# Research on heart and lung sound separation method based on DAE–NMF–VMD

Wenhui Sun[1,2], Yipeng Zhang[1,2] and Fuming Chen[2]*

*Correspondence:
cfm5762@126.com

[1] Medical Security Center, The 940th Hospital of Joint Logistics Support Force of Chinese People's Liberation Army, Lanzhou 730050, Gansu, China
[2] School of Information Engineering, Gansu University of Chinese Medicine, Lanzhou 730000, Gansu, China

**Abstract**

Auscultation is the most effective method for diagnosing cardiovascular and respiratory diseases. However, stethoscopes typically capture mixed signals of heart and lung sounds, which can affect the auscultation effect of doctors. Therefore, the efficient separation of mixed heart and lung sound signals plays a crucial role in improving the diagnosis of cardiovascular and respiratory diseases. In this paper, we propose a blind source separation method for heart and lung sounds based on deep autoencoder (DAE), nonnegative matrix factorization (NMF) and variational mode decomposition (VMD). Firstly, DAE is employed to extract highly informative features from the heart and lung sound signals. Subsequently, NMF clustering is applied to group the heart and lung sounds based on their distinct periodicities, achieving the separation of the mixed heart and lung sounds. Finally, variational mode decomposition is used for denoising the separated signals. Experimental results demonstrate that the proposed method effectively separates heart and lung sound signals and exhibits significant advantages in terms of standardized evaluation metrics when compared to contrast methods.

**Keywords:** Heart and lung sound separation, Deep autoencoder, Variational mode decomposition, Nonnegative matrix factorization, Deep learning

## 1 Introduction

Under normal circumstances, the frequency of heart sound signals falls within the range of 20 to 150Hz [1], while lung sound signals fall within the range of 50 to 2500Hz [2]. There exists a frequency overlap region between heart and lung sound signals, leading to mutual interference between them. When medical professionals use stethoscopes for auscultation, noise from the friction of the stethoscope with clothing, ambient environmental noise and the operation of the instrument all get collected along with heart and lung sounds into electronic stethoscopes. This significantly diminishes the effectiveness of auscultation and diagnosis. In recent years, research on classification algorithms for lung sounds has increased [3–6]. However, the primary challenge faced in current lung sound recognition research is that traditional classification methods struggle to extract crucial information from lung sound features, resulting in suboptimal recognition performance. Additionally, lung sound classification methods have a high dependence on data, and publicly available lung sound datasets on the internet often contain heart

sound interference. Pure lung sound data is scarce and challenging to obtain, making recognition networks prone to overfitting and less capable of achieving precise and efficient classification and recognition. To better implement lung sound classification algorithms and diagnose medical conditions, it is essential to perform preprocessing through heart and lung sound separation.

To date, researchers worldwide have developed various heart and lung sound separation algorithms. These include methods based on wavelet transformations [7, 8], but they suffer from poor adaptability and ineffective suppression of interference factors. Independent component analysis (ICA) and its extensions have also been explored [9, 10], requiring at least two sensors and therefore not suitable for single-channel devices. In recent years, nonnegative matrix factorization (NMF) has been used to separate different sound sources [11–13], with its ability to handle overlapping frequency bands recognized. Deep learning has also been employed in source separation, where these deep learning models directly decompose mixed sources into target sources, and their effectiveness surpasses that of NMF [14–16]. Since it is challenging to acquire pure heart and lung sounds as training data due to the limitations of stethoscope data collection, this paper proposes an unsupervised learning approach using deep autoencoders (DAE) and variational mode decomposition (VMD) to separate mixed heart and lung sound signals. The algorithm first utilizes a DAE model to extract highly informative representations of the mixed sounds. By applying the periodic clustering algorithm to the potential representation, the mixed cardiopulmonary sounds are separated. VMD boasts a clear mathematical theoretical framework and unique advantages in noise robustness and avoiding mode mixing, compared to other classical methods [17]. Therefore, VMD is employed to denoise the separated heart and lung sound signals. In contrast to other deep learning-based methods, the advantage of this paper's approach is that it does not require labeled training data. Leveraging periodic structures, it provides better separation performance compared to traditional methods. The main contributions of this study are summarized below.

1. Blind source separation heart and lung sound model based on deep autoencoders, nonnegative matrix decomposition and variational mode decomposition was established. Use the autoencoder to extract the potential height expression of the heart and lung sound signals, and then send the obtained potential height expression into a sparse nonnegative matrix to perform clustering according to the different periods of the heart sound signal and lung sound signal to achieve heart and lung sound separation. Finally, the obtained heart and lung sound signals are denoised and enhanced using variational mode decomposition to obtain clean heart and lung sound signals.

2. The blind source separation heart and lung sound model based on deep autoencoders, nonnegative matrix decomposition and variational mode decomposition uses signal distortion rate (SDR), perceptual evaluation of speech quality (PESQ) and short-term objective intelligibility (STOI). These three standard evaluation indicators to evaluate the effectiveness of the model were observed.

