Distinguishing Patients with Gastritis and Cholecystitis from the Healthy by Analyzing Wrist Radial Arterial Doppler Blood Flow Signals

Xiaorui Jiang[a], Dongyu Zhang[b], Kuanquan Wang[b], Wangmeng Zuo[b]
[a]*Key Lab of Intelligent Information Processing, Institute of Computing Technology, Chinese Academy of Sciences*
[b]*Department of Computer Science and Technology, Harbin Institute of Technology*
xiaoruijiang@kg.ict.ac.cn, {zhangdy, kqwang}@hit.edu.cn, cswmzuo@gmail.com

Abstract

This paper tries to fill the gap between Traditional Chinese Pulse Diagnosis (TCPD) and Doppler diagnosis by applying digital signal analysis and pattern classification techniques to wrist radial arterial Doppler blood flow signals. Doppler blood flows signals (DBFS) of patients with cholecystitis, gastritis and healthy people are classified by L2-soft margin SVM and 5 linear classifiers using the proposed feature - piecewise axially integrated bispectra (PAIB). A 5-fold cross validation is used for performance evaluation. The classification accuracies between either two groups of subjects are greater than 93%. Gastritis can be recognized with higher accuracy than cholecystitis. Cholecystitis can be recognized with higher accuracy on left hand data than right. The findings in this paper partly conform to the theory of TCPD. Though the sample size is relatively small, we could still argue that the methods proposed here are effective and could serve as an assistive tool for TCPD.

1. Introduction

In TCPD, doctors put their fingers on human wrist to feel vascular pulsation so as to collect pathological information of a patient's overall body condition [13, 15, 8]. It is also believed in TCPD that pulse waveform and its time/frequency properties can reflect the pathological changes of internal organs [10, 11].

Although modern medicine has revealed the mechanics of circulation system and blood flow, and their relationships with pulse waveform generation, western researchers mainly focus on applications in diagnosing and preventing cardiovascular diseases such as coronary heart disease, hypertension, arteriosclerosis etc through analysis of the DBFS waveform [1, 5, 2, 9]. To integrate the viewpoints of TCPD and western medicine should not only provide a nonintrusive medical diagnosis method but also help make TCPD more objective, scientific and acceptable. Because of the inter-relationships between pulse pressure (captured by digital pulse detector), blood velocity (reflected by DBFS), and the elastic deformation of arterial wall (which generates the pulse contour studied in TCPD), DBFS could serve as such a media.

In this paper, piecewise axially integrated bispectra (PAIB) were extracted from wrist radial arterial DBFS. L2-soft margin SVM and five linear classifiers were used for discriminating between healthy people, and patients with gastritis and cholecystitis. Results show that analysis of DBFS could be an assistive tool for computer-based TCPD in recognizing internal organ diseases.

2. Materials and Methods

The whole routine of the proposed method is illustrated in Figure 1. We focus mainly on Feature Selection and Pattern Classification in this paper.

2.1. Data Acquisition

142 Doppler sonograms (i.e. Doppler flow-velocity images or Doppler pulse waveforms) were collected. Data of at least 10-beats long were recorded and at least 2 recordings were made on both hands, if

1051-4651/10 $26.00 © 2010 IEEE
DOI 10.1109/ICPR.2010.610

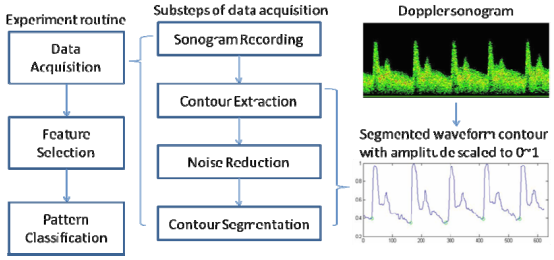


Figure 1: Block diagram for the routine of the proposed method

possible. Sample data were divided into three groups: Group H (19 healthy subjects), Group G (12 patients with gastritis) and Group C (14 cholecystitis patients) (see Table 1).

