Expert Systems With Applications 84 (2017) 220–231

Contents lists available at ScienceDirect

Expert Systems With Applications

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

Heart sound classification based on scaled spectrogram and tensor decomposition

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 November 2016
Revised 10 April 2017
Accepted 5 May 2017
Available online 8 May 2017

Keywords:
Heart sound
Scaled spectrogram
Tensor decomposition

ABSTRACT

Heart sound signal analysis is an effective and convenient method for the preliminary diagnosis of heart disease. However, automatic heart sound classification is still a challenging problem which mainly reflected in heart sound segmentation and feature extraction from the corresponding segmentation results. In order to extract more discriminative features for heart sound classification, a scaled spectrogram and tensor decomposition based method was proposed in this study. In the proposed method, the spectrograms of the detected heart cycles are first scaled to a fixed size. Then a dimension reduction process of the scaled spectrograms is performed to extract the most discriminative features. During the dimension reduction process, the intrinsic structure of the scaled spectrograms, which contains important physiological and pathological information of the heart sound signals, is extracted using tensor decomposition method. As a result, the extracted features are more discriminative. Finally, the classification task is completed by support vector machine (SVM). Moreover, the proposed method is evaluated on three public datasets offered by the PASCAL classifying heart sounds challenge and 2016 PhysioNet challenge. The results show that the proposed method is competitive.

1. Introduction

Many cardiac abnormalities are reflected in heart sound signals, which makes it possible to diagnose heart disease by analysing heart sound signals. Heart sound auscultation is a frequently used method to analysis heart sound signals using a stethoscope. As auscultation is convenient to implementation, it is widely used in the clinical diagnosis of heart disease (Hanna & Silverman, 2002; Rangayyan & Lehner, 1986). However, accurate auscultation needs a long physician experience, which is not easy to obtain (Jiang & Choi, 2006). Therefore, a computer assist tool for heart sound analysis is needed to help diagnose heart disease. Phonocardiogram (PCG) signal analysis is another method of analysing 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 literatures (Herzig, Bickel, Eitan, & Intrator, 2015; Jiang & Choi, 2006). With the PCG, the automatic analysis of heart sound signals using computer technology is becoming popular. Moreover, the telemedicine is becoming available with the development of electronic stethoscopes and smart phones (Deng & Han, 2016). 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 to preliminary diagnosis.

Heart sound classification usually involves three steps. The first step is heart sound segmentation. It attempts to detect the location of the fundamental heart sounds (FHs), including the first (S1) and second (S2) heart sounds, which are the important physiological characteristics of heart sounds. According to the accurate localization of the FHs, the systolic and diastolic regions of the heart sounds are detected. In addition, the heart cycles are also identified by FHs. Many methods have been developed for heart sound segmentation, such as, the amplitude threshold based methods (Chen, Kuan, Celi, & Clifford, 2010; Liang, Lukkarinen, & Hartimo, 1997; Moukadem, Dieterlen, Hueber, & Brandt, 2013; Sun, Jiang, Wang, & Fang, 2014) and probabilistic models based methods (Schmidt, Holst-Hansen, Graff, Toft, & Struijk, 2010; Springer, Tarassenko, & Clifford, 2015). However, heart sound segmentation remains a challenging task, and it is difficult to segment the FHs accurately in a noisy environment.

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

http://dx.doi.org/10.1016/j.eswa.2017.05.014
0957-4174/

Table 1
Summary of the heart sound feature extraction methods.

Author	Features	Advantages	Disadvantages
(Ari et al., 2010)	Time	Easy to calculate and quantify the features.	Physiological and pathological information is not fully extracted.
(Safara et al., 2013)	Frequency		
(Ari et al., 2010) (Deng & Han, 2016) (Soeta & Bito, 2015) (Gokhale, 2016) (Her & Chiu, 2016) (Teo et al., 2016) (Thomae & Dominik, 2016) (Zhang et al., 2017)	Time-frequency	Discriminative features are extracted automatically according to demand, more physiological and pathological information is available.	Discriminative features are hard to extract.

The second step is feature extraction. It attempts to extract features for classification based on the segmentation results. Many features have been proposed in the literatures. The three main types are time (Ari, Hembram, & Saha, 2010), frequency (Safara, Doraisamy, Azman, Jantan, & Ramaiah, 2013) and time-frequency complexity-based features (Ari et al., 2010; Deng & Han, 2016; Maglogiannis, Loukis, Zafiropoulos, & Stasis, 2009). Although the time or frequency based features are easy to understand and calculate according to the physiological characteristics of the heart sound signals, such as the heart rate, the systolic and diastolic period, they may lose some important pathological information which is not easy to quantify in the time or frequency domain independently, such as the shape of S2 in the time-frequency domain. In order to improve the performance of heart sound classification, extracting more discriminative features from the time-frequency domain is becoming more and more popular in recent research (Gokhale, 2016; Her & Chiu, 2016; Teo, Yang, Feng, & Su, 2016; Thomae & Dominik, 2016; Zhang, Han, & Deng, 2017). Moreover, the PCG signals need to be represented in the transformed time-frequency domain to extract features in these methods. The commonly used transform methods for PCG signals are wavelets (Ari et al., 2010), S-transform (Livanos, Ranganathan, & Jiang, 2000; Moukadem et al., 2013) and short time Fourier transform (STFT) (Soeta & Bito, 2015). Since the spectrogram calculated by STFT is easy to implement and convenient to scale, it is used in this paper. For a better comparison, the feature extraction methods and their advantages and disadvantages are summarized in Table 1. As shown in Table 1, the heart sound feature extraction is still a challenging task which is caused by the non-stationary and diversity of the PCG signals.

The last step is classification. It aims at choosing a suitable classifier to evaluate the extracted features and complete the classification task. The commonly used methods in heart sound classification are k-Nearest Neighbors (Quiceno-Manrique, Godino-Llorente, Blanco-Velasco, & Castellanos-Dominguez, 2010), Hidden Markov Model (SaraçOğLu, 2012), Artificial Neural Network (Uğuz, 2012), decision trees (Deng & Bentley, 2012) and SVM (Zheng, Guo, & Ding, 2015). The classifier is generally chosen according to the features extracted from the heart sound signal.

The primary goal of heart sound classification is to identify different heart sound categories. It is not necessary for the detailed segmentation in some situations, especially when the heart cycles are known (Avendano-Valencia, Godino-Llorente, Blanco-Velasco, & Castellanos-Dominguez, 2010). Therefore, the heart sound signal is segmented into heart cycles instead of explicitly S1 and S2, as we did in the previous research (Zhang et al., 2017). In this way, the physiological and pathological information in the heart cycles is reserved and the segmentation is performed in a more efficient approach. Then, the spectrogram is extracted from each heart cycle after the heart cycles are detected. 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. To overcome the problem, the bilinear interpolation (Hariharan, Arbeláez, Girshick, & Malik, 2015) method is used to scale the size of the spectrogram, thus the direct comparison is enabled. 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. The commonly used method of dimension reduction is the principal component analysis (PCA) (Jolliffe, 2002) method. In order to perform the PCA, the spectrogram needs to be expanded into a vector form. Thus, during the dimension reduction process, the intrinsic structure information in the time-frequency domain is lost. However, the intrinsic structure of the scaled spectrograms, such as the relative position between the S1 and S2 in the time-frequency domain, contains the physiological and pathological information of the heart sound signal, which is important for heart sound classification. To keep the intrinsic structure of the scaled spectrograms during the dimension reduction process, Avendano-Valencia et al. (2010) proposed the 2D-PCA method. Whereas, the time or frequency information is extracted independently in the 2D-PCA method. As a result, the intrinsic structure is not fully extracted. In order to extract the relevant intrinsic structure information more efficiently during the dimension reduction process, the Tucker-2 tensor decomposition (Kolda & Bader, 2009) method is used in our study for its ability to keep the intrinsic structure information. Moreover, the extracted features are competitive compared with the features calculated from the literatures by evaluation. Finally, the classification is performed using the SVM classifier.