3. Conduct comparative experiments with other heart and lung sound separation algorithms. The experimental results verify the effectiveness of using the proposed algorithm for heart and lung sound separation. The spectrograms and spectrograms after

Sun *et al. EURASIP Journal on Advances in Signal Processing*      (2024) 2024:59

Page 3 of 17

experiments using real heart and lung sound data prove that using this algorithm to separate heart and lung sounds is very effective and the separated signals have only small background noise interference.

## 2 Method

### 2.1 Model frame

The model framework adopted in this study is shown in Fig. 1. Firstly, the feature representation is obtained by training DAE model, and the periodic coding matrix is generated by discrete Fourier transform of the feature representation. Then, the clustering results are obtained by sparse NMF clustering, and then the separated heart sounds and lung sounds are encoded. Finally, the preliminary separated heart sounds and lung sounds are denoised by VMD to obtain pure heart sounds and pure lung sounds.

### 2.2 Heart–lung sound separation model

#### 2.2.1 Deep autoencoder (DAE)

DAE includes encoder and decoder, and the framework of DAE model is shown in Fig. 2. The encoder compresses the input data into a low dimension to provide a feature representation of a smaller dimension, and the decoder decodes this low-dimensional representation into an output as similar as possible to the original input. Both encoder and decoder are composed of full-scroll layers. In the training process, the purpose of DAE is to minimize the reconstruction error between input and output.

The internal structure of the encoder and decoder of DAE is shown in Fig. 3, which is a convolution and deconvolution process from left to right. The input signal is sent to the encoder composed of convolution layer and activation function to get the feature representation, and then, the reconstructed signal is obtained after passing through the encoder composed of deconvolution and activation function.
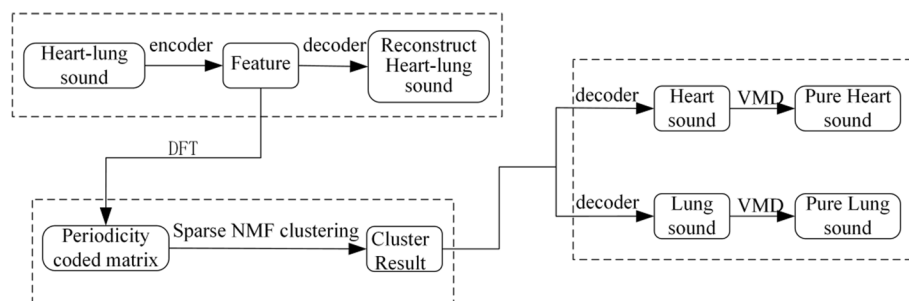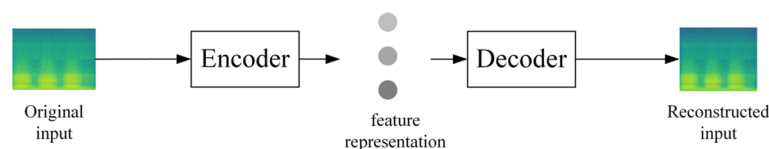


**Fig. 1** DAE–NMF–VMD algorithm block diagram
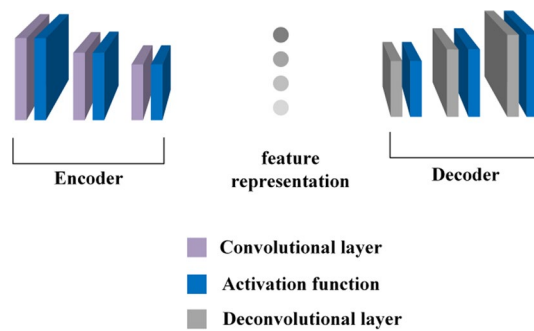


**Fig. 2** DAE model architecture

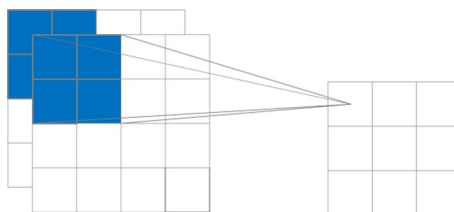**Fig. 3** Decoder structure and encoder structure



**Fig. 4** Convolution operation

The encoder in the DAE algorithm is composed of convolutional units that perform convolution functions, and the computation process is as shown in Eq. (1).

$$f_j^{(1)} = \sigma\left(C\left(W_{ji}^{(0)}, x\right) + b_j^{(0)}\right)$$

$$f_j^{(k+1)} = \sigma\left(\sum_{i=1}^{I} C\left(W_{ji}^{(k)}, f^{(k)}\right) + b_j^{(k)}\right) \tag{1}$$

In Eq. (1), $C$ stands for convolution process, $f_j^{(k)}$ represents the j-th feature map in the $k$-th layer, and I denotes the total number of channels. Each coding layer has $j$ convolution kernels, and the convolution kernel size is $L^*1$. $W_{ji}$ represents the $i$-th channel of $W_j$. Each neuron $f_j^{k+1}$ in the feature map of the $(k+1)$-th layer is calculated as the weighted sum of elements obtained by performing convolution operations with the receptive fields of all previous feature maps $f^{(k)}$, using the weights from $W_j$, and $b_j^{(k)}$ represents the bias. The corresponding convolution operation is illustrated in Fig. 4. The local area of input data is weighted and summed by sliding convolution kernel to extract the feature representation of this area.

