SVM Classification for Discriminating Cardiovascular Disease Patients from Non-cardiovascular Disease Controls Using Pulse Waveform Variability Analysis*

Kuanquan Wang[1], Lu Wang[1], Dianhui Wang[2], and Lisheng Xu[1]

[1] Department of Computer Science and Engineering,
Harbin Institute of Technology (HIT), Harbin, 150001 China
{wangkq, wanglu, xulisheng}@hit.edu.cn
[2] Department of Computer Science and Computer Engineering,
La Trobe University, Melbourne, VIC 3086, Australia
csdhwang@ieee.org

Abstract. This paper analyzes the variability of pulse waveforms by means of approximate entropy (*ApEn*) and classifies three group objects using support vector machines (SVM). The subjects were divided into three groups according to their cardiovascular conditions. Firstly, we employed *ApEn* to analyze three groups' pulse morphology variability (PMV). The pulse waveform's *ApEn* of a patient with cardiovascular disease tends to have a smaller value and its variation's spectral contents differ greatly during different cardiovascular conditions. Then, we applied a SVM to discriminate cardiovascular disease patients from non-cardiovascular disease controls. The specificity and sensitivity for clinical diagnosis of cardiovascular system is 85% and 93% respectively. The proposed techniques in this paper, from a long-term PMV analysis viewpoint, can be applied to a further research on cardiovascular system.

1 Introduction

More and more noninvasive measurements of physiological signals, such as ECG, heart sound, wrist pulse waveform, can be acquired for the assessment of physical condition. Among these methods, the ECG provides information about the electrical activity of the heart [1], while the wrist pulse waveform affords the information on the pressure variation in the wrist vessel. Various civilizations in the past have used arterial pulse as a guide to diagnose and treat various diseases.

The Chinese art of pulse feeling, which is still being practiced, has more than 2,000 years of history. According to traditional Chinese pulse diagnosis, the pulse not only can deduce the positions and degree of pathological changes, but is also a convenient, inexpensive, painless, and noninvasive method promoteded by the U.N. [2, 3]. Recording and analyzing the pressure wave in the radial artery of the wrist provide a non-

* Supported by the National Natural Science Foundation of China under Grant No.90209020.

G.I. Webb and Xinghuo Yu (Eds.): AI 2004, LNAI 3339, pp. 109–119, 2004.

invasive measure of the arterial pressure wave in proximal aorta. The radial artery pulse wave can reveal central systolic, diastolic and mean arterial pressures, as well as supply an assessment of arterial wave reflection, which is closely related to cardiovascular condition and the degree of stiffness of arteries. Recently, increasingly numbers of western medicine researchers have begun to pay more attention to pulse diagnosis [4-5].

Pulse waveform is analyzed usually by traditional time and frequency domain methods. Having analyzed the pulse waveform with the conventional methods, we find that some dynamic characters of the pulse waveform are undiscovered [6, 7]. Few papers on pulse waveform's nonlinear analysis can be found [8]. Currently, a number of nonlinear methods have been recently developed to quantify the dynamics of physiological signals such as ECG, EEG and so on. These have achieved some meaningful results that the conventional statistics cannot achieve [9]. Consequently, we investigate the pulse's variability through nonlinear methods.

There are many methods that can disclose the dynamic characters of physiological signal, such as K-S entropy, the largest Lyapunov exponent, approximate entropy, coarse-grained entropy and so on. However, K-S entropy and largest Lyapunov exponent assume that the time series have enough length. It appears that ApEn has potential wide spread utility for practical data analysis and clinical application due to its five salient features [10]. Furthermore, the *ApEn* can be applied in both deterministic and stochastic processes. At present, whether pulse waveform's nature is deterministic chaos or not has not been proved yet. Therefore we employ the *ApEn* to disclose some clinical value of pulse variability [11].

This paper applies SVM to discriminate cardiovascular disease patients from non-cardiovascular controls. The technique of SVM, developed by Vapnik, was proposed essentially for classification problems of two classes. SVM use geometric properties to exactly calculate the optima separating hyper plane directly from the training data [12-14]. Based on structure risk minimum principal, SVM can efficiently solve the learning problem, with the strengths of good generalization and correct classification. It is important to emphasize that SVM have been employed in a number of applications [15]. However, few of them belong to the bioengineering field, and in particular to pulse waveform variability discrimination.