Table 1: Data set in experiments

Category	#Samples		
	Left Hand	Right Hand	Left and Right
Healthy (H)	18	18	19
Gastritis (G)	10	8	12
Cholecystitis (C)	11	11	14

DBFS should be filtered. In diastole there are many small burrs. They are mostly high frequency noises that spoil time-frequency property of DBFS. Soft-threshold wavelet packet filtering was used to smooth the signal [4]. The best wavelet packet bases were chosen based on Shannon entropy. "Db10" was used to decompose the signal. The adaptive threshold value selection rule is Rigrsure thresholding rule, a Stein's unbiased risk estimator. Only the coefficients of high frequency components were filtered (see Figure 2).

After noise reduction, the waveform contour was segmented into periods. The algorithm for finding the starting point of each period was that proposed in [6] with some adaptation to DBFS.

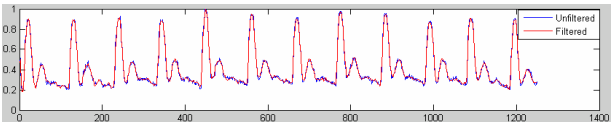


Figure 2: Example of wavelet packet filtering

2.2. Axially integrated bispectra

It is widely recognized that feature extraction using higher order statistics (cumulants or polyspectra) offers many advantages compared to traditional techniques based on second order statistics [14]:

(1) retaining both amplitude and phase information;
(2) being translation or shift invariant;
(3) being robust to additive Gaussian noise.

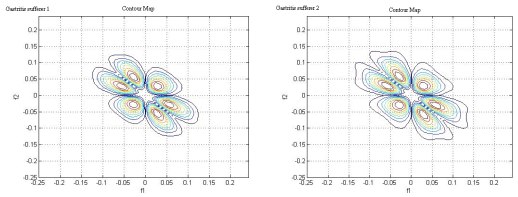
For signal $x(t)$, we define the unified 3rd order cumulant using samples $x(1)$, …, $x(N)$ in Eq. (1) where τ_k means time delay.

$$\hat{c}_{3x}(\tau_1,\tau_2)=\frac{1}{N}\sum_{n=1}^{N}x(n)x(n+\tau_1)x(n+\tau_2)\,. \tag{1}$$

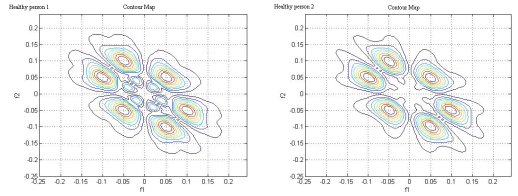
Bispectra $B_x(\omega_1,\omega_2)$ is thus defined as

$$B_x(\omega_1,\omega_2)=S_{3x}(\omega_1,\omega_2)=\sum_{\tau_1=-\infty}^{\infty}\sum_{\tau_2=-\infty}^{\infty}c_{3x}(\tau_1,\tau_2)e^{-j(\omega_1\tau_1+\omega_2\tau_2)}\,. \tag{2}$$

Bispectra can be represented by the contour map of $|B_x(\omega_1,\omega_2)|$ estimates (see Figure 3). The distinctions between healthy persons and both groups of patients lie in where the $|B_x(\omega_1,\omega_2)|$ culminates. In Figure 3, frequency points (values along X and Y axes) are multiples of f_s (sampling frequency), e.g. f_1= 0.15 means f_1=0.15*f_s. We use f and ω interchangeably in this paper. $|B_x(\omega_1,\omega_2)|$ of healthy persons culminate near (0.05*f_s,0.05*f_s), yet the peaks shift to around (0.025*f_s, 0.025*f_s) in the contour maps of patients with gastritis. The peak shift could be a strong evidence implying impacts of pathological changes and help differentiating the two categories of samples.