The decoder is composed of deconvolution units, and the calculation process is shown in formula (2).

$$f_j^{(k+1)} = \sigma\left(\sum_{i=1}^{I} D(W_{ji}^{(k)}, f^{(k)}) + b_j^{(k)}\right)$$

$$\hat{x} = \sigma\left(D(W_{ji}^{(K_{\text{All}})}, f^{(K_{\text{All}})}) + b_j^{(K_{\text{All}})}\right) \tag{2}$$
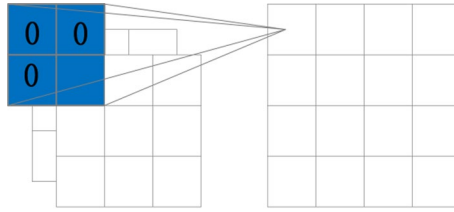
**Fig. 5** deconvolution operation

Here, $K_{All} = K_E + K_D$ represents the total number of layers in the DAE, where $K_E$ and $K_D$ are the numbers of layers in the encoder and decoder, respectively. $D$ is deconvolution process. Each decoding layer has $j$ convolution kernels, and the convolution kernel size is $L*1$. $W_{ji}$ represents the i-th channel of $W_j$. In the $k+1$-th layer's feature map, each neuron $f_j^{(k+1)}$ is the weighted sum of the element-wise deconvolution of $W_j$ with the receptive field from all previous feature maps $f^{(k)}$, with $b_j^{(k)}$ denoting the bias. The corresponding deconvolution operation is shown in Fig. 5. In the deconvolution process, all feature maps $f^{(k)}$ of the $k$-th layer are zero-filled, and then the deconvolution process is carried out to reconstruct the data with the same size as the original signal.

Initially, the mixed heart and lung sound signal is transformed into the frequency and phase components using short-time Fourier transform (STFT). Then, the spectral features are converted into logarithmic power spectra (LPS). $X = [x_1,...,x_n,...,x_N]$ represents the input, where $N$ is the number of frames in $X$. DAE then encodes the mixed heart and lung sound LPS through encoder, transforming $X$ into a matrix representing feature representations $F^{(K_E)} = [f_1^{K_E}, \cdots, f_n^{K_E}, \cdots, f_N^{K_E}]$. The decoder reconstructs the matrix of feature representations back into the original spectral features. The parameters of the DAE are trained using the back-propagation algorithm to minimize mean squared error (MSE). Due to the input and output being the same, the DAE is trained in an unsupervised manner.

### 2.2.2 NMF periodic clustering method

Because the heart sound signal and lung sound signal have different periods, the heart–lung sound is separated and mixed by using the different periods of the heart sound signal and lung sound signal. By training DAE model, we get the potential features, and the set of potential feature representation and time series is matrix $F$. We transpose the original L to obtain $S^{mix} = F^T$. Based on $S^{mix}$, the entire set of neurons is divided into two groups: one corresponds to heart sounds, and the other corresponds to lung sounds. To analyze the periodicity of each submatrix $s_j^{mix}$, we apply discrete Fourier transform (DFT) to $s_j^{mix}$[18], forming a periodic encoding matrix $P = [p_1,...,p_j,...,p_M]$.

$$\text{where } p_j = \left| \text{DFT}\left(s_j^{mix}\right) \right| \tag{3}$$

sparse NMF clustering is employed to cluster the vectors in $P$ into two groups. Equation (4) illustrates the NMF clustering process, which is achieved by minimizing the error function. Based on the encoding matrix $H_P$ (the transpose of $P^T$) with the highest scores, the clustering assignment of $S^{mix}$ can be determined.

Sun *et al. EURASIP Journal on Advances in Signal Processing*      (2024) 2024:59

Page 6 of 17

$$H_P = \arg\min[\|P - W_P H_P\|^2 + \lambda \|H_P\|_1] \tag{4}$$

where $W_P$ represents the cluster centroid, $H_P$ is the transpose of matrix p, and $H_P = [h_1,...,h_p,...,h_M]$ stands the cluster members. $\lambda$ represents the sparsity penalty factor, this study selects an $\lambda$ value of 1. $\|\cdot\|_1$ represents the *L1* norm, and $\|\cdot\|^2$ represents the Frobenius distance. On the basis of the highest score in $H_P$, the cluster allocation of $S^{\mathrm{mix}}$ can be determined. The separation of heart sounds and lung sounds is realized by stopping the submatrix that does not belong to the target, and the separation results are $S^c$ and $S^r$. After obtaining the coding matrix of each source, we decode it to get the separated heart sounds and lung sounds.