In Section 2, the long-term pulse data collection and preprocessing are stated firstly. Then the *ApEn*s analysis of long-term pulse and their corresponding experimental results are presented in Section 3. Having extracted 12 features on pulse variability, we apply a SVM classifier to discriminate cardiovascular disease patients from non-cardiovascular controls in this section. Section 4 draws our conclusions.

2 Material and Methods

This section describes the data collection and our analysis methodology.

2.1 Study Protocol and Data Collection

In this study, all the pulse data are acquired by our pulse monitoring and diagnosis system, illustrated in Fig. 1 [7].

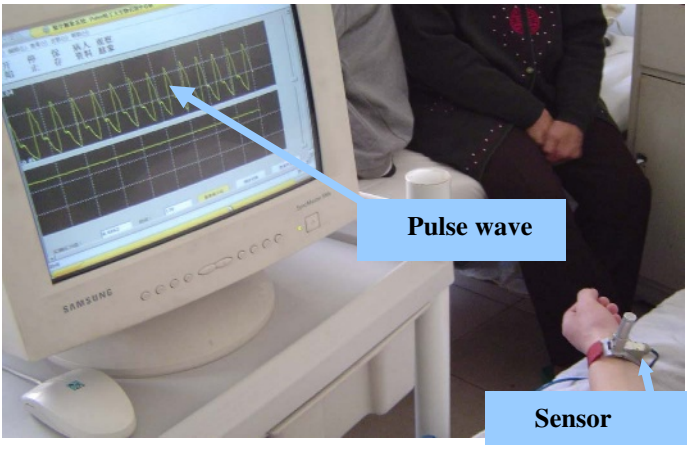

Fig. 1. Our pulse acquisition system

Pulse waveform recordings are acquired from 90 volunteers. Three groups are studied, each including 30 subjects, matched for age and gender. All of them are examined by ultrasonic test. They are confirmed to be without neural system problems.

- **Group1** is 30 patients with cardiovascular disease (15 females and 15 males, age 60±12 years);
- **Group2** is 30 patients hospitalized for non-cardiac cause (15 females and 15 males, age 55±12 years);
- **Group3** contains 30 healthy subjects who are selected as control subjects matched for sex and age (15 females and 15 males, age 55±12 years). Those selected control subjects have no documented history of cardiovascular diseases and disorders, and have been examined by ultrasonic, X-ray examination and so on.

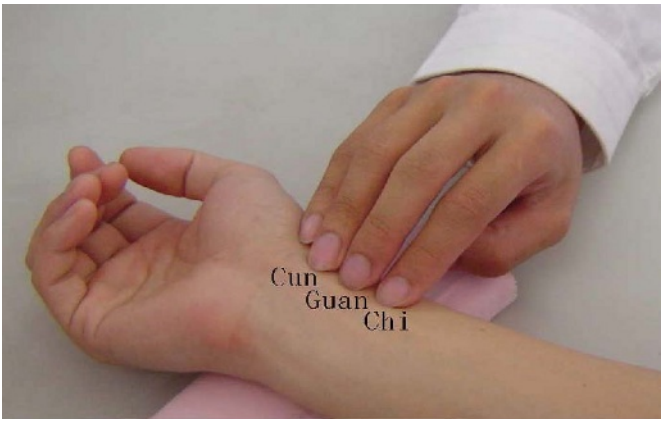

Fig. 2. The positions of *"Cun"*, *"Guan"*, *"Chi"*

The pulses of all subjects are acquired for 600 seconds long at the sampling rate of 100 Hz. Each subject was asked to relax for more than 5 minutes before pulse acquisition. According to the traditional Chinese pulse diagnosis, we can acquire pulse at the positions of *"Cun"*, *"Guan"*, *"Chi"*, which are demonstrated in Fig. 2. All of the subjects were lying on their backs during pulse acquisition. According to the theory of Traditional Chinese pulse diagnosis, the pulse in *"Cun"* position reflects the condition of the heart. As a result, we put our pulse sensor on the *"Cun"* positions of the subjects' left wrists to study the relationship between cardiovascular condition and pulse waveform variability.

2.2 Methods

We utilize an approximate entropy and SVM classifier techniques to analyze and classify wrist pulse waveform variability. The whole procedure is illustrated in Fig. 3. At first, we use the designed filter to remove the interference and baseline wander of pulse waveform. Then we segment the pulse waveform into 200 partitions and apply the approximate entropy to analyze the variability of pulse morphology. After that, we extract 12 features from the approximate entropies. Finally, we employ the SVM classifier to discriminate the cardiovascular disease patients from non-cardiovascular disease controls.

