. J. L. Marriott, N. L. Schwartz, and H. H. Bix, "Ventricular Fusion Beats," *Circulation*, vol. 26, pp. 880-884, 1962.

[9] C. Thanawattano and S. T. A. Ram, "ECG classification using modified support vector machine," *JICT2007*, Vientiane, Lao PDR, 2007

[10] R. Mark and G. Moody. MIT-BIH Arrhythmia Database. [Online]. Available: http://ecg.mit.edu/ dbinfo.html.

[11] S. L. Marple, "Digital spectral analysis with applications," *Prentice Hall*, Englewood Cliffs, New Jersey 1987

[12] L. Ljung, "System Identification: Theory for the user," *Prentice Hall*, Englewood Cliffs, New Jersey, 1999

[13] V. N. Vapnik, "The Nature of Statistical Learning Theory," *2nd ed.*, Springer-Verlag, New York, 1999

[14] D. Ge *et al*, Cardiac arrhythmia classification using autoregressive modeling. BioMedical Engineering. [Online]. Available: http://www.biomedical-engineering-online.com/content/1/1/5

Chusak Thanawattano was born in Bangkok, Thailand, in 1973. He received the electrical engineering degree from Chiang Mai University in 1996. He received M.S. and Ph.D. in electrical engineering degrees from Vanderbilt University, TN, in 2001 and 2002, respectively.

Since 2002, he has been a researcher at National Electronic and Computer Technology Center (NECTEC), Pathumthani, Thailand. Since 2008, he has been a researcher of the biomedical signal processing laboratory at the same center. His research interests are in biomedical signal processing, biomedical instruments and embedded system. His specific current interests include tremor analysis for Parkinson's disease, sleep disorder signal acquisition and analysis, ECG wireless monitoring and classification.

Thaweesak Yingthawornsuk was born in Bangkok, Thailand, in 1970. He received the electrical engineering degree from King Mongkut's Institute of Technology Thonburi in 1993. He received M.S. and Ph.D. in electrical engineering degrees from Vanderbilt University, TN, in 2003 and 2007, respectively. Since 2007, he has been a lecturer at King Mongkut's University of Technology Thonburi, Bangkok, Thailand. Since 2008, he has served as a Chairman of Media Technology Program and Head of Division of Biomedical Media Technology at the same university. His research interests are in speech processing, biomedical signal processing, physiological system identification, modeling and classification. Currently his ongoing research involves prediction of psychiatric related disorder based on applied speech processing techniques, acoustical parameter modeling and classification.