By training DAE model, we get the potential feature $F$, and then, the coding matrix is transformed into a coding matrix $P$ with obvious periodicity by discrete Fourier transform (DFT). A sparse nonnegative matrix factorization (NMF) clustering method is employed to separate the encoding matrix $P$ into representative encoding matrices corresponding to heart and lung sound signals. Then, the source encoding matrix is reconstructed using the encoder. Finally, the obtained heart sound LPS (log power spectrum) sequence and lung sound LPS sequence are transformed into heart and lung sound signals using the inverse short-time Fourier transform (ISTFT).

### 2.2.3 Variational mode decomposition (VMD)

The VMD method decomposes the signal $x$ into a series of intrinsic mode functions (IMFs) with limited bandwidth, adaptively updating the optimal center frequencies and bandwidths for each IMF. The constrained variational problem generated based on $u_1, u_2, ..., u_k$ and the predetermined scaling parameter $K$ is shown in Eq. (5):

$$\begin{cases} \min_{\{u_k\},\{w_k\}} \left\{ \sum_k \left\| \partial_t \left[ \left( \delta(t) + \frac{j}{\pi t} \right) * u_k(t) \right] e^{-jw_k t} \right\|^2 \right\} \\ s.t. \qquad \sum_k u_k = s \end{cases} \tag{5}$$

where $\{u_k\} = \{u_1,...,u_k\}$ and $\{w_k\} = \{w_1,...,w_k\}$ represent the decomposed IMF components and the center frequencies of each component, respectively. $\partial(t)$ represents the partial derivative of the function with respect to time $t$, $\delta(t)$ is the unit impulse function, and * denotes convolution operation.

The introduction of the enhanced Lagrangian $\zeta$, as given below, allows for the transformation of the constrained variational problem into an unconstrained variational problem:

$$\zeta(u_k, w_k, \lambda) = \alpha \sum_k \left\| \partial_t \left[ \left( \delta(t) + \frac{j}{\pi t} \right) \right] e^{-jw_k t} \right\|^2 + \left\| s(t) - \sum_k u_k(t) \right\|^2 + \left\langle \lambda(t), s(t) - \sum_k u_k(t) \right\rangle \tag{6}$$

In Eq. (6), $\lambda$ and $\alpha$ are, respectively, the Lagrange multipliers and second-order penalty factors. The solution to the original minimization problem is now found as a saddle point of the enhanced Lagrangian during a series of alternating-direction optimizations of the multipliers, known as the method of alternating direction for multipliers.

Sun *et al. EURASIP Journal on Advances in Signal Processing*    (2024) 2024:59

Page 7 of 17

The derivatives of the IMF components $u_k$ and center frequencies $w_k$ are then derived as shown in Eqs. (7) and (8):

$$\hat{u}_k^{n+1}(\omega) = \frac{\hat{s}(\omega) - \sum\limits_{i \neq k} \hat{u}_i(\omega) + \hat{\lambda}(\omega)/2}{1 + 2\alpha(\omega - \omega_k)^2} \tag{7}$$

$$\omega_k^{n+1} = \frac{\int_0^\infty \omega \left| u_k^{n+1}(\omega) \right|^2 d\omega}{\int_0^\infty \left| u_k^{n+1}(\omega) \right|^2 d\omega} \tag{8}$$

In the above equations, $\omega$ represents frequency, and $\hat{u}_k^{n+1}(\omega)$, $\hat{s}(\omega)$ and $u_k^n(t)$ represent the Fourier transforms of $s(t)$, $u(t)$ and $\lambda(t)$.

By using the algorithm described above, for a convergence tolerance $e > 0$, the decomposition stops when Eq. (9) is satisfied, and the final modal components and their center frequencies $w_k$ are obtained.

$$\sum_k \left\| \hat{u}_k^{n+1} - \hat{u}_k^n \right\|_2^2 / \left\| \hat{u}_k^n \right\|_2^2 < e \tag{9}$$

The VMD algorithm is used to decompose the heart sound and lung sound signals separated by DAE into a series of IMFs with finite bandwidth. The choice of the predetermined number of decomposition modes $K$ and the penalty factor $\alpha$ directly affects the accuracy of the VMD decomposition results. Therefore, selecting suitable values for $K$ and $\alpha$ is crucial for obtaining pure heart and lung sounds.

(1) Selection of $K$.

 In the VMD algorithm, the value of $K$ represents the number of IMF components into which the signal is decomposed. If the optimal value of $K$ is obtained, it means that the center frequency distribution between adjacent IMF components is reasonable, and there will be no similar or mixed results in the decomposition. In this study, the value of $K$ is determined using empirical mode decomposition (EMD) [19]. Based on experiments, this study selects a value of $K$ as 7.