(a) Countour maps of 2 patients with gastritis



(b) Countour maps of 2 healthy subjects

Figure 3: Bispectral contour maps

For the convenience of computation, we define the feature vectors as follows. The first feature set is computed based on the subspectrum in the 1st and 4th quadrants. We find that the energy of bispectra mainly locates in the frequency band of 0~0.25*f_s, so we extract the subspectrum with f_1 ranging from 0 to 0.25*f_s and f_2 ranging from -0.25*f_s to 0.25*f_s. There are altogether 64×32 frequency points. Let $B(f_1, f_2)$ be the bispectra of signal, we define feature vector $\mathbf{F_1}$ as follows:

$$\mathbf{F_1} = [PAIB(i,j)]_{i=1,...,32;\ j=1,...,16}, \quad (3)$$

where

$$PAIB(i,j) = \sum_{k=1+(j-1)*4}^{j*4} B(i,k) \quad i = 1,...,32;\ j = 1,...,16. \quad (4)$$

Here *PAIB* stands for piecewise axially integrated bispectra [14]. $\mathbf{F_1}$ is a 512-dimensional vector.

For the second feature set the extracted subspectrum has f_1: 0~0.25*f_s and f_2: 0~0.25*f_s, which contains 32×32 frequency points. The feature vector $\mathbf{F_2}$ with a reduced dimension of 256 is constructed in similar way,

$$\mathbf{F_2} = [PAIB(i,j)]_{i=1,...,32;\ j=1,...,8}, \quad (5)$$

where

$$PAIB(i,j) = \sum_{k=1+(j-1)*4}^{j*4} B(i,k) \quad i = 1,...,32;\ j = 1,...,8. \quad (6)$$

$\mathbf{F_1}$ and $\mathbf{F_2}$ were computed as in Figure 4.

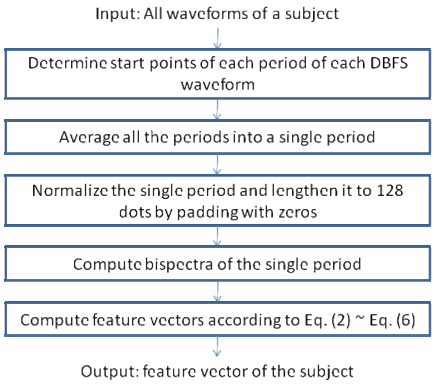


Figure 4: Steps for calculating *PAIB*

3. Results

In the experiments, six classifiers including Fisher linear discriminant (FLD), quadric-programming Fisher linear discriminant (FLD-QP), batch-mode perceptron (Perc), Kozinec's perceptron (Kozi), the K-nearest neighbor (KNN), and L2-soft support vector machine (SVM) [3] are used. The learning rate of the perceptron algorithm is η=0.1. Model parameters of SVM are C=10, σ=1.

We made a 5-fold cross-validation to reduce the generalized error rate and make the evaluation less biased [12]. The overall performance is the average of the results of all 5 iterations.

We denote BIS1 as the classification task between Group H and Group G, BIS2 as task between Group H and Group C, and BIS3 as task between Group G and Group C.

In Task BIS1, the L2-SVM got a 100% recognition rate on both left hand and right hand data. Most linear classifiers all got a high recognition rate greater than 89% (see Table 2). In general, the recognition rate was slightly greater on right hand data than left.

In task BIS2, the situation was just the other way. Except for PERC using $\mathbf{F_1}$, performances on left hand data were substantially greater than right. The highest recognition rate of 93% was obtained on left hand data using L2-SVM and $\mathbf{F_2}$ (see Table 3). Another observation is that all classifiers performed better using $\mathbf{F_2}$ than $\mathbf{F_1}$ or at least equally well using either $\mathbf{F_1}$ or $\mathbf{F_2}$.

In Task BIS3, we only used $\mathbf{F_2}$. G and C samples were classified with a high accuracy of about 95%, with FLD and PERC being exceptions. Left hand data had better performance than right hand data (see Table 4). In all three tasks, L2-SVM was the classifier with best performance.

