Biomedical Signal Processing and Control 32 (2017) 20–28

Contents lists available at ScienceDirect

Biomedical Signal Processing and Control

journal homepage: www.elsevier.com/locate/bspc

Heart sound classification based on scaled spectrogram and partial least squares regression

Wenjie Zhang [a], Jiqing Han [a,*], Shiwen Deng [b,*]

[a] School of Computer Science and Technology, Harbin Institute of Technology, Harbin, China
[b] School of Mathematical Sciences, Harbin Normal University, Harbin, China

ARTICLE INFO

Article history:
Received 30 May 2016
Received in revised form 9 September 2016
Accepted 13 October 2016
Available online 31 October 2016

Keywords:
Scaled spectrogram
Heart sound
Partial least squares regression

ABSTRACT

Phonocardiogram (PCG) signal analysis is an effective and convenient method for the preliminary diagnosis of heart disease. In this study, a scaled spectrogram and partial least squares regression (PLSR) based method was proposed for the classification of PCG signals. Proposed method is mainly comprised of four stages, namely as being heart cycle estimation, spectrogram scaling, dimension reduction and classification. At the heart cycle estimation stage, the short time average magnitude difference of the Shannon energy envelope is applied. Then the spectrogram of the obtained heart cycle is calculated for feature extraction. However, the sizes of the spectrograms between different PCG signals are usually not the same. In order to overcome the difficulty of direct comparison, the bilinear interpolation is used for the spectrogram to get the scaled spectrogram with a fixed size. Nevertheless, the scaled spectrogram contains a large quantity of redundant and irrelevant information. To extract the most relevant features from the scaled spectrogram, we adopt the PLSR to reduce the dimension of the scaled spectrograms. Since PLSR has the advantage of using the category information during the dimension reduction process, the extracted features are more discriminative. Then the classification results are obtained via support vector machine (SVM). The proposed method is evaluated on two public datasets offered by the PASCAL classifying heart sounds challenge, and the results are compared to those obtained using the best methods in the challenge, thereby proving the effectiveness of our method.

1. Introduction

Many pathological conditions of the cardiovascular system are reflected in heart sound signals, which makes it possible to diagnose heart disease by analyzing heart sound signals. Heart sound auscultation is a method used to analyze heart sound signals using a stethoscope. Because of its easy implementation, auscultation is widely used in the clinical diagnosis of heart disease [1,2]. However, the accuracy of auscultation depends on the skill and subjective experience of the physician [3]. Therefore, an objective analysis of heart sound signals is necessary. PCG signal analysis is another method of analyzing heart sound signals using phonocardiograms. The physiological and pathological information has been extracted from the PCG signal using signal processing and artificial intelligence techniques in the literature [3,4]. With the PCG, the objective analysis of heart sound signals using computer technology is becoming popular. Moreover, telemedicine is becoming available with the development of electronic stethoscopes and smart phones [5]. Overall, the analysis of PCG signals has important significance for the diagnosis of heart disease. Heart sound classification aims at the automatic classification of PCG signals. It is very important for preliminary diagnosis.

Heart sound classification usually involves two steps. The first step is heart sound segmentation, which attempts to detect the location of the fundamental heart sounds (FHs). The FHs include the first (S1) and second (S2) heart sounds, which are the important physical characteristics of heart sounds. The accurate localization of the FHs shows the systolic and diastolic regions of the heart sounds. In addition, the heart cycles are identified by the FHs. Thus, the characteristics of different pathological situations in the region of one heart cycle are used to classify different heart sound categories. Many methods, such as the envelope-based method [6], the method using dynamic clustering [7] and the logistic regression-hsmm based method [8], have been developed for this task. However, heart sound segmentation remains a challenging task, and it is difficult to segment the FHs accurately in a noisy environment.

* Corresponding authors.
E-mail addresses: jqhan@hit.edu.cn (J. Han), dengswen@gmail.com (S. Deng).

http://dx.doi.org/10.1016/j.bspc.2016.10.004
1746-8094/

The second step of heart sound classification is to extract the features in one heart cycle and use the features for classification. Many features have been proposed in the literature. The three main types are time [5], frequency [9] and time-frequency complexity-based features [10,11]. Although the time-frequency-based features are more computationally complex than features based on only time or frequency, they provide more comprehensive information about the PCG signal. Thus, time-frequency-based features usually outperform other features. The commonly used time-frequency feature extraction methods for PCG signals are wavelets [10], S-transform [12] and short time Fourier transform (STFT) [13]. The magnitude of the STFT yields the spectrogram. This spectrogram is used in this paper since it is easy to implement and convenient to scale.

The primary goal of heart sound classification is to identify different heart sound categories. This is not necessary for segmentation in some situations, especially when the heart cycles are known. So the estimation of heart cycle duration and alignment methods based on the envelope are proposed to obtain the heart cycles instead of locating both S1 and S2. The calculation process is simplified in this way. Although the correct segmentation information can improve the classification performance, it requires a lot of computing. More importantly, the segmentation results are not correct in many cases which greatly affect the accuracy of the classification.

The spectrogram is extracted for each heart cycle after the heart cycles are estimated. However, the sizes of the spectrograms are different since the heart rates of different PCG signals are usually not the same. This prohibits a direct comparison between the spectrograms of different PCG signals. A bilinear interpolation [14] method is used to scale the size of the spectrogram, thus enabling the direct comparison. Nevertheless, the scaled spectrogram contains a large quantity of redundant and irrelevant information. In order to extract the most relevant information, a dimension reduction process of the scaled spectrogram is adopted. In addition, the heart sound category provides valuable information to distinguish between different categories and it helps to improve the classification performance. As a result, the extracted features will be more discriminative if the category information is fully utilized during the dimension reduction process. PLSR [15] maximizes the correlation between the PCG signals and their corresponding category information during the dimension reduction process. Thus the category information is utilized. Also, PLSR is capable to robustly handle more descriptor variables than the number of samples. These are the advantage of PLSR compared with other dimensionality reduction method, such as principle component analysis (PCA) [16], linear discriminant analysis (LDA) [17]. Therefore, the discriminative features of the scaled spectrogram are extracted using PLSR in this paper. Finally, the classification is performed using the SVM classifier [11].