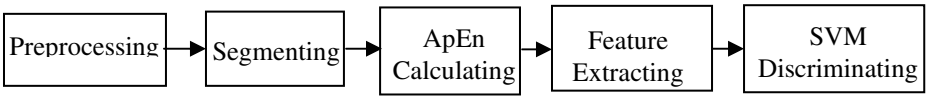

Fig. 3. The schematic figure on the procedure of pulse waveform

2.2.1 Pulse Waveform Preprocessing Based on Cascaded Adaptive Filter

The bandwidth of the acquiring system is with almost linear response from the 0.05Hz to 100Hz, causing no distortion of pulse waveform. However, distortion may arise from the subject's movement, respiration and so on. Thus, the baseline wander introduced in the acquisition process must be removed before computing the pulse waveform's *ApEn*. We apply the cascade adaptive filter as described in the paper [16] to remove this wander.

2.2.2 Waveform Variability Analysis Using ApEn

Over the past few decades, thanks to the advance of computer technology, the recording and storage of massive datasets of pulse waveform is possible. As a result, some nonlinear analysis methods can be used to extract useful clinical information from pulse data.

Nonlinear dynamical analysis is a powerful approach to understand biological system. Pincus introduced *ApEn* as a set of measures of system complexity, which has easily been applied to clinical cardiovascular and other time series. *ApEn* may contain the information that is neither visually apparent nor extractable with conventional methods of analysis.

ApEn is a measure of complexity and regularity. For instance, a small *ApEn* means a high degree of regularity. The approximate entropy, *ApEn(m, r, N)*, can be estimated as a function of the parameters *m*, *r* and *N*, where *m* is the dimension to which the signal will be expanded, *r* is the threshold and *N* is the length of the signal to be analyzed. Both theoretical analysis and clinical applications conclude that when m=2 or 3, and *r* is between 10% and 25% of the standard derivation of the data to be analyzed, the *ApEn(m,r,N)* produces good statistical validity. In this paper, we use m=2, r=0.2, N=300 (that means every segment includes 300 sampling points).

The procedure of pulse morphology variability (PMV) analysis is as follows:

- Dividing each 10 minutes pulse recording into 200 segments. Each segment contains data corresponding to a 3-second portion of the recording (300 sampling points);
- Calculating *ApEn* of every segment and obtaining 200 *ApEn*s for each subject.

Having applied the *ApEn* for PMV analysis of three-groups, we illustrate the *ApEn* mean values of three groups in Fig. 4. The y-coordinate is the average of every subject's 200 *ApEn*s. Each group contains 30 subjects and their *ApEn* Means all vary from 0.08 to 0.25. On average, the *ApEn* Means of **Group1** are smaller than **Group2** and **Group3**'s. But the *ApEn* Means of three groups don't have significant difference.

The *ApEn* averages of PMV don't have significant difference, but the fluctuation of their *ApEn* consequences differs notably. In Fig. 5, ***ApEn1***, ***ApEn2*** and ***ApEn3*** is the typical *ApEns* of subject in **Group1**, **Group2** and **Group3** respectively. The y-axis is the value of *ApEn* and the x-axis is the segment's sequence number. From Fig. 5, we can find that the ***ApEn1*** fluctuates faster and more regularly than ***ApEn2*** and ***ApEn3***. This means that the healthier the person's cardiovascular system is, the more complex his PMV is.

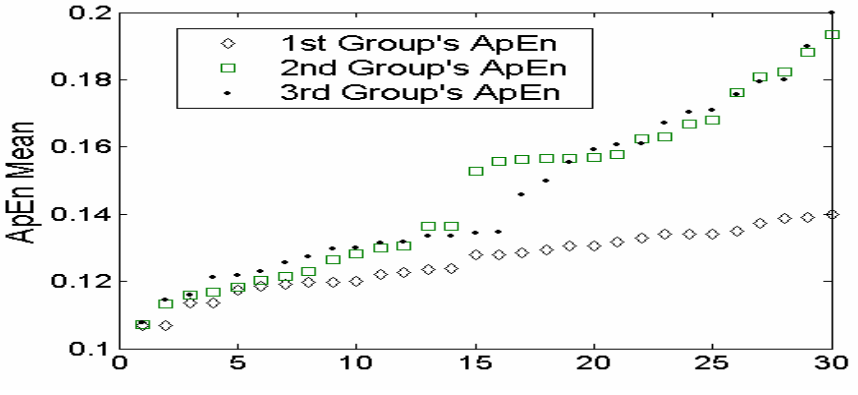

Fig. 4. The comparison of three groups' *ApEn*s averages. Each group contains 30 persons. Each person's 10 minutes pulse waveform was portioned into 200 segments. Each point stands for the average of a person's 200 *ApEn*s