The main framework of this paper is shown in Fig. 1 and consists of three steps: heart cycle segmentation, feature extraction, including spectrogram scaling and tensor decomposition, and classification. The main contribution of this paper is that we proposed a more efficient method to extract the intrinsic structure of the scaled heart cycle spectrograms during the dimension reduction process using tensor decomposition. Thus, more useful physiological and pathological information is reserved during the dimension reduction process. As a result, the extracted features are more discriminative. Moreover, the extracted features are adopted for heart sound classification which can provide a primary diagnosis in the primary health center and home care. In addition, the tensor decomposition method has not been applied for heart sound spectrogram analysis in the literatures to our knowledge.

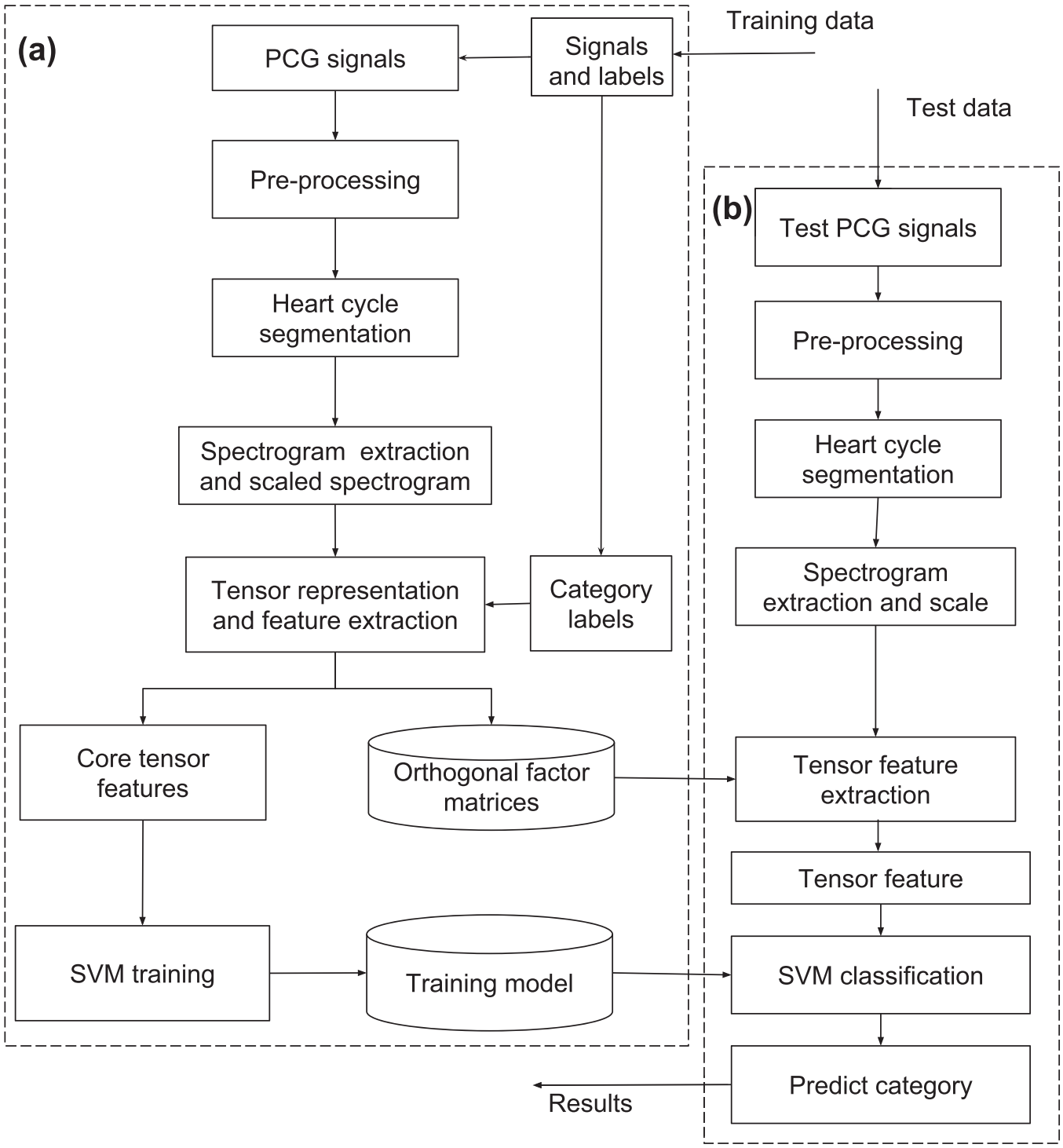

Fig. 1. The framework of heart sound classification. (a) The training phase and (b) the test phase.

2. Method

2.1. Data collection

Three datasets are used in this paper, including Dataset-A, Dataset-B and Dataset-C. Dataset-A and Dataset-B (Marques, Almeida, Rocha, & Coimbra, 2013) are collected from the classifying heart sounds PASCAL challenge competition. Dataset-C is collected from the 2016 PhysioNet challenge (Liu et al., 2016). 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 (Senior, 2011). Dataset-A includes 176 records of a 44100 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 (Gomes & Pereira, 2012). 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

Table 2
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
Dataset-C	*Normal*	2575	–
	Abnormal	665	–

(Gomes & Pereira, 2012). 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. Dataset-C is contributed by nine different research organizations. There are two categories are labeled, including normal and abnormal heart sound recordings. The categories of the abnormal heart sounds mainly contain mitral valve prolapse, innocent or benign murmurs, aortic disease, miscellaneous pathological conditions, coronary artery disease, mitral regurgitation and some pathological recordings without specific labels. The heart sound signals are all sampled to 2000 Hz. The number of samples used in training and testing dataset is shown in Table 2.

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.

2.3. Heart cycle segmentation

In order to make a fair comparison, the heart cycles are segmented. Then, the classification task is based on each segmented heart cycle. First, the normalized signal $x[n]$ is decomposed and reconstructed using the discrete wavelet transform. Since the Daubechies wavelet is morphologically similar to the PCG signal, the 4 level detail wavelet coefficients of the 6th-order Daubechies decomposed wavelet is used for reconstruction. Then the average Shannon energy envelopes are extracted based on the reconstructed signal. Finally, the heart cycle duration is estimated using the short time average magnitude difference function of the average Shannon energy envelopes.

$$\gamma[c] = \frac{1}{M}\sum_m |e[m+k] - e[m]| \tag{2}$$

where $e[m]$ is the extracted envelopes and M is the number of $e[m]$. The value of $\gamma[c]$ is limited in a certain interval because the duration of a heart cycle is usually between 0.4 s and 1.5 s (Elharrar & Surawicz, 1983). Suppose that the corresponding values of c are limited to between a and b. Then, the value of c is estimated by $\{c : \min_c \gamma[c]. a \le k \le b\}$.

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 S1 detection of other heart cycles. The S1 detection of the first heart cycle is based on the envelope. As shown in Fig. 2(a), 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 (Wang, Kim, Ling, & Soh, 2006), 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, as shown in Fig. 2(b), next S1 detection is based on the previous determined S1 and the sliding heart cycle window. 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 segmented.

2.4. Spectrogram extraction and scaled spectrogram

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. Suppose that x is the single interpreted heart cycle from the PCG signal; then, the STFT of x is

$$S[t, f] = \sum_n x[n]w[n-t]e^{-i2\pi nf} \tag{3}$$

where $w[t]$ is a window function. Taking the logarithmic power of S, the spectrogram of this heart cycle is obtained. However, the size of the spectrogram is usually different between different PCG signals. Thus, the spectrograms cannot be directly compared. To solve this problem, the sizes of the spectrograms are scaled to be equal through the bilinear interpolation method (Zhang et al., 2017).