The main framework of this paper is shown in Fig. 1 and consists of four steps: estimation of heart cycle duration and alignment, spectrogram scaling of each heart cycle, PLSR and classification. PLSR consists of two parts, i.e., dimension reduction and regression. The contributions of this paper are threefold. First, the heart cycles are estimated and aligned instead of locating both S1 and S2 to simplify the calculation process. Second, the spectrograms of heart cycles of different lengths can be compared directly using the bilinear scaling process which has not been applied in heart sound researches to our knowledge. Third, the category information is utilized during the dimension reduction process. In this way, the extracted features are best correlated with their categories in the dimension reduction process which makes the features more discriminative.

Table 1
The number of samples in the training and testing dataset.

Dataset	Category	Training	Testing
Dataset-A	*Normal*	31	14
	Murmur	34	14
	Extra Heart Sound	19	8
	Artifact	40	16
Dataset-B	*Normal*	200	136
	Murmur	66	39
	Extrasystole	46	20

2. Method

2.1. Data collection

The datasets used in this paper, including Dataset-A and Dataset-B [18], are collected from the classifying heart sounds Pascal challenge competition. Dataset-A is collected by volunteers using iStethoscope which is an iPhone application that enables an iPhone to use its microphone as a digital stethoscope [19]. Dataset-A includes 176 records with a 44,100 Hz sampling frequency and it can be grouped into four categories: *Normal*, *Murmur*, *Extra Heart Sound* and *Artifact*. A normal heart sound has a clear lub dub, lub dub pattern, with the time from lub to dub shorter than the time from dub to the next lub [20]. In the *Murmur* category, the heart murmurs sound as though there is a whooshing, roaring, rumbling, or turbulent fluid noise in one of two temporal locations: (1) between lub and dub, or (2) between dub and lub [20]. A regular additional sound can be identified as an extra heart sound. A wide range of different sounds are contained in the *Artifact* category, including speech, music and noise. With the approval of the RHP Ethics Committee, Dataset-B was collected at the Real Hospital Portugues using a Littmann Model 3100 electronic stethoscope with a 4000 Hz sampling frequency. Dataset-B includes 507 records grouped into three categories: *Normal*, *Murmur* and *Extrasystole*. The *Extrasystole* heart sound is not the same as the extra heart sound in Dataset-A because the additional sound is not regularly occurring. Besides, no information is available on the auscultated subjects, such as gender, age, and condition. The number of samples used in the training and testing dataset is shown in Table 1.

2.2. Preprocessing

The collected PCG signals are often contaminated with high frequency noise. The main information of the PCG signals is concentrated at low frequencies. Therefore, the PCG signals are resampled to 2000 Hz before further processing. Additionally, the resampled signals are filtered with a band-pass (50–950 Hz), 6th-order Butterworth filter to further eliminate the noise. Then, the PCG signals are normalized to a fixed scale of [−1 1]:

$$x[n] = \frac{x'[n]}{\max_n(|x'[n]|)} \tag{1}$$

where $x'[n]$ is the resampled and filtered signal and $x[n]$ is the normalized signal. A murmur PCG signal after filtering and normalization is shown in Fig. 2 (a).

2.3. Estimation of heart cycle duration and alignment

The classification task is based on each heart cycle. First, the normalized signal $x[n]$ is decomposed and reconstructed using the discrete wavelet transform to accurately estimate the heart cycle. Because the Daubechies wavelet is morphologically similar to the PCG signal, it is widely used to decompose the PCG

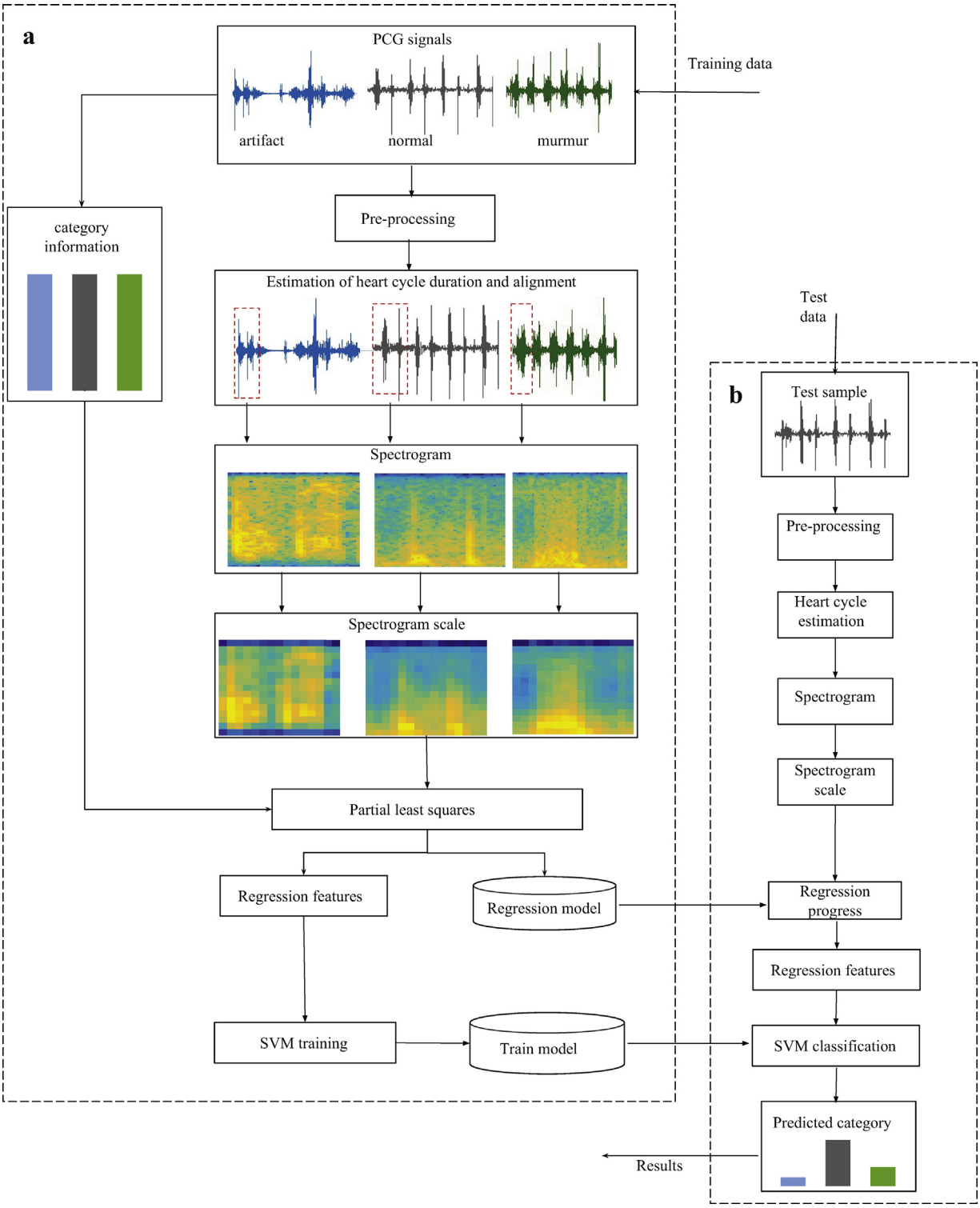

Fig. 1. The framework of heart sound classification. (a) The training phase and (b) the test phase.