(2) Selection of $\alpha$.

 $\alpha$ is another important parameter to be set during the VMD decomposition process, and it determines the bandwidth of the IMF. A larger $\alpha$ value results in smaller bandwidths for each IMF component obtained by VMD. The value of $\alpha$ should neither be too large nor too small. Additionally, it is found that within a relative range, this parameter has a minimal impact on the results. For heart and lung sound signals, this study selects an $\alpha$ value of 1500.

Using the above algorithm, the heart and lung sound signals are decomposed into seven IMF components by the VMD algorithm, and the high-frequency IMF components are summed to obtain denoised heart or lung sound signals.

The separation algorithm is as follows:

Algorithm 1:Cardiopulmonary sound separation

---

**Input:** mixed heart-lung $X$

**Output:** heart coded matrix $S^c$, lung coded matrix $S^r$

1:Feature extraction of vector $[x_1, \cdots, x_n, \cdots, x_N]$ in X.

2:Obtain the characteristic matrix $F^{(K_E)} = [f_1^{K_E}, \cdots, f_n^{K_E}, \cdots, f_N^{K_E}]$.

3:Transpose matrix $F$ to get $S^{mix}$.

4: for $j = 1$ to $M$ do

5:     $p_j = |\mathrm{DFT}(s_j^{mix})|$

6: end for

7: Perform clustering on vectors $[p_1, \cdots, p_j, \cdots, p_M]$ in P

8: Obtain labels of P: c = $[c_1, \ldots, c_M]$, where there are only two labels for $c_j$ {c or r}.

9: Set $\sigma_{\min}$, where $\sigma_{\min}$ is a vector whose coefficients are the latent neuron's minimum values

10: foreach $t$ = [c; r] do

11:     Initialize $S^t = S^{mix}$

12:     for $j$ =1 to M do

13:        if $c_j \neq t$ then

14:           do $s_j^t = \sigma_{\min}$

15: end if

16: end for

17: return $S^t$

18: end foreach

---

## 3  Experiment and discussion

In this section, the experiment and performance evaluation of the proposed heart and lung sound separation method are discussed. The evaluation metrics used include signal-to-distortion ratio (SDR), perceptual evaluation of speech quality (PESQ) and short-time objective intelligibility (STOI) to validate the effectiveness of the proposed method.

### 3.1  Experimental parameters

The proposed method takes the spectrogram of the mixed signal as input and outputs separated heart and lung sound signals (Table 1). The DAE model's structure is as follows:

**Table 1** DAE model structure

|         | Convolution layer | Filter | Convolution kernel |
|---------|-------------------|--------|--------------------|
| Encoder | First layer       | 64     | $1 \times 4$       |
|         | Second layer      | 32     | $1 \times 3$       |
|         | Third layer       | 16     | $1 \times 3$       |
|         | Fourth layer      | 8      | $1 \times 2$       |
| Decoder | First layer       | 8      | $1 \times 2$       |
|         | Second layer      | 16     | $1 \times 3$       |
|         | Third layer       | 32     | $1 \times 3$       |
|         | Fourth layer      | 64     | $1 \times 4$       |
|         | Fifth layer       | 1      | $1 \times 1$       |

The DAE model uses a stride of 1 for both the convolutional and deconvolutional units. Activation function is ReLU. The optimizer utilized is Adam. Unsupervised NMF method is employed as the baseline, with the L2 norm serving as the loss function.

### 3.2 Experimental data

The experimental data in this study were obtained from real heart and lung sounds. The dataset used in this research is sourced from publicly available datasets [20, 21]. Heart and lung sounds were collected under conditions with relatively low noise and can be considered as clean heart and lung sound signals. These clean heart and lung sounds were linearly mixed to create the mixed heart–lung sound signals. Assuming $x_c$ represents the heart sound signal and $x_r$ represents the lung sound signal, the mixed signal takes the form: signal $= x_c + ax_r$, where $a$ is a coefficient. Based on the signal-to-noise ratio formula (Eq. 10), we have:

$$10 \lg \frac{p_c}{p_r} = 10 \lg \frac{(x_c)^2}{(ax_r)^2} = r \tag{10}$$

In the equation, $r$ represents the logarithm of the ratio between the energy of the output mixed signal's heart sound and lung sound. If the energy of the output heart sound and lung sound is equal, then $r$ equals 0. The coefficient $a$ can be determined by using the equation provided. Finally, the signal is normalized.

### 3.3 Evaluation indicators

For the heart and lung sound separation method studied in this paper, we obtain separated heart and lung sounds. We use pure heart and lung sounds as references to calculate the separation performance, and we employ three standardized evaluation metrics: signal-to-distortion ratio (SDR), perceptual evaluation of speech quality (PESQ) and short-time objective intelligibility (STOI) to assess the separation performance. The formulas for calculating SDR, PESQ and STOI are as follows.