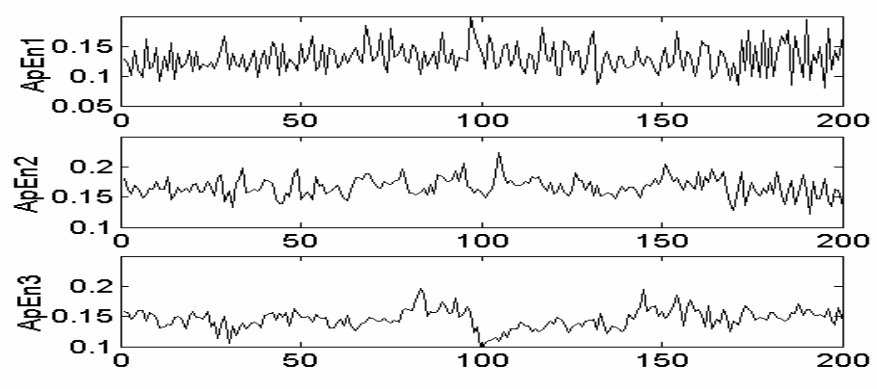

Fig. 5. The comparison of *ApEn*s

2.2.3 Pulse Morphology Variability Features Extraction

In the above part we analyze the PMV of the three groups and find that PMV has notable clinical value to differentiate cardiovascular conditions.

From the spectral point of view, we can discover some useful relationship between PMV and the cardiovascular system. Fig. 7 illustrates the power spectrum of PMV. All of them are computed from the 200 *ApEn*s of 10 minutes' pulse waveforms. The x-axis is the Nyquist frequency and the y-axis is the amplitude of its spectrum. The first row ***PSD1*** is the spectrum of one patient in **Group1**; the second ***PSD2*** is the spectrum of one patient in **Group2**; the third row ***PSD3*** is the spectrum of **Group3**'s. We can find that the healthy person's *ApEn* has more low frequency content as shown in ***PSD3***. The ***PSD1*** has more high frequency content than ***PSD2*** and ***PSD3***.