signal. The 6th-order Daubechies wavelet has been proven to obtain the best performance [21]; therefore, it is used. The 4 level detail wavelet coefficients, which mainly contain information in the frequencies 62.5–125 Hz, are used to reconstruct the signal. Because the main energies of S1 and S2 are also in this frequency band, it is robust for estimating the heart cycle using these coefficients. A reconstructed murmur signal is shown in Fig. 2(b).

Then, the average Shannon energy envelopes are extracted based on the reconstructed signal:

$$e[m] = \frac{1}{R}\sum_{n_m} -x_r^2[n_m]\log x_r^2[n_m] \tag{2}$$

where R is the length of the sliding window and $x_r[n]$ is the reconstructed signal based on the wavelet coefficients.

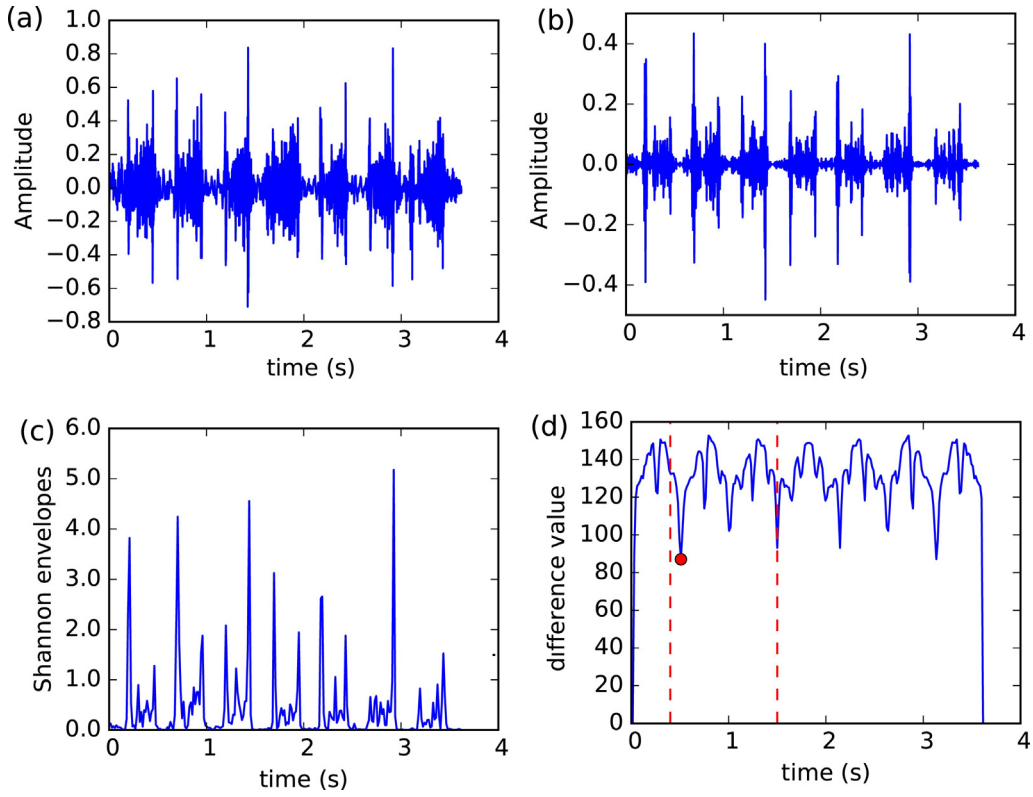

Fig. 2. (a) A murmur PCG signal after filtering and normalization; (b) the reconstructed signal using the four-level-of-detail wavelet coefficients; (c) the normalized average Shanon envelopes of (b); (d) the estimation of the cardiac cycle, where the dot is the estimated value and the dashed lines are the limited interval.

Finally, the heart cycle is estimated by the short time average magnitude difference method, which is a fast and effective method for estimating the signal period [22]. The difference values are calculated by

$$\gamma[k] = \frac{1}{M}\sum_{m}|e[m+k] - e[m]| \tag{3}$$

where M is the number of $e[m]$. The value of $\gamma[k]$ is limited in a certain interval because the duration of a heart cycle is usually between 0.4 s and 1.5 s [23]. Suppose that the corresponding values of k are limited to between a and b. Then, the value of k is estimated by $\{k : \min_k \gamma[k], a \le k \le b\}$, as shown in Fig. 2(d). Finally, the estimated duration of the heart cycle is $h = k \cdot R/2000$ s.

After the estimation of heart cycle duration, it is essential to align the beginning and end of each heart cycle. The process of alignment main contains two steps: (1) the S1 detection of the first heart cycle and (2) the S1s detection of other heart cycles. As shown in Fig. 3 (a), the S1 detection of the first heart cycle is based on the envelope. The first heart cycle of the envelope is intercepted based on the estimated heart cycle. Then the first two largest peaks of the first heart cycle is detected as the middle of S1 and S2, the interval between them is called S interval. As the systole period is usually no longer than the diastole period [24], the S1 is determined by the ratio between S interval and estimated heart cycle. If the ratio is greater than 0.5, the second peak is S1, otherwise the first peak is S1. In the second step, next S1 detection is based on the previous determined S1 and the sliding heart cycle window. As shown in Fig. 3 (b), the next S1 is determined by the highest peaks around the sliding window. Then the heart cycle is intercepted between the S1 and the next S1 near the previous one. Thus the heart cycles are aligned according to the detected S1s.