Suppose that the spectrogram before interpolation is $S \in \mathbb{R}^{T\times F}$, after interpolation is $X \in \mathbb{R}^{I_1\times I_2}$. The main steps are as follows:

1. Set $\eta_T = T/I_1$ and $\eta_F = F/I_2$.
2. Suppose that $X(i_1, i_2)$ is one element in X.
   Set $\mu = i_1 \cdot \eta_T$ and $\xi = i_2 \cdot \eta_F$.
   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 $X(i_1, i_2)$ is represented as

$$\begin{aligned} X(i_1, i_2) = {} & S(\mu, \xi)\cdot(1-\Delta\mu)\cdot(1-\Delta\xi) \\ & + S(\mu+1, \xi)\cdot\Delta\mu\cdot(1-\Delta\xi) \\ & + S(\mu, \xi+1)\cdot(1-\Delta\mu)\cdot\Delta\xi \\ & + S(\mu+1, \xi+1)\cdot\Delta\mu\cdot\Delta\xi \end{aligned}$$

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

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

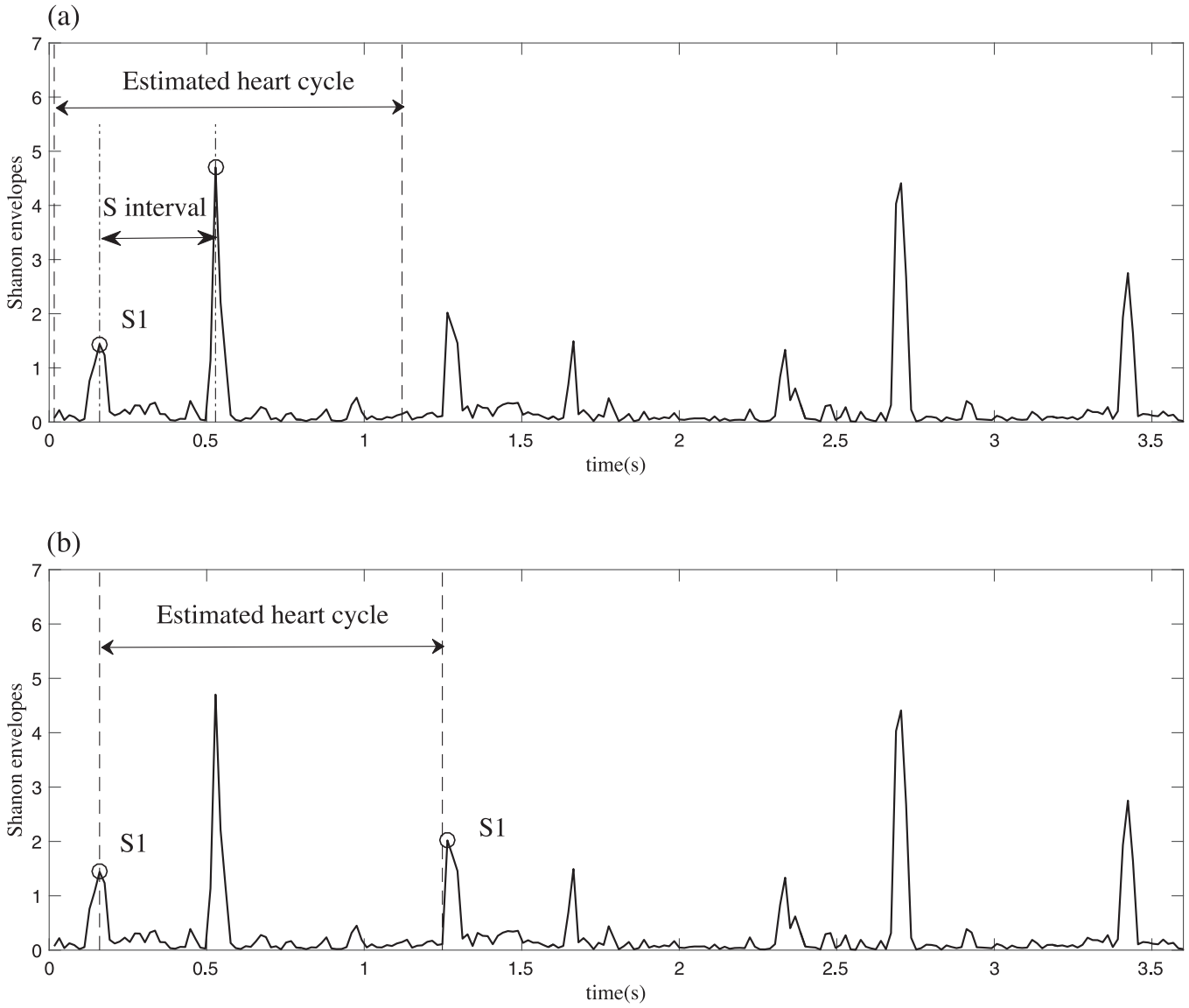

Fig. 2. (a) S1 detection of the first heart cycle; (b) S1 detection of the next heart cycle.

2.5. Feature extraction based on tensor decomposition

In order to extract the intrinsic structure information in the scaled spectrograms, the tensor decomposition method is used. First, the scaled spectrograms are concatenated as a three-way tensor $\mathcal{X} \in \mathbb{R}^{I_1 \times I_2 \times K}$. As in Kolda and Bader (2009), we denote tensor by calligraphic letters, i.e., $\mathcal{X}$. The scaled spectrograms are represented as $\{\mathcal{X}^{(k)} \in \mathbb{R}^{I_1 \times I_2}, k = 1, 2, \ldots, K\}$, K is the number of the samples in the training dataset.

To extract the intrinsic structure features in tensor $\mathcal{X}$, $\mathcal{X}$ is decomposed using the Tucker-2 tensor decomposition method (Liu, Glänzel, & De Moor, 2011). As shown in Fig 3, $\mathcal{X}$ is decomposed into a core tensor $\mathcal{G} \in \mathbb{R}^{J_1 \times J_2 \times K}$ and two factor matrices $U^{(1)} \in \mathbb{R}^{I_1 \times J_1}$ and $U^{(2)} \in \mathbb{R}^{I_2 \times J_2}$. The size of the core tensor $J_1 \times J_2 \times K$ is pre-defined according to the dimension of the extracted features.

Thus, the concatenated scaled spectrograms tensor $\mathcal{X}$ is decomposed as,

$$\mathcal{X} = \mathcal{G} \times_1 U^{(1)} \times_2 U^{(2)} + \mathcal{E} \tag{4}$$

where $\mathcal{E}$ is the residual tensor, $\mathcal{G} \times_1 U^{(1)}$ and $\mathcal{G} \times_2 U^{(2)}$ represent the mode-1 and mode-2 product of a tensor with a matrix, respectively. The mode-1 product $(\mathcal{G} \times_1 U^{(1)}) \in \mathbb{R}^{J_1 \times I_2 \times K}$ and mode-2 product $(\mathcal{G} \times_2 U^{(2)}) \in \mathbb{R}^{J_1 \times I_2 \times K}$ are computed by,

$$\begin{cases} (\mathcal{G} \times_1 U^{(1)})_{i_1 j_2 k} = \sum_{j_1=1}^{J_1} \mathcal{G}_{j_1 j_2 k} U^{(1)}_{i_1 j_1} \\ (\mathcal{G} \times_2 U^{(2)})_{j_1 i_2 k} = \sum_{j_2=1}^{J_2} \mathcal{G}_{j_1 j_2 k} U^{(2)}_{i_2 j_2} \end{cases} \tag{5}$$

In the decomposition process, the factor matrices $U^{(1)}$ and $U^{(2)}$ are generally forced to be orthogonal for algorithmic simplicity. Moreover, a more effective object function is considered for the feature extraction and classification task through maximizing the high order generalization of the Fisher score corresponding to the core tensor $\mathcal{G}$. And the object function is,

$$\varphi = \arg \max_{U^{(1)}, U^{(2)}} = \frac{\sum_{c=1}^{C} K_c ||\bar{\mathcal{G}}^{(c)} - \bar{\bar{\mathcal{G}}}||_F^2}{\sum_{k=1}^{K} ||\mathcal{G}^{(k)} - \bar{\mathcal{G}}^{(c_k)}||_F^2} \tag{6}$$

where C is the number of heart sound classes, K_c is the number of samples in class c, $\mathcal{G}^{(c)}$ is the average core tensor over samples in class c, $\bar{\bar{\mathcal{G}}}$ is the average core tensor over all training samples , $\bar{\mathcal{G}}^{(c_k)}$