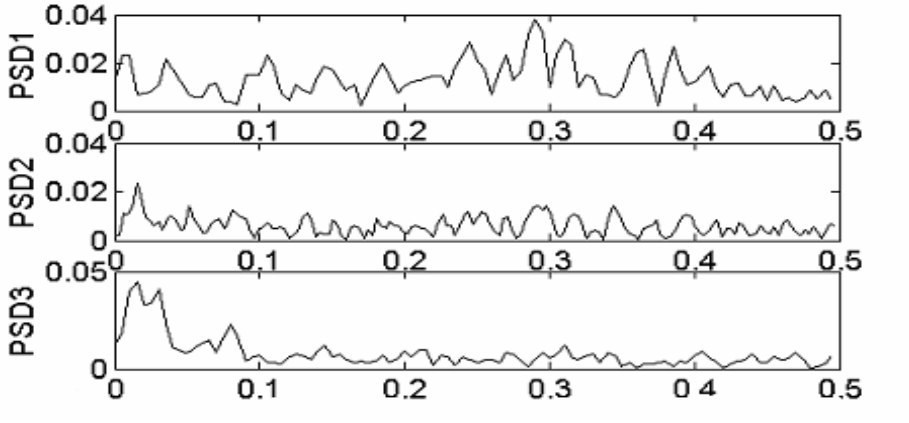

Fig. 6. The comparison on the spectral distribution of three groups' ApEn

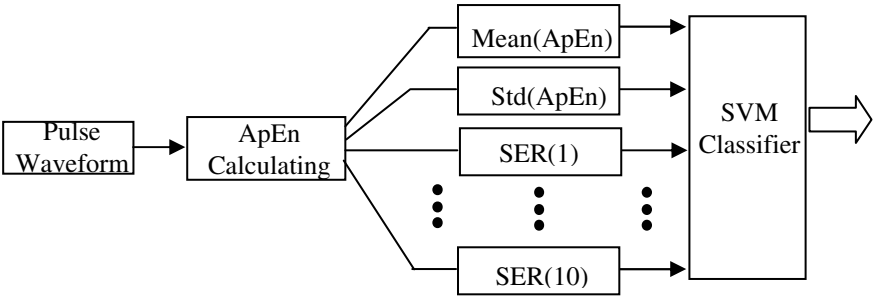

Fig. 7. The schematic figure of features extraction

In this part, we will extract some features from the PMV ApEns. Fig. 7 lists the features such as the mean, standard derivation of the ApEns and spectral energy ratio (***SER***) of ApEns. This paper partitions the power spectrum of the *ApEn* into 10 equi-distant segments as $[0 \sim 0.05f_s, 0.05 \sim 0.1f_s, \cdots, 0.4 \sim 0.45f_s, 0.45 \sim 0.5f_s]$. Then we can get their spectral rates of those 10 segments. The PMV's ***SER***s are computed as shown in Formula (1).

$$SER(i) = \sum_{f_i=(i-1)*0.05f_s}^{i*0.05f_s} A^2(f_i) \Big/ \sum_{f_j=0}^{0.5f_s} A^2(f_j) \text{ , i=1,…, 10.} \quad (1)$$

where f_i is the spectral and $A(f_i)$ is its corresponding amplitude.

2.2.4 SVM Classifiers

Support Vector Machines were invented by Vapnik. They are learning machines that can create functions from a set of labeled training data. For classification, SVMs operate by finding a hypersurface in the space of possible inputs. This hypersurface will attempt to split the positive examples from the negative examples. The split will be chosen to have the largest distance from the hypersurface to the nearest of the positive and negative examples.

The discriminant equation of the SVM classifier is a function of kernel $k(\boldsymbol{x}_i, \boldsymbol{x})$ and is given by:

$$D(x) = sign(\sum_{i=1}^{N_{sv}} \alpha_i y_i k(X_i, X) + b) . \quad (2)$$

where X_i are the support vectors, N_{sv} is the number of support vectors, α_i *is* the weight parameters, *b is* the biased parameter, and $y \in \{-1,+1\}$ depending on the class. In the present study the two degree non-homogeneous polynomial function was used for the linear kernel, given by $K(x, y) = x^T \cdot y$, or Polynomial Kernel at the degree of two $K(x, y) = (x \cdot y + 1)^d$ with d = 2, resulting in the discriminant function of the SVM classifier.

3 PMV's SVM Classifier Discrimination Results

The SVM classifier has better generalization ability than neural network and other classifiers, especially for small training data sets. In this study, we apply a SVM to classify **Group1** with **Group2**, and **Group1** with **Group3**. As listed in Table1, we name the **Group1** as Cardiovascular Disease Group, **Group2** and **Group3** as Non-Cardiovascular Disease Group. We name the subjects who are classified into cardiovascular patients as positive and those subjects who are classified into non-cardiovascular person as negative. If a subject who was labeled as cardiovascular patient is indeed so afflicted, this situation is referred to as a true positive (TP); a non-cardiovascular disease subject erroneously labeled as cardiovascular patient is referred to as a false positive (FP). We define negative outcomes that are true (TN) and false (FN) in an analogous manner [17-18]. We calculate some characters according to Formulas (3) - (5).

$$specificit\ y = \frac{TN}{TN + FP} \ , \tag{3}$$

$$sensitivity = \frac{TP}{TP + FN} \ , \tag{4}$$

$$accuracy = \frac{TN + TP}{ALL} \ . \tag{5}$$

The results list as Table2. The specificity, sensitivity, accuracy of **Group1/Group2** is 85%, 93%, and 89% respectively. They are slightly less than that of **Group1/Group3**.