2.4. Spectrogram scale

Once the heart cycle has been estimated, the PCG signal is divided into a number of heart cycles according to the estimated duration. Then, the time-frequency analysis is implemented for each heart cycle separately. Because the fast Fourier transform is used to speed up the discrete Fourier transform (DFT) process, it is very fast and easy to implement the STFT. Suppose that x is the single interpreted heart cycle from the PCG signal; then, the STFT of x is

$$S[t,f] = \sum_{n} x[n]w[n-t]e^{-i2\pi nf} \tag{4}$$

where $w[t]$ is a window function. Taking the log power of S, the spectrogram of this heart cycle is obtained and is denoted as Ψ. Suppose that the size of Ψ is $T \times F$, T is determined by the length of the heart cycle, and F is determined by the window length of the STFT. Thus, $\Psi \in \mathbb{R}^{T\times F}$.

However, the size of Ψ is usually different between different PCG signals. Thus, the spectrograms cannot be directly compared. To solve this problem, the bilinear interpolation method is used. Suppose that the spectrogram after interpolation is $\Psi' \in \mathbb{R}^{T'\times F'}$. The main steps are as follows:

1. Set $s_T = T/T'$ and $s_F = F/F'$.
2. Suppose that $\Psi'(t', f')$ is one element in Ψ'.
   Set $\mu = t' \cdot s_T$ and $\xi = f' \cdot s_F$.
   Set $\Delta\mu = \mu - |\mu|$ and $\Delta\xi = \xi - |\xi|$.
   Round down the value of μ and ξ as $\mu = \lceil \mu \rceil$ and $\xi = \lceil \xi \rceil$.

3. Then, the element $\Psi'(t', f')$ is represented as

$$\begin{aligned}\Psi'(t', f') = {} & \Psi(\mu, \xi)\cdot(1-\Delta\mu)\cdot(1-\Delta\xi) + \Psi(\mu \\ & +1, \xi)\cdot\Delta\mu\cdot(1-\Delta\xi) + \Psi(\mu, \xi+1)\cdot(1-\Delta\mu)\cdot\Delta\xi \\ & + \Psi(\mu+1, \xi+1)\cdot\Delta\mu\cdot\Delta\xi\end{aligned}$$

4. Repeat step 2 and step 3 until all the elements in Ψ' are obtained.

Take the sizes of the Ψ' between different heart cycles to be equal. Thus, the sizes of the spectrograms are scaled to be equal through the bilinear interpolation method. Ψ' is called the scaled spectrogram. Thus, a direct comparison of the time-frequency features between different heart cycles is enabled.

2.5. Partial least squares regression

Although the features can be compared directly through the spectrogram scale, the dimension of Ψ' is relatively high. A dimension reduction process is necessary to improve the classification performance. Besides, the category information also provides significant value for classification. In order to utilize the category information during the dimension reduction process, PLSR is used to extract the most valuable information from Ψ'. The category information between different categories is supposed as the same. However, the training samples of each category are different. To solve the unbalanced problem, the weights between different heart sound categories are not equal.

Suppose that the number of heart sound categories is Λ and that the sample size of each category $\{\lambda : \lambda = 1, \ldots, \Lambda\}$ is N_λ. Then, the total size of the samples is

$$N = \sum_{\lambda=1}^{\Lambda} N_\lambda \tag{5}$$

In addition, the weight of the λ-th category is N/N_λ. The category information is denoted by $\boldsymbol{Y}$ and is represented as

$$\boldsymbol{Y} = \begin{bmatrix} N/N_1 & 0 & 0 & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ N/N_1 & 0 & 0 & 0 & 0 & 0 \\ 0 & N/N_2 & 0 & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & N/N_2 & 0 & 0 & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & N/N_\lambda & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & N/N_\lambda & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & 0 & 0 & N/N_\Lambda \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & 0 & N/N_\Lambda \end{bmatrix} \tag{6}$$

where the number of non-zero elements in the λ-th column is N_λ.

In order to prevent over-fitting, the scaled time-frequency features Ψ' are averaged for each sample by

$$\Psi'_V(t', f') = \frac{1}{V}\sum_{v=1}^{V}\Psi'_v(t', f') \tag{7}$$

where Ψ'_v is the v-th scaled spectrogram in the sample and V is the number of heart cycles in the sample. Thus the features of each sample are represented by one averaged $\Psi'_V(t', f')$. Although there will be some beat to beat variation in the heart cycle duration, meaning that the heart sounds in each heart cycle window will not be fully aligned, the variation usually is small and will not affect the classification performance. When the variation is big, it may indict some pathological conditions. Through averaging, it is reflected in the averaged spectrograms. Besides, through averaging the auscultation artifact noises in the sample are weakened.

Write Ψ'_V as a column vector, denoted by $\boldsymbol{x}$. Then, $\boldsymbol{x}$ is the feature for this sample. For the N samples, the corresponding features are $\boldsymbol{X} = [\boldsymbol{x_1}, \boldsymbol{x_2}, \ldots, \boldsymbol{x_N}]^T$.

The category information $\boldsymbol{Y}$ is used during the dimension reduction process for $\boldsymbol{X}$ through the following objective function:

$$\max_{||\boldsymbol{p}||=||\boldsymbol{q}||=1} Cov(\boldsymbol{Xp}, \boldsymbol{Yq}) \tag{8}$$

where the objective function attempts to maximize the correlation between $\boldsymbol{X}$ and $\boldsymbol{Y}$. The solution to the objective function is obtained as follows:

1. Set $r = 1$, $\boldsymbol{X}^{(r)} = \boldsymbol{X}$ and $\boldsymbol{Y}^{(r)} = \boldsymbol{Y}$.
2. Calculate the first left singular vector of ${\boldsymbol{X}^{(r)}}^T\boldsymbol{Y}^{(r)}$; the result is denoted as ${\boldsymbol{p_1}}^{(r)}$.
   Set $\boldsymbol{u}^{(r)} = \boldsymbol{X}^{(r)}{\boldsymbol{p_1}}^{(r)}$.
3. Regress $\boldsymbol{X}^{(r)}$ and $\boldsymbol{Y}^{(r)}$ based on $\boldsymbol{u}^{(r)}$ and obtain the rank-one approximation of the data and the residuals

$$\begin{aligned}\boldsymbol{X}^{(r)} &= \boldsymbol{u}^{(r)}{\boldsymbol{c}^{(r)}}^T + \boldsymbol{E}^{(r+1)} \\ \boldsymbol{Y}^{(r)} &= \boldsymbol{u}^{(r)}{\boldsymbol{d}^{(r)}}^T + \boldsymbol{F}^{(r+1)}\end{aligned} \tag{9}$$

   The values of $\boldsymbol{c}^{(r)}$ and $\boldsymbol{d}^{(r)}$ are obtained by

$$\begin{aligned}\boldsymbol{c}^{(r)} &= \frac{{\boldsymbol{X}^{(r)}}^T\boldsymbol{u}^{(r)}}{||\boldsymbol{u}^{(r)}||^2} \\ \boldsymbol{d}^{(r)} &= \frac{{\boldsymbol{Y}^{(r)}}^T\boldsymbol{u}^{(r)}}{||\boldsymbol{u}^{(r)}||^2}\end{aligned} \tag{10}$$

4. Set $r = r + 1$, $\boldsymbol{X}^{(r)} = \boldsymbol{E}^{(r)}$ and $\boldsymbol{Y}^{(r)} = \boldsymbol{F}^{(r)}$.
5. Repeat steps 2 to 4 until certain conditions are satisfied.