(1)   Formula for calculating SDR:

SDR was put forward by Vincent and others in 2006, and it is an evaluation index of blind source separation task [22]. In source separation tasks, there are three types of noise: interference due to missed separation ($e_{\text{interf}}$), artifacts due to the reconstruction process ($e_{\text{artif}}$) and residual noise ($e_{\text{noise}}$). Here, $\hat{s}(t)$ represents the estimated result, and $s_{\text{target}}(t)$ represents the target. SDR is calculated as shown in formula (12), where $\|\cdot\|^2$ represents the signal energy value.

$$\hat{s}(t) = s_{\text{target}}(t) + e_{\text{interf}} + e_{\text{noise}} + e_{\text{artif}} \tag{11}$$

$$\text{SDR} = 10 \log_{10} \frac{\left\| s_{t\,\arg et}(t) \right\|^2}{\left\| e_{\text{interf}} + e_{\text{noise}} + e_{\text{artif}} \right\|^2} \tag{12}$$

(2) PESQ calculation formula:

PESQ evaluation was proposed by Rix et al. for evaluating the quality of sound signals, which has been defined by the ITU-T recommendation P.862 [23]. As shown in formula (13), where $d_{\text{SYM}}$ and $d_{\text{ASYM}}$ represent symmetric and asymmetric disturbances, respectively, providing a good balance between prediction accuracy and generalization capability.

$$\text{PESQ} = 4.5 - 0.1 d_{\text{SYM}} - 0.0309 d_{\text{ASYM}} \tag{13}$$

The values of PESQ range from $-0.5$ to $4.5$. In cases of severe distortion, the PESQ value may be below $1.0$.

(3)   STOI calculation formula:

The STOI was proposed by Cees et al. for predicting the intelligibility of noisy speech [24]. As shown in formula (14), where $j = 1,2,...,J$ represents the index of one-third octave bands, $N$ is typically set to 30, and $d_{j,n}$ is the correlation coefficient of the short-time spectral vectors between the test speech and clean speech.

$$\text{STOI} = \frac{1}{JN} \sum_{j,n} d_{j,n} \tag{14}$$

For these three metrics, higher scores indicate better source separation results.

### 3.4 Experiment

A randomly selected heart sound (as the signal) and lung sound (as the noise) were mixed at a signal-to-noise ratio of 0 dB to create a mixed heart–lung sound signal. After encoding and decoding with DAE, the Waveform and Spectrogram of the reconstructed signal are shown in Figs. 6 and 7, respectively. Comparing the original signal with the reconstructed signal, it can be observed that the spectrogram and cepstrogram of the signal reconstructed by the DAE model closely match the original signal, demonstrating the effectiveness of the DAE model training.

The Waveform and Spectrogram of the separated heart sound and lung sound signals obtained through the DAE–NMF model are shown in Figs. 8.

Comparing the mixed signal (a1) with the separated heart sound signal (b1) and lung sound signal (c1) in Fig. 8, it is evident that the heart and lung sounds have been effectively separated, demonstrating the effectiveness of the DAE algorithm in separating the signals. Observing the Waveform in Fig. 7a1 and Spectrogram in Fig. 7a2 of the mixed signal, it can be seen that the heart and lung sounds overlap significantly in the low-frequency region. After separation using the DAE algorithm, the heart sound signal primarily concentrates in the low-frequency part, while the lung sound signal mainly concentrates in the high-frequency part. Examining the Spectrogram in Figs. 8b2, c2, it is noticeable that there is some minor interference of high-frequency lung sounds in the separated heart sound signal and vice versa in the separated lung sound signal.
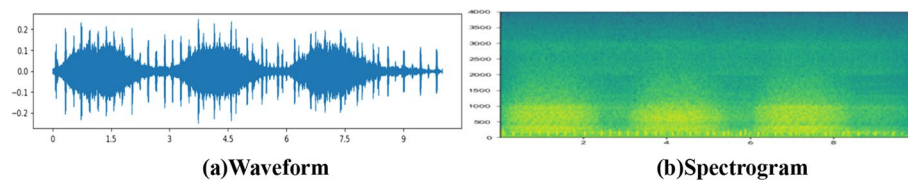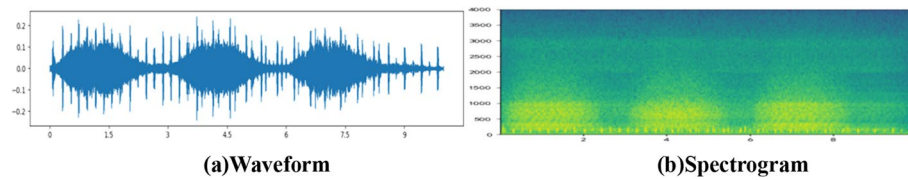