Table 1. Definitions on this discrimination between these Groups

	Non-ardiovascular Disease	Cardiovascular Disease
"Non-Cardiovascular Disease"	TN	FN
"Cardiovascular Disease"	FP	TP

Table 2. The discrimination results of three Groups

	Specificity	Sensitivity	Accuracy
Group1/ Group2	85%	93%	89%
Group1/ Group3	90%	93%	92%

As the 12 dimensional features cannot be illustrated, of the 12 features, we demonstrate only two dimensional features: ***mean(ApEn)*** and ***SER(1)*** in Figs. 8, 9 and 10. Fig. 8 is SVM classifier's result to classify **Group1** and **Group2**. Fig. 9 is the linear kernel SVM classifier's result to classify **Group1** and **Group3**. We can find that during the two features, **Group1** can be discriminated from **Group2** and **Group3** with high accuracy. Fig. 10 is SVM classifier's result to classify **Group2** and **Group3**. In Fig. 10, **Group2** and **Group3** cannot be differentiated with each other: all the vectors are support vectors. These results demonstrate that the variability of pulse waveform morphology has a powerful ability in discriminating the cardiovascular disease patients from the non-cardiovascular controls.

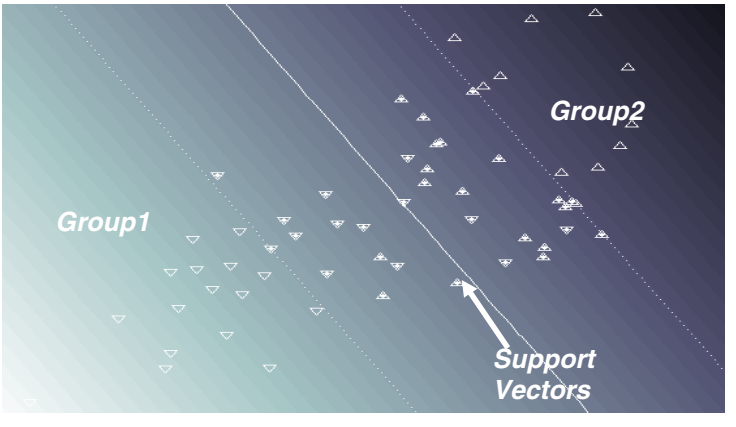

Fig. 8. The classification of **Group1/Group2** by SVM classifier

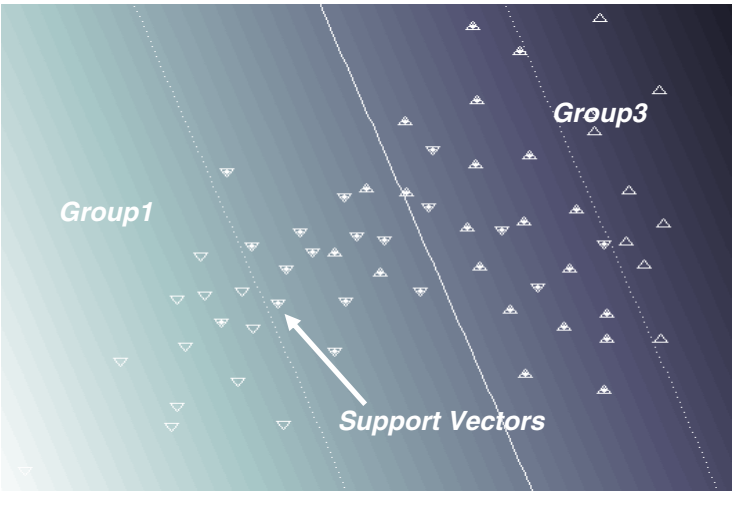

Fig. 9. The classification of **Group1/Group3** by SVM classifier

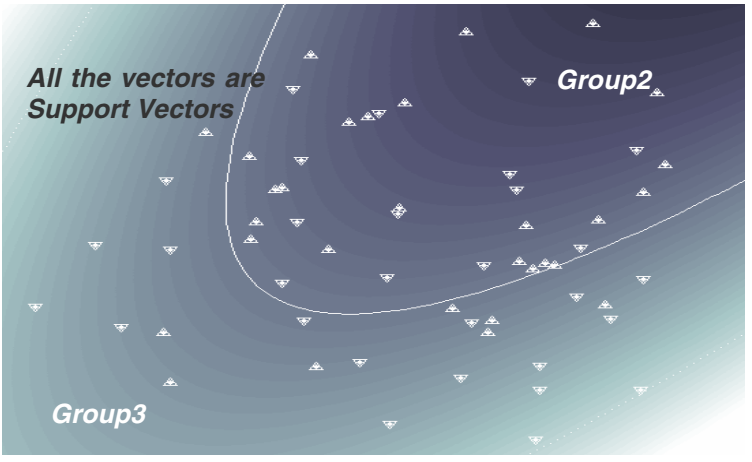

Fig. 10. The classification of **Group2/Group3** by SVM classifier (Polynomial Kernel at the degree of two)

4 Conclusions

This paper studies the variability of long-term pulse waveform and analyzes its clinical value for cardiovascular systems. Analysis of the dynamic behavior of pulse signal has opened up a new approach towards the assessment of normal and pathological cardiovascular behavior.