Finally, $\boldsymbol{X}$ and $\boldsymbol{Y}$ are denoted by

$$\begin{aligned}\boldsymbol{X} &= \boldsymbol{u}^{(1)}{\boldsymbol{c}^{(1)}}^T + \cdots + \boldsymbol{u}^{(r)}{\boldsymbol{c}^{(r)}}^T + \boldsymbol{E}^{(r+1)} \\ \boldsymbol{Y} &= \boldsymbol{u}^{(1)}{\boldsymbol{d}^{(1)}}^T + \cdots + \boldsymbol{u}^{(r)}{\boldsymbol{d}^{(r)}}^T + \boldsymbol{F}^{(r+1)}\end{aligned} \tag{11}$$

or simply by

$$\begin{aligned}\boldsymbol{X} &= \boldsymbol{U}\boldsymbol{C}^T + \boldsymbol{E}^{(r+1)} \\ \boldsymbol{Y} &= \boldsymbol{U}\boldsymbol{D}^T + \boldsymbol{F}^{(r+1)}\end{aligned} \tag{12}$$

where $\boldsymbol{U} = [\boldsymbol{u}^{(1)}, \cdots, \boldsymbol{u}^{(r)}]$, $\boldsymbol{C} = [\boldsymbol{c}^{(1)}, \cdots, \boldsymbol{c}^{(r)}]$, and $\boldsymbol{D} = [\boldsymbol{d}^{(1)}, \cdots, \boldsymbol{d}^{(r)}]$. $\boldsymbol{U}$ is the obtained features after the PLSR process. According to the PLSR result, the regression feature $\boldsymbol{x'}_t$ of a new sample $\boldsymbol{x}_t$ from the test set is

$${\boldsymbol{x'}_t}^T = {\boldsymbol{x}_t}^T\boldsymbol{P} \tag{13}$$

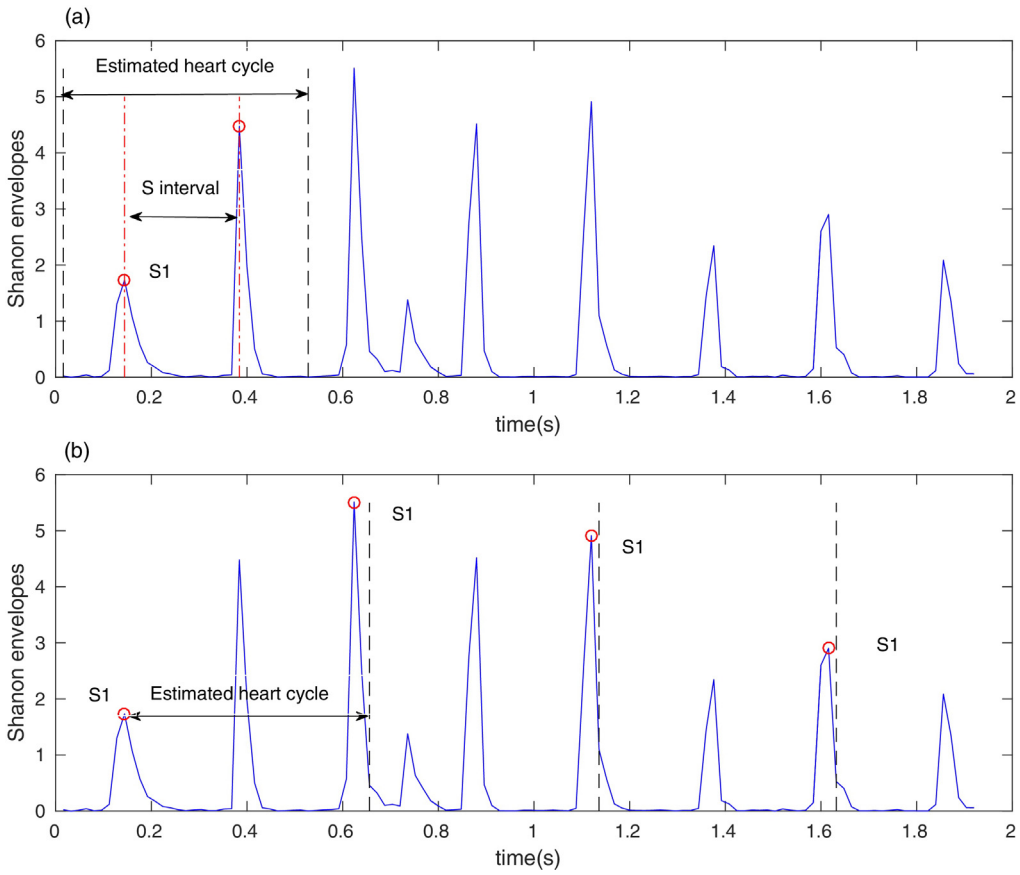

Fig. 3. (a) S1 detection of the first heart cycle; (b) S1s detection of other heart cycles.

where $\boldsymbol{P} = [\boldsymbol{p_1}^{(1)}, \boldsymbol{p_1}^{(2)}, \cdots, \boldsymbol{p_1}^{(r)}]$.

3. Results

The experimental results on the two datasets are compared with the three best methods in the challenge competition: J48 [20], MLP [20] and CS UCL [25]. All the methods are evaluated on the same datasets for the same evaluation criteria.

In the methods of J48 and MLP, only the temporal features are used. The difference is that J48 uses the decision trees as the classifier and MLP uses multi layer perceptron for classification [20]. In the CS UCL method, the wavelet decomposition and spectrogram is used for feature extraction and the decision trees is used for classification [25].