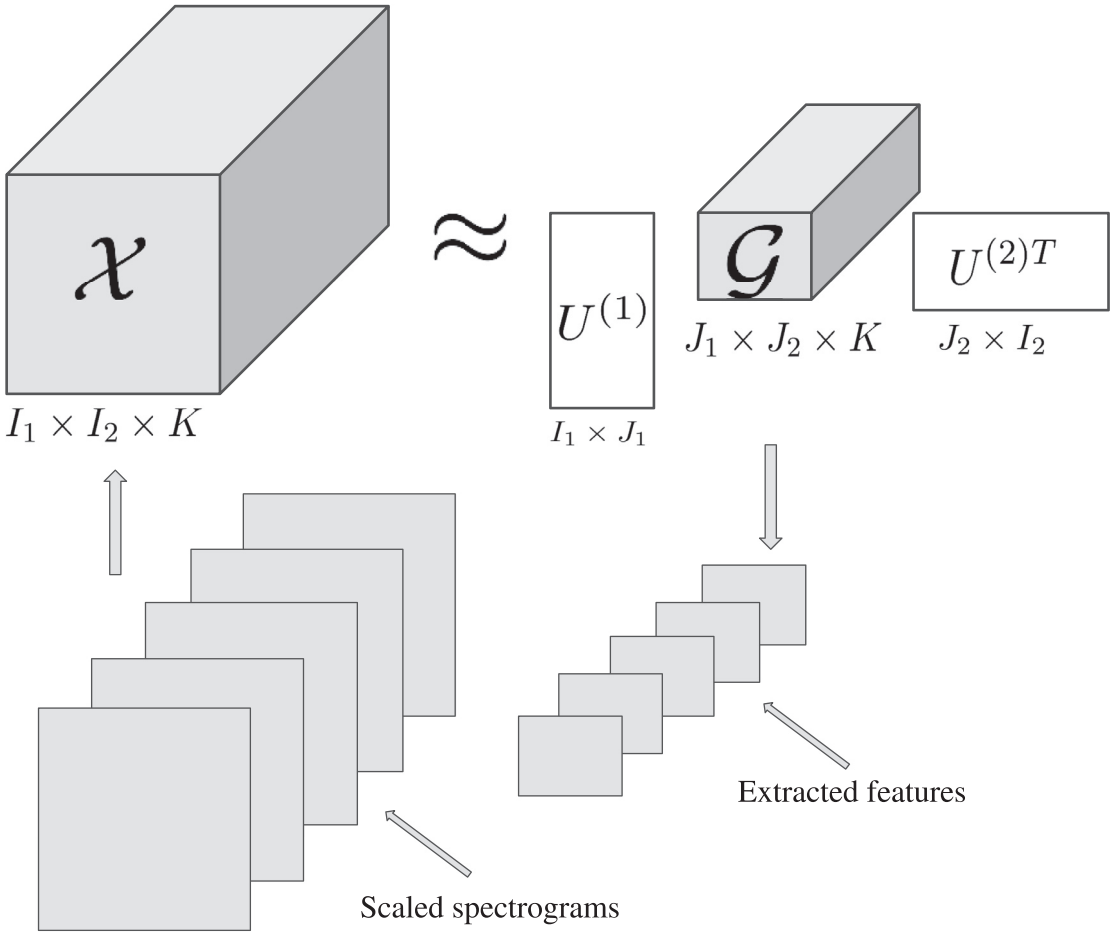

Fig. 3. Tucker-2 decomposition of the concatenated scaled spectrograms tensor.

is the average core tensor over class c_k and c_k is the class that the kth sample belongs to.

The objective function can be solved through the High Order Discriminant Analysis (HODA) algorithm (Lai, Xu, Yang, Tang, & Zhang, 2013). Given the concatenated scaled spectrograms tensor $\mathcal{X} \in \mathbb{R}^{I_1 \times, I_2, \times K}$ and the rank of the core tensor $\mathcal{G} \in \mathbb{R}^{J_1 \times J_2 \times K}$, the solution to the objective function Eq. (6) is obtained as follows:

1. Initialize $U^{(1)}$, $U^{(2)}$ randomly.
2. Set $n = 1$.
3. Calculate the within class scatter matrix A_n^w and between class scatter matrix A_n^b.
   The within class scatter matrix is obtained by,

$$\begin{aligned} A_n^w &= \sum_{k=1}^{K} ||\mathcal{G}^{(k)} - \bar{\mathcal{G}}^{(c_k)}||_F^2 \\ &= \sum_{k=1}^{K} \langle (\mathcal{X}^{(k)} - \bar{\mathcal{X}}^{(c_k)}) \times_d U^{(d)}, (\mathcal{X}^{(k)} - \bar{\mathcal{X}}^{(c_k)}) \times_d U^{(d)} \rangle_d \\ d &= \begin{cases} 2, & n = 1 \\ 1, & n = 2 \end{cases} \end{aligned} \tag{7}$$

   where $\bar{\mathcal{X}}^{(c_k)}$ is the average scaled spectrograms tensor over the class c. The tensor inner product of mode-d $\langle \mathcal{X}, \mathcal{X} \rangle_d$ is calculated by,

$$\langle \mathcal{X}, \mathcal{X} \rangle_d = \mathcal{X}_{(d)} \mathcal{X}_{(d)}^T \tag{8}$$

   where $\mathcal{X}_{(d)} \in \mathbb{R}^{I_d \times I_n K}$ is the mode-d matricized unfolding of tensor $\mathcal{X}$.

   The between class scatter matrix is obtained by,

$$\begin{aligned} A_n^b &= \sum_{c=1}^{C} K_c ||\bar{\mathcal{G}}^{(c)} - \bar{\bar{\mathcal{G}}}||_F^2 \\ &= \sum_{c=1}^{C} \langle \sqrt{K_c}(\bar{\mathcal{X}}^{(c)} - \bar{\bar{\mathcal{X}}}) \times_d U^{(d)}, \sqrt{K_c}(\bar{\mathcal{X}}^{(c)} - \bar{\bar{\mathcal{X}}}) \times_d U^{(d)} \rangle_d \\ d &= \begin{cases} 2, & n = 1 \\ 1, & n = 2 \end{cases} \end{aligned} \tag{9}$$

   where $\bar{\bar{\mathcal{X}}}$ is the average scaled spectrograms tensor over the training samples.
4. Apply the singular value decomposition for $A_n^w A_n^b$ as follows,

$$A_n^w A_n^b \approx U^{(n)} \Sigma^{(n)} V^{(n)T} \tag{10}$$

   Update $U^{(n)}$ using the decomposed result from Eq. (10).
5. Set $n = 2$, and repeat steps 3 to 4.
6. Repeat steps 2 to 5 until certain conditions are satisfied.

Thus, the orthogonal factor matrices $U^{(1)}$ and $U^{(2)}$ are obtained. According to the factor matrices, the features in the training samples are obtained based on tensor decomposition, as shown in Fig. 4(a). The decomposed tensor features, which are represented by the corresponding core tensor $\mathcal{G}$, are calculated by,

$$\mathcal{G} = \mathcal{X} \times_1 U^{(1)T} \times_2 U^{(2)T} \tag{11}$$

It is also important to generate the tensor representation feature in the test sample. As shown in Fig. 4(b), for a test scaled