4. Discussions and Conclusions

In this study, we investigated how to adopt signal processing and pattern classification techniques on DBFS to assist traditional Chinese medicine (TCM). In the tradition of TCM, symptoms rather than disease is the widely accepted concept. However we targeted at two digestive diseases, gastritis and cholecystitis here so that the results can be accepted and understood by a larger group of audience. Based on the results in this study, we argue the following two points. Firstly, pathological changes caused by certain diseases can be detected at wrist radial arterial, as is believed in TCM. Secondly, the features proposed in this paper were useful in detecting the pathological changes in the energy distribution of DBFS at wrist radial arterial. We now conclude with the following remarks:

1) $\mathbf{F_2}$ performed better than $\mathbf{F_1}$ in general speaking. Redundancy reduction is important.
2) The fact that using PCA to reduce the dimension of feature vector got a poor result (we omitted the details in this paper), implies that a redundancy of a certain degree may be profitable. And this leaves us concern of interest in our future work.
3) PAIB takes into account all the frequency pairs of f_1 and f_2 within a frequency band. It means that the shape of the amplitude map of bispectra is taken into consideration. More sophisticate methods for extracting geometric information from amplitude maps are of concern in our future work.
4) An interesting observation is that when healthy subjects and cholecystitis patients were being

classified, all the classifiers performed much better on the left hand data than right except PERC using $\mathbf{F_2}$. This conforms to the theory in TCPD which tells that diseases of liver and gallbladder should be diagnosed by radial arterial pulse waveform of left hand [13].

Table 2: Experiment results of Task BIS1

		Left Hand			Right Hand		
		H	C	Avg	H	C	Avg
FLD	$\mathbf{F_1}$	61.11	70	64.29	33.33	87.5	50
	$\mathbf{F_2}$	66.67	60	60.71	50	37.5	46.15
QFLD	$\mathbf{F_1}$	88.89	100	92.86	100	100	100
	$\mathbf{F_2}$	88.89	100	92.86	94.44	100	96.15
PERC	$\mathbf{F_1}$	88.89	90	89.29	88.89	100	92.31
	$\mathbf{F_2}$	83.33	100	89.29	100	100	100
KOZI	$\mathbf{F_1}$	100	90	96.43	94.44	100	96.15
	$\mathbf{F_2}$	100	100	100	94.44	100	96.15
KNN	$\mathbf{F_1}$	94.44	100	96.43	100	100	100
	$\mathbf{F_2}$	94.44	100	96.43	94.44	100	96.15
SVM	$\mathbf{F_1}$	100	100	100	100	100	100
	$\mathbf{F_2}$	100	100	100	100	100	100

Table 3: Experiment results of Task BIS2

		Left Hand			Right Hand		
		H	C	Avg	H	C	Avg
FLD	$\mathbf{F_1}$	77.77	45.45	67.86	38.89	63.64	48.28
	$\mathbf{F_2}$	61.11	72.73	67.86	66.67	27.27	51.72
QFLD	$\mathbf{F_1}$	83.33	100	89.66	77.78	63.64	72.41
	$\mathbf{F_2}$	83.33	100	89.66	61.11	63.64	62.07
PERC	$\mathbf{F_1}$	66.67	81.82	72.41	94.44	81.82	86.21
	$\mathbf{F_2}$	83.33	90.91	86.21	94.44	45.45	75.86
KOZI	$\mathbf{F_1}$	77.77	81.82	79.31	83.33	63.64	75.84
	$\mathbf{F_2}$	88.89	90.91	89.66	83.33	72.73	79.31
KNN	$\mathbf{F_1}$	83.33	72.73	79.31	83.33	63.64	75.86
	$\mathbf{F_2}$	94.44	81.82	89.66	83.33	63.64	75.86
SVM	$\mathbf{F_1}$	82.33	81.82	82.76	77.78	63.64	72.41
	$\mathbf{F_2}$	94.44	90.91	93.10	83.33	54.55	72.41