In order to evaluate our method, the scaled spectrogram features are used for comparison and the method is abbreviated as SS-method for convenience. However, the scaled spectrogram features are hard to extract the decision trees. Therefore, the decision trees classifier is not suitable for the classification of the scaled spectrogram features. Instead, the SVM classifier is used. To evaluate the PLSR features, our approach using PLSR features is experimented and the method is abbreviated as SS-PLSR for convenience.

3.1. Evaluation criteria

The evaluation criteria used in this paper are the precision of each heart sound category in the test sets and the evaluation of the problem heart sound categories. To measure the precision of each category, the number of true positives (TP) and the number of false positives (FP) are first obtained. Thus, the precision is defined as

$$P_r = \frac{TP}{TP + FP} \tag{14}$$

The sensitivity and specificity are used to evaluate the specific problem heart sound categories:

$$\begin{aligned} S_e &= \frac{TP}{TP + FN} \\ S_p &= \frac{TN}{FP_{TN} + TN} \end{aligned} \tag{15}$$

where S_e is the sensitivity, S_p is the specificity, TN is the number of true negatives, and FN is the number of false negatives, FP_{TN} is the number of false positives according to the classification results of negative samples. Note that the FP in Equation (14) is calculated according to the classification results of positive samples. Since the experiment is evaluated for multi classification problem which means only one category is recognized for each sample, FP and FP_{TN} may be different.

In Dataset-A, the sensitivity and specificity are used to evaluate the *Artifact* category. A wide range of different sounds, including feedback squeals and echoes, speech, music and noise are in the *Artifact* category. It is important to distinguish such noise from the other three categories. In addition, the F-score [26] is used to evaluate the problematic heartbeats, including the *Murmur* and *Extra Heart Sound* categories:

$$F_s = \frac{(\beta^2 + 1)P_r S_e}{\beta^2 P_r S_e} \tag{16}$$

where the parameter is set as $\beta = 0.9$.

Table 2
Results on Dataset-A.

Evaluation criteria	J48	MLP	CS UCL	SS-method	SS-PLSR
Precision of *Normal*	0.25	0.35	0.46	**0.67**	0.60
Precision of *Murmur*	0.47	0.67	0.31	**0.91**	**0.91**
Precision of *Extra Heart Sound*	0.27	0.18	0.11	0.37	**0.44**
Precision of *Artifact*	0.71	0.92	0.58	0.76	**0.94**
Artifact sensitivity	0.63	0.69	0.44	**1.00**	**1.00**
Artifact specificity	0.39	0.44	0.44	0.58	**0.64**
Youden index of *Artifact*	0.01	0.13	−0.09	0.58	**0.64**
F-score	0.20	0.20	0.14	0.28	**0.30**
Total precision	1.71	2.12	1.47	2.71	**2.89**

The bold values represent the best scores among the methods.

Table 3
Results on Dataset-B.

Evaluation criteria	J48	MLP	CS UCL	SS-method	SS-PLSR
Precision of *Normal*	0.72	0.70	**0.77**	0.74	0.76
Precision of *Murmur*	0.32	0.30	0.37	**0.66**	0.65
Precision of *Extrasystole*	0.33	**0.67**	0.17	0.24	0.33
Heart problem sensitivity	0.22	0.19	**0.51**	0.24	0.34
Heart problem specificity	0.82	0.84	0.59	0.89	**0.90**
Youden index of heart problem	0.04	0.02	0.01	0.13	**0.24**
Discriminant power	0.05	0.04	0.09	0.24	**0.36**
Total precision	1.37	1.67	1.31	1.57	**1.75**

The bold values represent the best scores among the methods.

In Dataset-B, the precision and sensitivity are used to evaluate the heart problem categories, including the *Murmur* and *Extrasystole* categories. In addition, the discriminant power [26] is used to evaluate the heart problem categories. The measure mainly evaluates how well an algorithm distinguishes between positive and negative examples:

$$D_p = \frac{\sqrt{3}}{\pi}\left(\log\frac{S_e}{1-S_e} + \log\frac{S_p}{1-S_p}\right) \tag{17}$$

Youden's index [26] is used to evaluate the performance the specific heart categories. It is used for both datasets:

$$Y_i = S_e - (1 - S_p) \tag{18}$$

3.2. Results of Dataset-A

The main purpose of evaluating Dataset-A is to distinguish between the *Normal*, *Murmur*, *Extra Heart Sound* and *Artifact* categories. One sample from each category is shown in Fig. 4.

The main experimental results are shown in Table 2. The results are analyzed from three perspectives:

1. The precision of the *Normal*, *Murmur*, *Extra Heart Sound* and *Artifact* categories.
2. The performance of the *Artifact* category.
3. The evaluation of the problematic heartbeats, including the *Murmur* and *Extra Heart Sound* categories.

The highest precision score on the *Normal* category is 67%, which is achieved by SS-method. And the highest precision scores on the *Murmur*, *Extra Heart Sound* and *Artifact* categories are all achieved by SS-PLSR method, the corresponding scores are 91%, 44% and 94%, respectively. The highest total precision 2.89 is also achieved by SS-PLSR method. In particular, the precision performance of SS-method outperforms the US UCL in all aspects, including the individual category precision and the total precision. This indicates that the scaled spectrogram features outperform the spectrogram features extracted from the method of US UCL. After the PLSR progress, the perform of our method has improved with a total precision increase of 0.18 which proves the effectiveness of the introduced labels information by PLSR.

Because the speech, music and noise are mixed in the *Artifact* category, the *Artifact* category should be very different from the other three categories. It is important to distinguish this category from the other categories. The performance on the *Artifact* category is further evaluated in terms of sensitivity and specificity. Youden's index is also considered. The Youden's index of our method is much higher than the other three methods. Overall, our method has a good ability to detect the *Artifact* category.

Finally, the problematic heartbeats are evaluated in terms of the F-score. The result obtained by SS-method and SS-PLSR methods are 0.28 and 0.30, which both are higher than the J48, MLP and CS UCL methods. This indicates that our method is better at detecting the problematic heartbeats.

3.3. Results on Dataset-B

There are only three categories in Dataset-B: the *Normal*, *Murmur* and *Extrasystole* categories. Note that the *Extrasystole* category in Dataset-B is different from the *Extra Heart Sound* category in Dataset-A. The extra heart sounds in the *Extrasystole* category are irregular, whereas they are regular in the *Extra Heart Sound* category. The main experimental results are shown in Table 3. The results are mainly analyzed from two perspectives:

1. The precision of the *Normal*, *Murmur* and *Extrasystole* categories.
2. The performance of the heart problem categories, including the *Murmur* and *Extrasystole* categories.