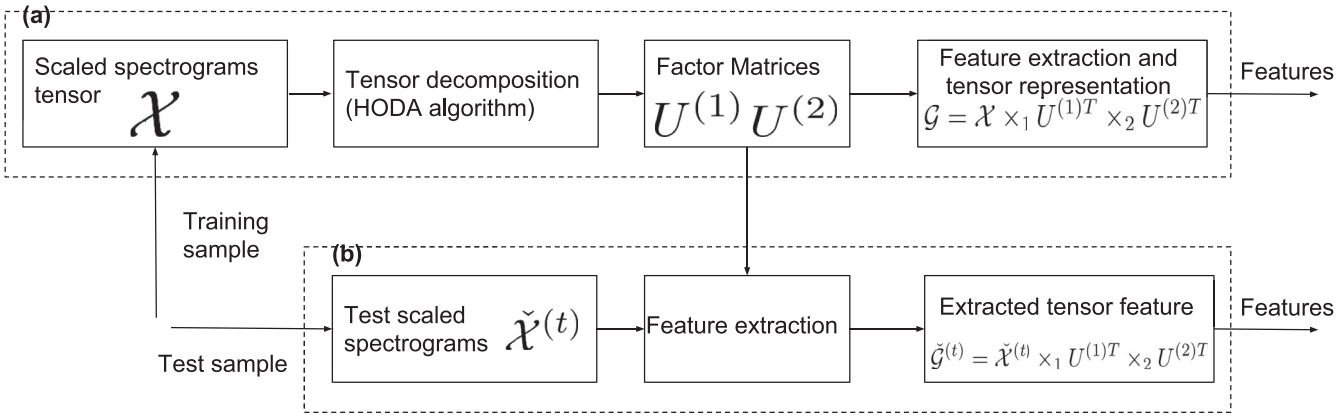

Fig. 4. The feature extraction process based on tensor decomposition. (a) Training sample feature extraction; (b) Test sample feature extraction.

spectrogram $\check{\mathcal{X}}^{(t)} \in \mathbb{R}^{I_1 \times I_2}$, the extracted feature $\check{\mathcal{G}}^{(t)} \in \mathbb{R}^{J_1 \times J_2}$ is represented as,

$$\check{\mathcal{G}}^{(t)} = \check{\mathcal{X}}^{(t)} \times_1 U^{(1)T} \times_2 U^{(2)T} \tag{12}$$

3. Experimental results

Using the Dataset-A and Dataset-B, the proposed method is compared with the three best methods in the PASCAL challenge competition: J48 (Gomes & Pereira, 2012), MLP (Gomes & Pereira, 2012), CS UCL (Deng & Bentley, 2012) and other methods: MBS (Oliveira, Gomes, & Jorge, 2014), 2D-PCA (Avendano-Valencia et al., 2010), SS-PLSR (Zhang et al., 2017) and SVM-DM (Deng & Han, 2016). Using the Dataset-C, the proposed method is also compared with the methods in the 2016 PhysioNet challenge: MLBI (Gokhale, 2016), TFF (Her & Chiu, 2016), PSA (Teo et al., 2016), DGRE (Thomae & Dominik, 2016) and other methods: 2D-PCA (Avendano-Valencia et al., 2010) and SS-PLSR (Zhang et al., 2017). SVM-DM is computationally complex both in training and testing process (Deng & Han, 2016) and is not suitable for large dataset, such as Dataset-C. 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 (Gomes & Pereira, 2012). MBS method is the improvement of J48. However, MBS is only suitable for Dataset-B (Oliveira et al., 2014). In the CS UCL method, the wavelet decomposition and spectrogram is used for feature extraction and the decision trees classifier is used for classification (Deng & Bentley, 2012). SVM-DM method merges the features into a unified representation using diffusion map. And SS-PLSR takes advantage of the label information of the heart sound categories during the dimension reduction process. The compared methods in the 2016 PhysioNet challenge try to learn features in the time-frequency domain through machine learning methods. The difference is that MLBI learns features through envelop, TFF uses clustering method, PSA detects the peaks in the spectrogram and DRGE generates features by the neural network.

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. Moreover, the methods SS-PLSR and SVM-DM are also evaluated by the SVM classifier. To compare with the 2D-PCA method, the features are extracted from the scaled spectrogram through 2D-PCA method and classified using the SVM method. To evaluate the tensor decomposed features, our approach using tensor decomposed features is experimented and the method is abbreviated as SS-TD for convenience.

3.1. Evaluation criteria

The evaluation criteria used in Dataset-A and Dataset-B 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. And the precision is defined as

$$P_r = \frac{TP}{TP + FP} \tag{13}$$

The sensitivity and specificity are used to evaluate the specific problem heart sound categories (Bentley, Nordehn, Coimbra, & Mannor, 2011):

$$\begin{aligned} S_e &= \frac{TP}{TP + FN} \\ S_p &= \frac{TN}{FP_{TN} + TN} \end{aligned} \tag{14}$$

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 Eq. (13) 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.

Since the test sample is not balanced, the normalized precision P_n is calculated to further evaluate the precision and defined by,

$$\begin{cases} P_n = \sum_{q=1}^{Q} W_q P_r^{(q)} \\ W_q = \frac{N_q}{\sum_{q=1}^{Q} N_q} \end{cases} \tag{15}$$

where Q is the total number of sample category in the testing dataset, N_q is the number of sample in the qth category and $P_r^{(q)}$ is the corresponding precision of the qth category.

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

Table 3
Results on Dataset-A.

Evaluation criteria	J48	MLP	CS UCL	SS-Method	SS-PLSR	2D-PCA	SVM-DM	SS-TD
Precision of *Normal*	0.25	0.35	0.46	**0.67**	0.60	0.56	0.62	**0.67**
Precision of *Murmur*	0.47	0.67	0.31	0.91	0.91	**1.00**	0.91	**1.00**
Precision of *Extra Heart Sound*	0.27	0.18	0.11	0.37	0.44	0.30	**1.00**	0.43
Precision of *Artifact*	0.71	0.92	0.58	0.76	**0.94**	**0.94**	0.64	0.80
Artifact sensitivity	0.63	0.69	0.44	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**
Artifact specificity	0.39	0.44	0.44	0.58	**0.64**	0.58	0.58	**0.64**
Youden index of *Artifact*	0.01	0.13	−0.09	0.58	**0.64**	0.58	0.58	**0.64**
F-score	0.20	0.20	0.14	0.28	0.30	0.26	**0.31**	0.30
Total precision	1.71	2.12	1.47	2.71	2.89	2.80	**3.17**	2.90
Normalized precision	0.45	0.59	0.40	0.72	**0.76**	0.75	**0.76**	**0.76**

other three categories. In addition, the F-score (Bentley et al., 2011) 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 $\beta = 0.9$.

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 (Bentley et al., 2011) 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 (Bentley et al., 2011) is used to evaluate the performance of the specific heart categories. It is used for both Dataset-A and Dataset-B:

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

Only the precision of the normal , abnormal heart sound categories and the overall score are evaluated in the 2016 PhysioNet challenge (Liu et al., 2016). The overall score is calculated by

$$O_v = \frac{P_{rn} + P_{ra}}{2} \tag{19}$$

where P_{rn} and P_{ra} are the corresponding normal and abnormal heart sounds precision. In order to have a fair comparison, Dataset-C is evaluated by the same criterion as the 2016 PhysioNet challenge.

3.2. Results of Dataset-A

The purpose of evaluating Dataset-A is to distinguish between the *Normal, Murmur, Extra Heart Sound* and *Artifact* categories. The results are analysed 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 experimental results are shown in Table 3.

The highest precision scores on the *Normal* and *Murmur* categories are all achieved by SS-TD method, the corresponding scores are 67% and 100%, respectively. On the *Normal* category, SS-method also achieves the best score. And 2D-PCA method achieves the best score on the *Murmur* category. Although the highest precision scores of the *Extra Heart Sound* and *Artifact* categories are not achieved by SS-TD method, the highest normalized precision 0.76 is achieved by SS-TD method. This indicates that SS-TD method is competitive and effective.

Since 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 highest scores are all achieved by SS-TD method and SS-PLSR method at the same time, the corresponding sensitivity, specificity and Youden's index are 1.00, 0.64 and 0.64, respectively. Overall, the SS-TD method has a good ability to detect the *Artifact* category.

Finally, the problematic heartbeats are evaluated in terms of the F-score. The results obtained by SS-Method and SS-TD methods are 0.28 and 0.30, which both are higher than that of the J48, MLP and CS UCL methods. Although the highest score is achieved by SVM-DM method, it is only 0.01 higher than that of SS-TD method. This indicates that our method is competitive at detecting the problematic heartbeats.

Note that, the highest precision score *Artifact* category is achieved by the 2D-PCA method which only indicates the false negative number is less. However, the performance of the *Artifact* category should be further evaluated. The total and normalized precision of 2D-PCA method is higher than that of the SS-Method, and this proves the effectiveness of extract the principal components of the scaled spectrogram. Moreover, the total and normalized precision of SS-TD method is higher than that of the 2D-PCA method, this proves the effectiveness of extract the intrinsic structure information of the scaled spectrogram.