(a)Waveform                    (b)Spectrogram

**Fig. 6** Original signal



(a)Waveform                    (b)Spectrogram

**Fig. 7** Reconstructed signal



(a1) Mixed signal Waveform          (a2) Mixed signal Spectrogram

(b1) Heart sound signal Waveform    (b2) Heart sound signal Spectrogram

(c1) Lung sound signal Waveform     (c2) Lung sound signal Spectrogram
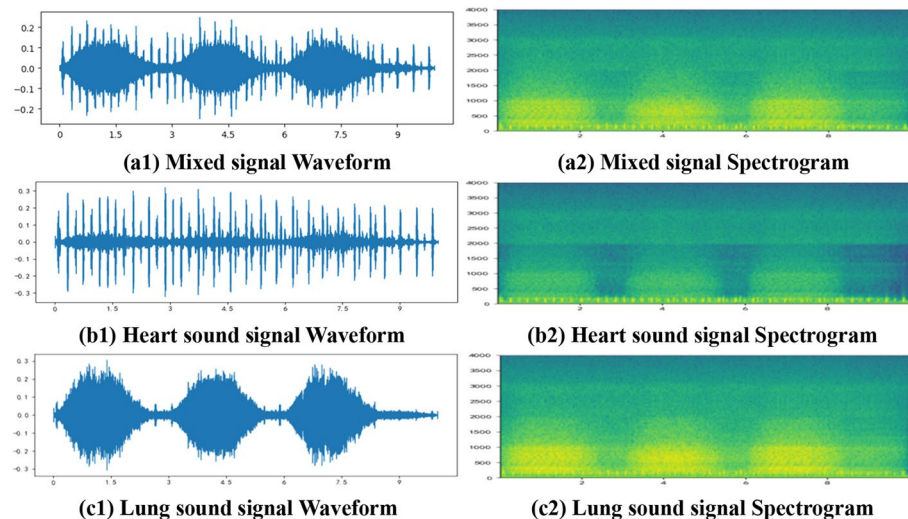
**Fig. 8** Signal after heart–lung sound separation

(a1) Use VMD to decompose heart sounds       (a2) Use VMD to decompose lung sounds

(b1) Heart sound IMF frequency                    (b2) Lung sound IMF frequency

(c1) Denoised heart sound Waveform               (c2) Denoised lung audio Waveform

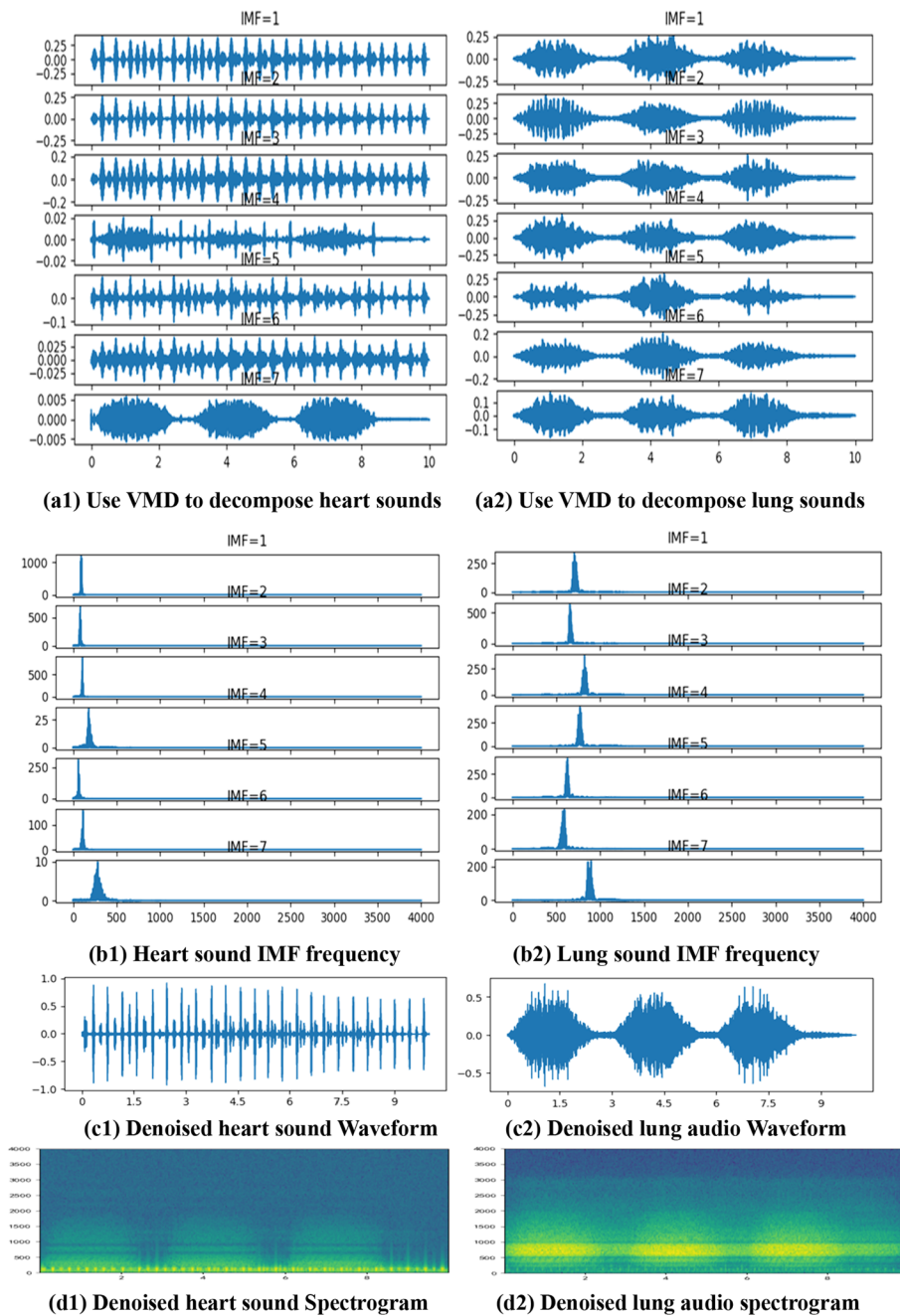(d1) Denoised heart sound Spectrogram            (d2) Denoised lung audio spectrogram