The overall total precision of our SS-PLSR method is 1.75, which is the highest among the methods. In particular, the precisions of the *Murmur* category in SS-method and SS-PLSR methods are 66% and 65%, which are much higher than that of the other three methods. Compared with CS UCL method, SS-method outperforms with a superior total precision 1.57. This indicts that the scaled spectrogram features outperform the spectrogram features extracted from the method of US UCL.

It is important to distinguish the *Normal* category from the heart problem categories. Thus, the *Murmur* and *Extrasystole* categories are evaluated and combined. The best specificity of the heart problems is 0.90 which is obtained by our SS-PLSR method. Moreover, the evaluation criteria of Youdens index and discriminant power achieved the best results through our method. In addition, the results are much better than those obtained by the other three methods. The Youdens index of SS-method and SS-PLSR methods are 0.13 and 0.24 which are much higher than the other best result 0.05. The discriminant power 0.24 and 0.36 are also more than two times of the third best result 0.09. These show that our method is competitive.

4. Discussion

The objective of this paper is to classify different heart sound signals automatically. Thus, the classification results provide a preliminary diagnosis, which helps to determine whether further diagnosis is necessary. The main categories of heart sounds in this paper are normal, murmur and some problematic heart beats. Their physiological and pathological information is contained in the heart cycles. It is reasonable to use the heart cycles information for classification instead of locating both S1 and S2. Moreover, locating both S1 and S2 is very difficult for some pathological PCG signals.

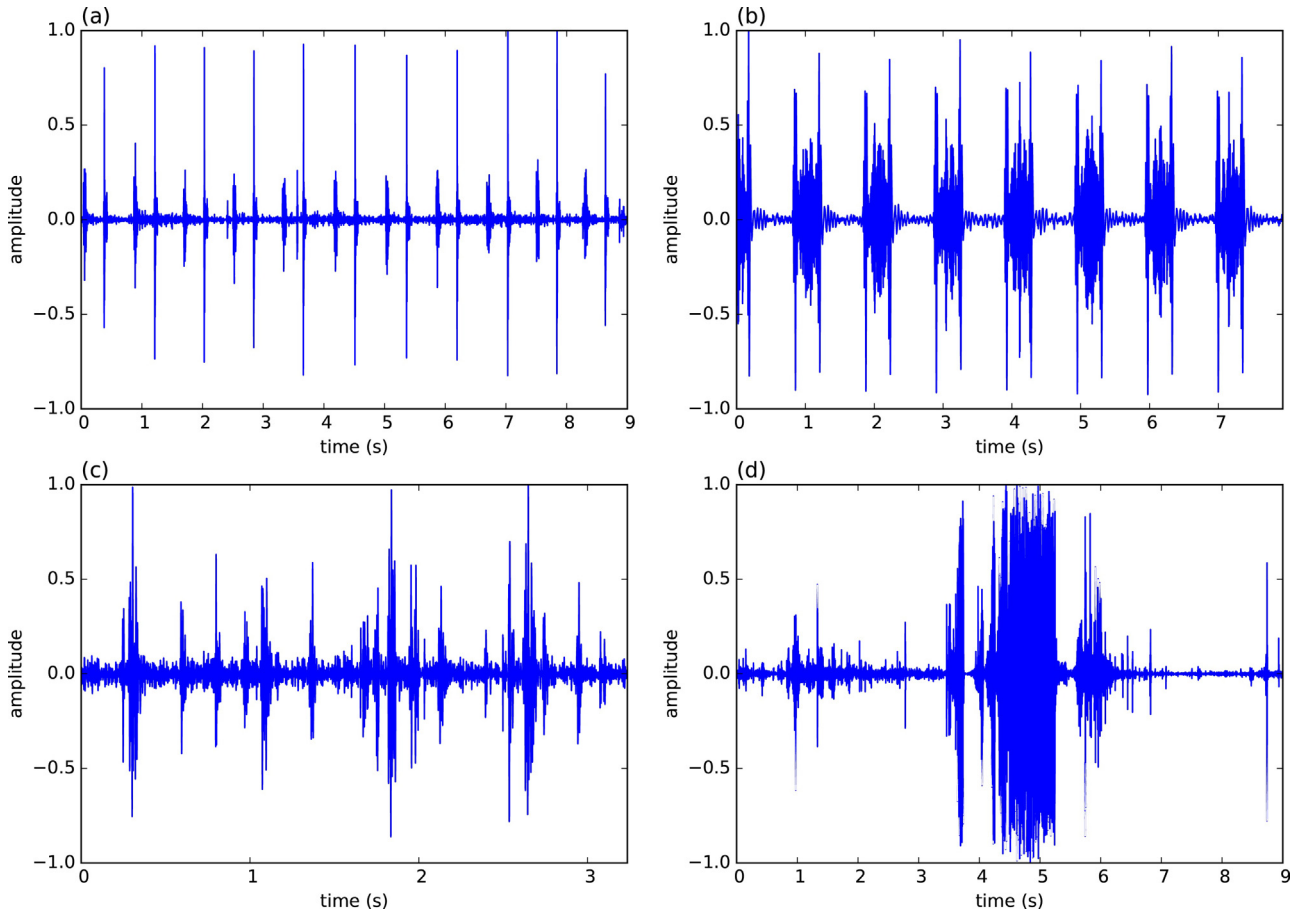

Fig. 4. The PCG samples of four heart sound categories in Dataset-A. (a) *Normal*, (b) *Murmur*, (c) *Extra Heart Sound*, and (d) *Artifact*.

The experimental results are based on the estimated and aligned heart cycles of the PCG signals. This will greatly simplify the classification process. Although the main energies of the PCG signals in the *Artifact* category of Dataset-A are not quasi-periodic because the signal-to-noise ratio (SNR) is very low, the heart sounds mainly concentrated in the low frequency range are quasi-periodic. The heart sounds in the low frequencies are able to estimate and align the heart cycle. The features are averaged for each sample based on the heart cycle numbers in the sample. It continues to be easy to detect the noise in the *Artifact* category because of the low SNR.