3.3. Results of 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 results are mainly analysed 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 experimental results are shown in Table 4.

The highest precision score 83% on the *Normal* category is achieved by SS-TD method. Although the highest precision score of *Murmur* and *Extrasystole* categories are achieved by other methods. Our method SS-TD achieves the best normalized precision along with SVM-DM method. This indicates our method is competitive. It can be seen that the total and normalized precision of 2D-PCA method is higher than that of SS-Method. This indicates that the 2D-PCA method is effective for dimension reduction. Moreover, our method SS-TD has the best performance compared with SS-Method and 2D-PCA which shows the superior performance of extracting the intrinsic structure.

Table 4
Results on Dataset-B.

Evaluation criteria	J48	MBS	MLP	CS UCL	SS-Method	SS-PLSR	2D-PCA	SVM-DM	SS-TD
Precision of *Normal*	0.72	0.77	0.70	0.77	0.74	0.76	0.78	0.77	**0.83**
Precision of *Murmur*	0.32	0.38	0.30	0.37	0.66	0.65	0.57	**0.76**	0.70
Precision of *Extrasystole*	0.33	0.12	**0.67**	0.17	0.24	0.33	0.23	0.50	0.15
Heart problem sensitivity	0.22	0.29	0.19	**0.51**	0.24	0.34	0.41	0.34	0.49
Heart problem specificity	0.82	0.51	0.84	0.59	0.89	0.90	0.84	**0.95**	0.84
Youden index of heart problem	0.04	−0.20	0.02	0.01	0.13	0.24	0.24	0.29	**0.33**
Discriminant power	0.05	−0.20	0.04	0.09	0.24	0.36	0.30	**0.54**	0.39
Total precision	1.37	1.27	1.67	1.31	1.57	1.75	1.58	**2.03**	1.68
Normalized precision	0.60	0.63	0.62	0.63	0.67	0.69	0.68	**0.74**	**0.74**

Table 5
Results on Dataset-C.

Evaluation criteria	MLBI	TFF	PSA	DRGE	SS-Method	SS-PLSR	2D-PCA	SS-TD
Normal precision	0.80	0.84	0.75	**0.96**	0.92	0.95	0.94	0.92
Abnormal precision	0.75	0.87	0.79	0.83	0.82	0.82	0.82	**0.88**
Overall score	0.77	0.85	0.77	0.89	0.87	0.88	0.88	**0.90**

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 0.95 which is obtained by SVM-DM. And the best sensitivity score of the heart problems is 0.51 which is obtained by CS UCL method. However, our method has the best overall performance with a specificity 0.84 and sensitivity 0.49. This is evaluated by the Youden's index where our method SS-TD achieves the best scores 0.33 which is higher the other methods. Furthermore, the discriminant power 0.39 obtained by SS-TD method is only lower than that of SVM-DM method. These show that our method is competitive.

3.4. Results of Dataset-C

There are only two categories in Dataset-C, including normal and abnormal heart sound signals. The results are analysed from the precision and overall score of the two categories. Since only the training dataset is available, the evaluation results are obtained using 10-fold cross-validation method (Pedregosa et al., 2011). The experimental results are shown in Table 5.

The highest normal precision 96% is obtained by DRGE method. Our method SS-TD achieved the highest abnormal precision 88%. Moreover, the highest overall score 90% is also obtained by our method SS-TD. This indicates that our method which benefits from the features extraction based on scaled spectrogram and tensor decomposition is competitive. Note that the performance of SS-PLSR, 2D-PCA and SS-TD methods are all better than that of SS-Method. It shows that the importance of dimension reduction.

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. Their physiological and pathological information is contained in the heart cycles. It is reasonable to use the features extracted from the segmented heart cycles for automatic heart sound classification.

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. The size and time-frequency information retained requires a trade off. In order to select a suitable size of the scaled spectrogram, we conduct the experiments. Dataset-A and Dataset-B are used to evaluate the suitable size scale. The performances for the scaled spectrograms of size 40×30, 35×25, 30×20, 25×15, 20×10, 10×10 and 5×5 are compared, and the results are shown in Fig. 5(a). It can be seen that the normalized precision of Dataset-A and Dataset-B both achieves the best score at the size 30×20. Therefore, the size of the scaled spectrograms used in this paper is 30×20. Note that, the parameter of the radial basis function (RBF) kernel used in the SVM classifier is set to be the reciprocal of the dimension length for a fair comparison. After the size of the scaled spectrogram is selected, the parameter is tuned for a better performance.

In the tensor decomposition process, the size of the core tensor is pre-defined. A suitable size would be preferred for feature extraction. The smaller the size, the extracted core tensor features are more principal. However, too small size may loose some discriminate information. Therefore, the size of the core tensor also requires a trade off. We also select the size of the core tensor by experiments. The performances for the core tensor of size 25×20, 20×20, 20×15, 20×10, 15×10, 10×10 and 5×5 are compared, and the results are shown in Fig. 5(b). It can be seen that the normalized precision has little change beyond the size 10×10 for both Dataset-A and Dataset-B. In order to have a fair comparison, the size of dimension reduced features in 2D-PCA method should be the same with the core tensor. Therefore, the performances of 2D-PCA method with different size are also evaluated, as shown in Fig. 5(c). It can be seen that normalized precision of Dataset-A and Dataset-B for 2D-PCA also has little change beyond the size 10×10. Moreover, the normalized precision all has a substantial decline after the size is declined to 5×5. Therefore, the size of the core tensor and the size of the dimension reduced features in 2D-PCA method are all 10×10 in this paper.

The features extracted by the tensor decomposition method has better ability to reserve the intrinsic structure information in the scaled spectrograms. As shown in Fig. 6(d), the spectrogram after tensor decomposition still has strong energy in the position of fundamental heart sounds. However, the spectrogram after 2D-PCA does not have obvious energy in the position of the second fundamental heart sound, as shown in Fig. 6(c). As a comparison, the tensor decomposition has better ability to reserve the intrinsic structured information during the dimension reduction process.

Moreover, the proposed method SS-TD is evaluated on three datasets and compared with other related methods. The results show that our method is competitive. Furthermore, the experimen-