This paper also presents PMV's spectral energy ratio for differentiating person's cardiovascular condition. The results conform that the PMV can be used to differentiate the subjects in different cardiovascular condition. Using SVM to construct classifiers the accuracy of **Group1** to **Group2** is 89% and the accuracy of **Group1** to **Group3** is 92%. For the purpose of probing the mechanism of manifestations of the pulse, further work needs to be performed to quantitatively analyze cardiovascular system's behavior.

Acknowledgements

This study is supported by National Natural Science Foundation of China under Grant No.90209020. We thank all of the volunteers for providing us with the invaluable pulse data.

References

1. T.C. Joseph, Guide to ECG analysis, Lippincott Williams & Wilkins Press, (2002)
2. Y.Z. Feng, Chinese Journal of Biomedical Engineering, (1983)1(1)
3. W.A. Lu, Y.Y.Lin Wang, and W.K. Wang, "Pulse analysis of patients with sever liver problems," IEEE Engineering in Medicine and Biology, Jan/Feb, (1999) 73-75

4. Rourke, et al. "Pulse wave analysis", Br J Clin Pharmacol, Vol 51, (2001) 507-522
5. M. Aritomo, Y. Yonezawa, "A wrist-mounted activity and pulse recording system," Proceedings of the First Joint BMES/EMBS Conference Serving Humanity, Advancing Technology, (1999) 693
6. K.Q. Wang, L.S. Xu and D. Zhang, "TCPD based pulse monitoring and analyzing," ICMLC2002, Nov.3-7, Beijing, (2002)
7. L.S. Xu, K.Q. Wang and D. Zhang, "Modern researches on pulse waveform of TCPD," 2002 International Conference on Communications Circuits and Systems and West Sino Expositions Proceedings, Chengdu, China, (2002) 1073-1078
8. Yoshio Maniwa, Tadashi Iokibe, Masaya Koyama, Motoki Yamamoto, Shoichi Ohta, "The Application of Pulse Wave Chaos in Clinical Medicine", 17 Fuzzy System Symposium, Chiba, Sept. 5-7, (2001)
9. J.E. Naschitz, R. Itzhak, N. Shaviv and et al, "Assessment of cardiovascular reactivity by fractal and recurrence quantification analysis of heart rate and pulse transit time," J Hum Hypertension, Vol. 17, N. 2,(2 003)111-118
10. S. Pincus, "Approximate entropy (ApEn) as a complexity measure," Chaos 5, (1995) 110-117
11. K.Q. Wang, L.S. Xu and D. Zhang, "Approximate entropy based pulse variability analysis," Proceedings of the IEEE Symposium on Computer-Based Medical Systems, (2003) 236-241
12. Cristianini, N., Shawe-Taylor, J., An Introduction to Support Vector Machines, Cambridge University Press (2000)
13. S. Gunn. "Support vector machines for classification and regression", ISIS technical report, Image Speech & Intelligent Systems Group, University of Southampton (1997)
14. BE Boser, IM Guyon, and VN Vapnik. "A training algorithm for optimal margin classifiers", Proceedings of the 5th Annual ACM Workshop on Computational Learning Theory, ACM Press, (1992) 144–152
15. M. Z. Rahman, SML Kabir and J. Kamruzzaman, "Design and implementation of an SVM-based computer classification system for discriminating depressive patients from healthy controls using the P600 component of ERP signal", Comput Methods Programs Biomed, Jul; 75(1): (2004) 11-22
16. L.S. Xu, K.Q. Wang, D. Zhang, "Adaptive baseline wander removal in the pulse waveform", IEEE Proceeding of CBMS2002 International Conference, June, (2002) 143-148
17. M. Akay, Nonlinear Biomedical Signal Processing, New York, IEEE Press (2000)
18. X.H. Zhou, A.O. Nancy, and K.M. Donna, Statistical Methods in Diagnostic Medicine, Wiley-Interscience publication (2002)