The spectrogram is calculated and scaled for each heart cycle. Thus, the features are directly comparable despite the different sizes of the spectrograms. The sizes of the scaled spectrograms should be set large enough to provide sufficient information. However, the larger the size, the more complex the spectrogram. In addition, it is unfavorable for the PLSR process if the sizes are too large because the number of samples is limited. The size and time-frequency information retained requires a trade off. The size of the scaled spectrograms used in this paper is 30×20. Finally, the category information is used during the PLSR process. The weights of samples from different categories are not the same because the numbers of samples in the training set are different. To solve the problem of unbalanced samples, the weights of the samples for a category with fewer samples are higher than those for a category with more samples. The methods are evaluated on two datasets and compared with three other methods. The highest total precision is achieved on the two datasets.

5. Conclusion

This paper proposed a novel method for classification based on scaled spectrograms and PLSR. This method can efficiently detect whether a PCG signal is problematic. Thus, it provides valuable information for deciding whether further treatment is necessary. Instead of characterizing the feature of a heart cycle obtained via explicit segmentation, the feature is extracted based on the estimated heart cycle to simplify the computation process. The spectrogram extracted from the heart cycle is scaled to the same size via a bilinear interpolation method. As a result, a direct comparison between different lengths of heart cycles is enabled. Finally, the category information of the heart sound samples is used during the PLSR process. The proposed framework is evaluated on the public datasets from the PASCAL classifying heart sounds challenge. The experiments show that our method significantly outperform the baseline methods.

Acknowledgements

This research is partly supported by the National Natural Science Foundation of China under grant Nos. 91120303 and 61471145.

References

[1] I.R. Hanna, M.E. Silverman, A history of cardiac auscultation and some of its contributors, Am. J. Cardiol. 90 (3) (2002) 259–267.
[2] R.M. Rangayyan, R.J. Lehner, Phonocardiogram signal analysis: a review., Crit. Rev. Biomed. Eng. 15 (3) (1986) 211–236.

[3] Z. Jiang, S. Choi, A cardiac sound characteristic waveform method for in-home heart disorder monitoring with electric stethoscope, Expert Syst. Appl. 31 (2) (2006) 286–298.
[4] J. Herzig, A. Bickel, A. Eitan, N. Intrator, Monitoring cardiac stress using features extracted from s1 heart sounds, IEEE Trans. Biomed. Eng. 62 (4) (2015) 1169–1178.
[5] S.-W. Deng, J.-Q. Han, Towards heart sound classification without segmentation via autocorrelation feature and diffusion maps, Future Gener. Comput. Syst. 60 (2016) 13–21.
[6] H. Liang, S. Lukkarinen, I. Hartimo, Heart sound segmentation algorithm based on heart sound envelogram, in: Computers in Cardiology Conference, IEEE, 1997, pp. 105–108.
[7] H. Tang, T. Li, T. Qiu, Y. Park, Segmentation of heart sounds based on dynamic clustering, Biomed. Signal Process. Control 7 (5) (2012) 509–516.
[8] D. Springer, L. Tarassenko, G. Clifford, Logistic regression-HSMM-based heart sound segmentation, IEEE Trans. Biomed. Eng. 63 (2015) 822–832.
[9] F. Safara, S. Doraisamy, A. Azman, A. Jantan, A.R.A. Ramaiah, Multi-level basis selection of wavelet packet decomposition tree for heart sound classification, Comput. Biol. Med. 43 (10) (2013) 1407–1414.
[10] S. Ari, K. Hembram, G. Saha, Detection of cardiac abnormality from PCG signal using LMS based least square SVM classifier, Expert Syst. Appl. 37 (12) (2010) 8019–8026.
[11] I. Maglogiannis, E. Loukis, E. Zafiropoulos, A. Stasis, Support vectors machine-based identification of heart valve diseases using heart sounds, Comput. Methods Program Biomed. 95 (1) (2009) 47–61.
[12] A. Moukadem, A. Dieterlen, N. Hueber, C. Brandt, A robust heart sounds segmentation module based on s-transform, Biomed. Signal Process. Control 8 (3) (2013) 273–281.
[13] Y. Soeta, Y. Bito, Detection of features of prosthetic cardiac valve sound by spectrogram analysis, Appl. Acoust. 89 (2015) 28–33.
[14] X. Li, M.T. Orchard, New edge-directed interpolation, IEEE Trans. Image Process. 10 (10) (2001) 1521–1527.
[15] P. Geladi, B.R. Kowalski, Partial least-squares regression: a tutorial, Anal. Chim. Acta 185 (1986) 1–17.
[16] I. Jolliffe, Principal Component Analysis, Wiley Online Library, 2002.
[17] S. Balakrishnama, A. Ganapathiraju, Linear discriminant analysis – a brief tutorial, Inst. Signal Inf. Process. 18 (1998) 1–9.
[18] N. Marques, R. Almeida, A.P. Rocha, M. Coimbra, Exploring the stationary wavelet transform detail coefficients for detection and identification of the s1 and s2 heart sounds, in: Computing in Cardiology Conference, IEEE, 2013, pp. 891–894.
[19] K. Senior, Smart phones: new clinical tools in oncology? Lancet Oncol. 12 (5) (2011) 429–430.
[20] E.F. Gomes, E. Pereira, Classifying heart sounds using peak location for segmentation and feature construction, in: Workshop Classifying Heart Sounds, La Palma, Canary Islands, 2012.
[21] D. Kumar, P. Carvalho, M. Antunes, R.P. Paiva, J. Henriques, An adaptive approach to abnormal heart sound segmentation, in: International Conference on Acoustics, Speech and Signal Processing, IEEE, 2011, pp. 661–664.
[22] M.J. Ross, H.L. Shaffer, A. Cohen, R. Freudberg, H.J. Manley, Average magnitude difference function pitch extractor, IEEE Trans. Acoust. Speech Signal Process. 22 (5) (1974) 353–362.
[23] V. Elharrar, B. Surawicz, Cycle length effect on restitution of action potential duration in dog cardiac fibers, Am. J. Physiol. Heart Circ. Physiol. 244 (6) (1983) 782–792.
[24] P. Wang, Y. Kim, L. Ling, C. Soh, First heart sound detection for phonocardiogram segmentation, in: IEEE Conference Engineering in Medicine and Biology, IEEE, 2006, pp. 5519–5522.
[25] Y. Deng, P.J. Bentley, A robust heart sound segmentation and classification algorithm using wavelet decomposition and spectrogram, in: Workshop Classifying Heart Sounds, La Palma, Canary Islands, 2012.
[26] P. Bentley, G. Nordehn, M. Coimbra, S. Mannor, The Pascal Classifying Heart Sounds Challenge, 2011, See: http://www.peterjbentley.com/heartchallenge/index.html.