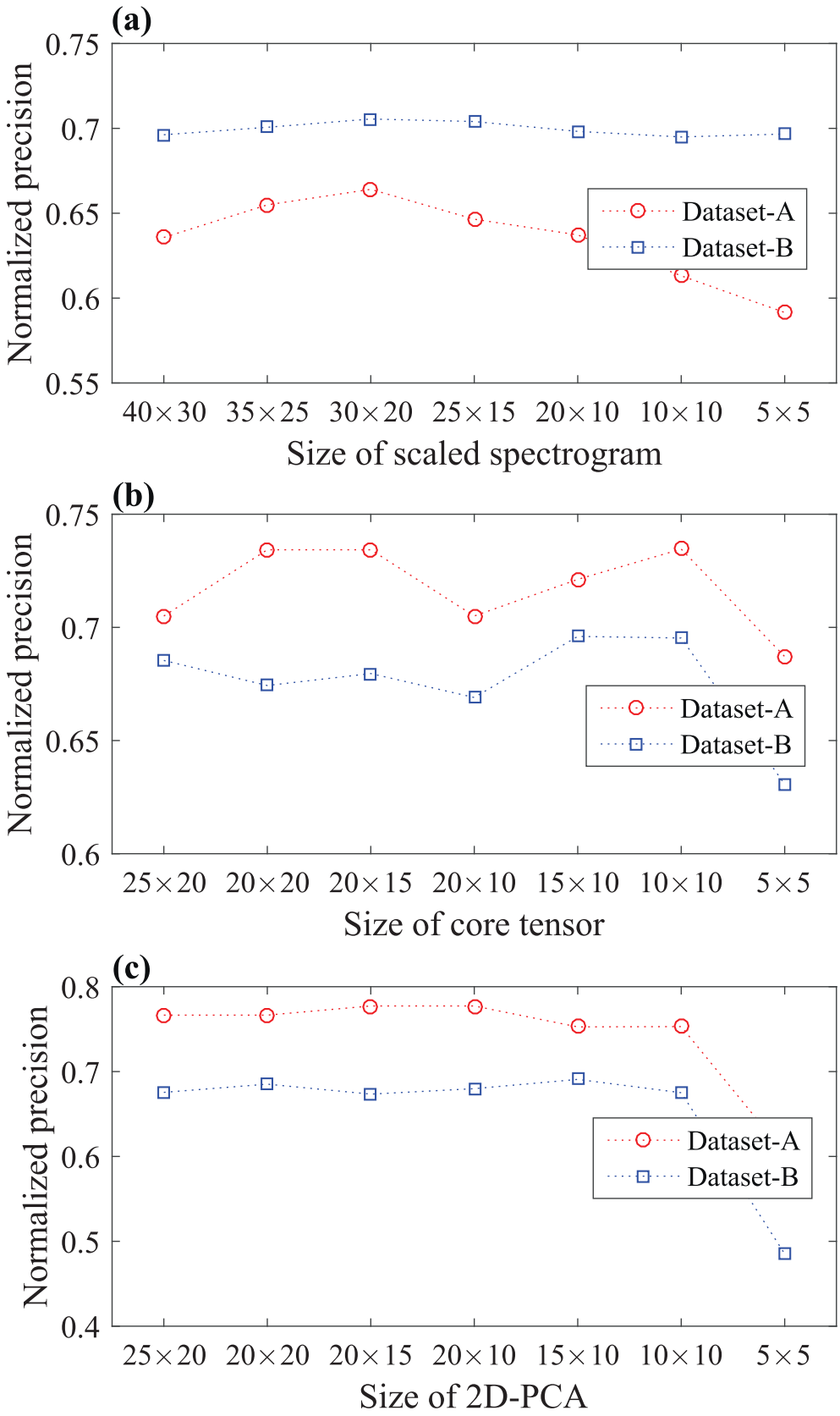

Fig. 5. The performance of different size. (a) The different size of scaled spectrograms; (b) The different size of core tensor for scaled spectrogram 30 × 20; (c) The different size of 2D-PCA for scaled spectrogram 30 × 20.

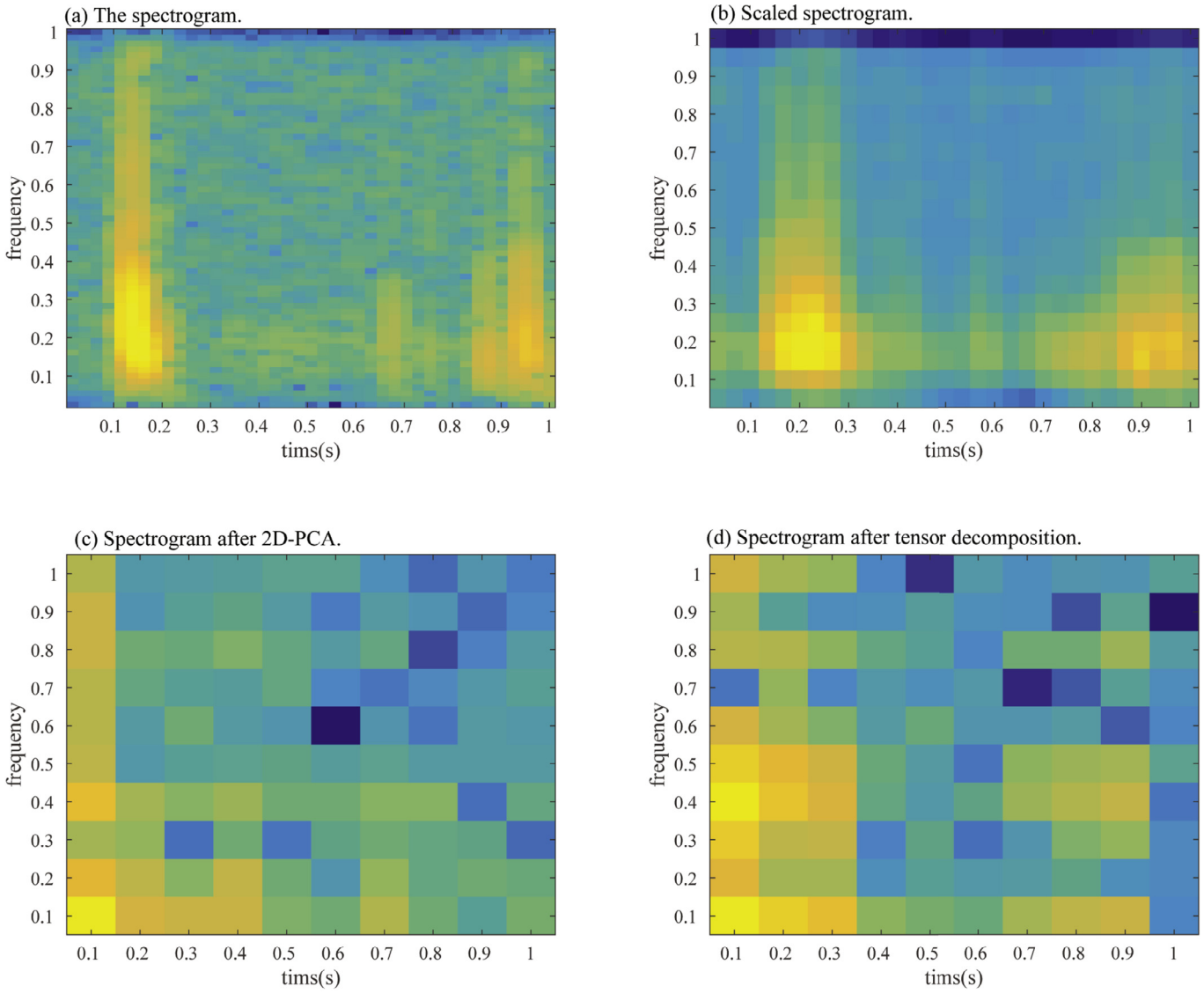

Fig. 6. The spectrogram of a normal heart sound signal. (a) The original spectrogram without scaling, the size is 65 × 45; (b) The spectrogram after scaling, the size is 20 × 30; (c) The spectrogram after dimension reduction using 2D-PCA, the size is 10 × 10; (d) The spectrogram after dimension reduction using tensor decomposition, the size is 10 × 10.

tal results also indicate that the scaled spectrogram and tensor decomposition based method has good ability to extract discriminative features for heart sound classification. And the good performance mainly benefits from the more efficient extraction method for the intrinsic structure information in the scaled spectrograms during the dimension reduction process. In this way, more useful physiological and pathological information is reserved. However, the intrinsic structure extraction is performed using tensor decomposition which is computationally complex. As a result, the factor matrices obtained in the training phase are time consuming. Nevertheless, the features in the testing phase are easy to compute since the factor matrices are already calculated in the training phase. Thus, the proposed method has practical significance.

5. Conclusion

This paper proposed a novel method for heart sound classification based on scaled spectrogram and tensor decomposition. This method can efficiently detect whether a PCG signal is problematic. Thus, it provides a valuable information for deciding whether further treatment is necessary. 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. During the dimension reduction process, the intrinsic structure information is extracted using the tensor decomposition method. Thus, the extracted features are more discriminative. The proposed framework is evaluated on the public datasets from the PASCAL classifying heart sounds challenge and 2016 PhysioNet challenge. The experiments show that our method is competitive compared with the baseline methods.

In future work, the internal and external noise of the heart sounds will be taken into consideration to improve the robustness of the proposed method. And more efficient and accurate heart cycle detection method is required to improve the performance. Moreover, proposing the feasible scaling methods for S-transform and wavelets would enhance the quality of the scaled spectrograms. In addition, more heart disease categories would be added to train the model. Thus the performance of the model is to be improved. And the method will be evaluated in real clinic through auscultation equipment.

Acknowledgments

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

References

Ari, S., Hembram, K., & Saha, G. (2010). Detection of cardiac abnormality from PCG signal using LMS based least square SVM classifier. *Expert Systems with Applications, 37*(12), 8019–8026.