Table 4: Experiment results of Task BIS3

	Left Hand			Right Hand		
	G	C	Avg	G	C	Avg
FLD	60	45.45	52.38	75	45.45	57.89
QFLD	100	90.91	95.24	87.5	54.55	68.42
PERC	90	81.82	85.71	75	54.55	63.16
KOZI	100	90.91	95.24	87.5	63.64	73.64
KNN	100	90.91	95.24	100	45.45	68.42
SVM	100	90.91	95.24	100	72.73	76.19

Note: feature set = $\mathbf{F_2}$

References

[1] J. N. Cohn, S. Finkelstein, G. McVeigh, D. Morgan, L. LeMay, J. Robinson and J. Mock, "Noninvasive pulse wave analysis for the early detection of vascular disease," *American Journal of Hypertension*, vol. 26, 1995, pp. 503–508.

[2] M. Crilly, C. Coch, M. Bruce, H. Clark and D. Williams, "Indices of cardiovascular function derived from peripheral pulse wave analysis using radial applanation tonometry: a measurement repeatability study," *Journal of Vascular Medicine,* vol. 12, 2007, pp. 189–197.

[3] N. Cristianini, J. Shawe-Taylor, *An introduction to support vector machine*, Cambridge University Press, 2000, ch. 6.

[4] D. L. Donoho, "De-noising by soft thresholding, " *IEEE Trans. on Information Theory*, vol. 41, no. 3, 1995, pp. 613–627.

[5] H. Greenspan, O. Shechner, M. Scheinowitz, and M. S. Feinberg, "Doppler echocardiography flow-velocity image analysis for patients with arterial fibrillation," *Ultrasound in Medicine and Biology*, vol. 31, no. 8, 2005, pp. 1031–1040.

[6] E. Kazanavicius, R. Gircys, and A. Vrubliauskas, "Mathematical methods for determining the foot point of the arterial pulse wave and evaluation of proposed methods," *Information Technology and Control*, vol. 34, no. 1, 2005, pp. 29-36.

[7] L. Khadra, A. Al-Fahoum and S. Binajjaj, "A new quantitative analysis technique for cardiac arrhythmia classification using bispectrum and bicohency", in *Proc. 26th Annu. IEEE Int. Conf. on Engineering in Medicine and Biology*, San Francisco, 2004, pp. 13–16.

[8] E. Lau, A. T. Chuwang, "Relationships between wrist-pulse characteristics and body conditions," in *Proc. 14th Engineering Mechanics Conference of the American Society of Civil Engineers,* Austin, USA, 2000.

[9] Y. J. Lee, J. Lee and J. Y. Kim, "A study on characteristics of Radial Arteries through ultrasonic waves," in *Proc. 30th Annu. IEEE Int. Conf. on Engineering in Medicine and Biology*, Vancouver, Canada, 2008, pp. 2453–2456.

[10] W. A. Lu, C. H. Cheng, Y. Y. Wang Lin, and W. K. Wang, "A pulse spectrum analysis of the hospital patients with possible liver problems," *American Journal of Chinese Medicine*, vol. 45, 1996, pp. 311–320.

[11] W. A. Lu, Y. Y. Wang Lin, and W. K. Wang, "Pulse analysis of patients with severe liver problems," *IEEE Engineering in Medicine and Biology Magazine*, vol. 18, no.1, 1999, pp. 73–75.

[12] S. Raudys, K. A. Jain, "Small sample size effects in statistical pattern recognition: Recommendations for practitioners," *IEEE Trans. On Pattern analysis and Applications*, vol. 13, no. 3, 1991, pp. 254–265.

[13] S. T. Shi, *Yi Yuan (Sources of Medicine)*, Jiangsu Science and Technology Publishing House, 1983.

[14] J. K. Tugnait, "Detection of non-Gaussian signals using integrated polyspectrum", *IEEE Trans. on Signal Processing*, vol. 42, no. 11, 1994, pp. 3137–3149.

[15] B. H. Wang, J. L. Xiang, "Detecting system and power-spectral analysis of pulse signals of human body," in *Proc. 4th Int. Conf. on Signal Processing*, Beijing, China, 1998, pp. 1646-1649.