Avendano-Valencia, L., Godino-Llorente, J., Blanco-Velasco, M., & Castellanos–Dominguez, G. (2010). Feature extraction from parametric time–frequency representations for heart murmur detection. *Annals of Biomedical Engineering, 38*(8), 2716–2732.

Bentley, P., Nordehn, G., Coimbra, M., & Mannor, S. (2011). The PASCAL classifying heart sounds challenge. See http://www.peterjbentley.com/heartchallenge/index.html.

Chen, T., Kuan, K., Celi, L. A., & Clifford, G. D. (2010). Intelligent heartsound diagnostics on a cellphone using a hands-free kit. In *AAAI spring symposium series* (pp. 26–31).

Deng, S.-W., & Han, J.-Q. (2016). Towards heart sound classification without segmentation via autocorrelation feature and diffusion maps. *Future Generation Computer Systems, 60*, 13–21.

Deng, Y., & Bentley, P. J. (2012). A robust heart sound segmentation and classification algorithm using wavelet decomposition and spectrogram. Workshop Classifying Heart Sounds, La Palma, Canary Islands.

Elharrar, V., & Surawicz, B. (1983). Cycle length effect on restitution of action potential duration in dog cardiac fibers. *American Journal of Physiology Heart and Circulatory Physiology, 244*(6), 782–792.

Gokhale, T. (2016). Machine learning based identification of pathological heart sounds. In *Computing in cardiology conference (CinC), 2016* (pp. 553–556). IEEE.

Gomes, E. F., & Pereira, E. (2012). Classifying heart sounds using peak location for segmentation and feature construction. Workshop Classifying Heart Sounds, La Palma, Canary Islands.

Hanna, I. R., & Silverman, M. E. (2002). A history of cardiac auscultation and some of its contributors. *The American Journal of Cardiology, 90*(3), 259–267.

Hariharan, B., Arbeláez, P., Girshick, R., & Malik, J. (2015). Hypercolumns for object segmentation and fine-grained localization. In *Conference on computer vision and pattern recognition* (pp. 447–456). IEEE.

Her, H.-l., & Chiu, H.-W. (2016). Using time-frequency features to recognize abnormal heart sounds. In *Computing in cardiology conference* (pp. 1145–1147). IEEE.

Herzig, J., Bickel, A., Eitan, A., & Intrator, N. (2015). Monitoring cardiac stress using features extracted from s1 heart sounds. *IEEE Transactions on Biomedical Engineering, 62*(4), 1169–1178.

Jiang, Z., & Choi, S. (2006). A cardiac sound characteristic waveform method for in-home heart disorder monitoring with electric stethoscope. *Expert Systems with Applications, 31*(2), 286–298.

Jolliffe, I. (2002). *Principal component analysis*. Wiley Online Library.

Kolda, T. G., & Bader, B. W. (2009). Tensor decompositions and applications. *SIAM Review, 51*(3), 455–500.

Lai, Z., Xu, Y., Yang, J., Tang, J., & Zhang, D. (2013). Sparse tensor discriminant analysis. *IEEE Transactions on Image Processing, 22*(10), 3904–3915.

Liang, H., Lukkarinen, S., & Hartimo, I. (1997). Heart sound segmentation algorithm based on heart sound envelogram. In *Computers in cardiology conference* (pp. 105–108). IEEE.

Liu, C., Springer, D., Li, Q., Moody, B., Juan, R. A., Chorro, F. J., … Johnson, A. E., et al. (2016). An open access database for the evaluation of heart sound algorithms. *Physiological Measurement, 37*(12), 2181.

Liu, X., Glänzel, W., & De Moor, B. (2011). Hybrid clustering of multi-view data via tucker-2 model and its application. *Scientometrics, 88*(3), 819–839.

Livanos, G., Ranganathan, N., & Jiang, J. (2000). Heart sound analysis using the s transform. In *Computers in cardiology conference* (pp. 587–590). IEEE.

Maglogiannis, I., Loukis, E., Zafiropoulos, E., & Stasis, A. (2009). Support vectors machine-based identification of heart valve diseases using heart sounds. *Computer Methods and Programs in Biomedicine, 95*(1), 47–61.

Marques, N., Almeida, R., Rocha, A. P., & Coimbra, M. (2013). Exploring the stationary wavelet transform detail coefficients for detection and identification of the s1 and s2 heart sounds. In *Computing in cardiology conference* (pp. 891–894). IEEE.

Moukadem, A., Dieterlen, A., Hueber, N., & Brandt, C. (2013). A robust heart sounds segmentation module based on s-transform. *Biomedical Signal Processing and Control, 8*(3), 273–281.

Oliveira, S. C., Gomes, E. F., & Jorge, A. M. (2014). Heart sounds classification using motif based segmentation. In *International database engineering & applications symposium* (pp. 370–371). ACM.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., … Dubourg, V., et al. (2011). Scikit-learn: Machine learning in python. *Journal of Machine Learning Research, 12*(Oct), 2825–2830.

Quiceno-Manrique, A., Godino-Llorente, J., Blanco-Velasco, M., & Castellanos–Dominguez, G. (2010). Selection of dynamic features based on time–frequency representations for heart murmur detection from phonocardiographic signals. *Annals of Biomedical Engineering, 38*(1), 118–137.

Rangayyan, R. M., & Lehner, R. J. (1986). Phonocardiogram signal analysis: a review. *Critical Reviews in Biomedical Engineering, 15*(3), 211–236.

Safara, F., Doraisamy, S., Azman, A., Jantan, A., & Ramaiah, A. R. A. (2013). Multi-level basis selection of wavelet packet decomposition tree for heart sound classification. *Computers in Biology and Medicine, 43*(10), 1407–1414.

SaraçOğLu, R. (2012). Hidden Markov model-based classification of heart valve disease with PCA for dimension reduction. *Engineering Applications of Artificial Intelligence, 25*(7), 1523–1528.

Schmidt, S., Holst-Hansen, C., Graff, C., Toft, E., & Struijk, J. J. (2010). Segmentation of heart sound recordings by a duration-dependent hidden Markov model. *Physiological Measurement, 31*(4), 513.

Senior, K. (2011). Smart phones: New clinical tools in oncology? *The Lancet Oncology, 12*(5), 429–430.

Soeta, Y., & Bito, Y. (2015). Detection of features of prosthetic cardiac valve sound by spectrogram analysis. *Applied Acoustics, 89*, 28–33.

Springer, D., Tarassenko, L., & Clifford, G. (2015). Logistic regression-HSMM-based heart sound segmentation. *IEEE Transactions on Biomedical Engineering, 63*, 822–832.

Sun, S., Jiang, Z., Wang, H., & Fang, Y. (2014). Automatic moment segmentation and peak detection analysis of heart sound pattern via short-time modified Hilbert transform. *Computer Methods and Programs in Biomedicine, 114*(3), 219–230.

Teo, S.-K., Yang, B., Feng, L., & Su, Y. (2016). Power spectrum analysis for classification of heart sound recording. In *Computing in cardiology conference (CinC), 2016* (pp. 1169–1172). IEEE.

Thomae, C., & Dominik, A. (2016). Using deep gated RNN with a convolutional front end for end-to-end classification of heart sound. In *Computing in cardiology conference* (pp. 625–628). IEEE.

Uğuz, H. (2012). A biomedical system based on artificial neural network and principal component analysis for diagnosis of the heart valve diseases. *Journal of Medical Systems, 36*(1), 61–72.

Wang, P., Kim, Y., Ling, L., & Soh, C. (2006). First heart sound detection for phonocardiogram segmentation. In *Engineering in medicine and biology conference* (pp. 5519–5522). IEEE.

Zhang, W., Han, J., & Deng, S. (2017). Heart sound classification based on scaled spectrogram and partial least squares regression. *Biomedical Signal Processing and Control, 32*, 20–28.

Zheng, Y., Guo, X., & Ding, X. (2015). A novel hybrid energy fraction and entropy-based approach for systolic heart murmurs identification. *Expert Systems with Applications, 42*(5), 2710–2721.