**Fig. 9** VMD noise reduction processing of heart sounds and lung sounds

In the separated Waveform of the heart–lung sound mixed signal obtained using the DAE-NMF algorithm, it can be observed that there is a slight presence of lung sound noise in the heart sound signal and vice versa. Therefore, denoising using the VMD algorithm was applied to the separated heart and lung sound signals. The Waveform and Spectrogram of the denoised heart and lung sound signals are shown in Fig. 9.

Sun *et al. EURASIP Journal on Advances in Signal Processing*     (2024) 2024:59

Page 13 of 17

Based on the number of decomposition modes, Fig. 9a1 shows that the heart sound, after being processed by the VMD algorithm, is decomposed into seven intrinsic mode functions (IMFs), and Fig. 9b1 displays the frequency corresponding to each IMF. Comparing the spectrogram of the heart sound in Fig. 8b and Fig. 9d1, it is observed that the heart sound signal after VMD denoising primarily consists of low-frequency components concentrated within the range of 20 to 150Hz. Similarly, comparing the lung sound spectrograms in Figs. 8c and 9d2, it is evident that the lung sound signals after VMD denoising are mainly composed of high-frequency components concentrated within the range of 50 to 2500 Hz, thus confirming the effectiveness of the VMD denoising algorithm.

### 3.5 Experimental comparison and evaluation

In this study, we compared four methods:

  (1) NMF-$K$-means algorithm.
  (2) DAE-$K$-means algorithm.
  (3) DAE-NMF algorithm.
  (4) DAE–NMF–VMD algorithm.

We randomly selected one heart sound signal and one lung sound signal, mixed them to create a mixed heart–lung sound signal with a signal-to-noise ratio (SNR) of 0 dB. We then fed this mixed signal into different heart–lung sound separation model for simulated experiments. Subsequently, we evaluated and compared the separated clean heart and lung sounds using the performance metrics SDR, PESQ and STOI. The results are shown in Tables 2 and 3.

Observing and comparing the heart sound evaluation indexes of four methods of heart–lung sound separation, we found that under the same clustering method, the SDR, PESQ and STOI values of the separation method using DAE to extract features are higher than those of the separation method using NMF to extract features. The SDR, PESQ and STOI values of DAE-$K$-means algorithm are 0.947944, 0.403321 and

**Table 2** Comparison of Heart Sound Evaluation Metrics

| Model | SDR | PESQ | STOI |
| --- | --- | --- | --- |
| NMF-$K$-means algorithm | 3.578298 | 0.541848 | 0.477938 |
| DAE-$K$-means algorithm | 4.526242 | 0.945169 | 0.586775 |
| DAE-NMF algorithm | 6.026242 | 1.131177 | 0.774058 |
| DAE–NMF–VMD algorithm | 9.561472 | 1.683339 | 0.945169 |

**Table 3** Comparison of lung sound evaluation indicators

| Model | SDR | PESQ | STOI |
| --- | --- | --- | --- |
| NMF-$K$-means algorithm | 3.061995 | 1.665166 | 0.516236 |
| DAE-$K$-means algorithm | 3.993437 | 1.885919 | 0.722660 |
| DAE-NMF algorithm | 4.667300 | 1.982141 | 0.733597 |
| DAE–NMF–VMD algorithm | 7.639952 | 3.886267 | 0.791355 |

0.108837 higher than those of NMF-*K*-means algorithm, respectively. Under the same feature extraction algorithm, the separation method using NMF periodic clustering algorithm has higher SDR, PESQ and STOI values than the separation method using *K*-Means clustering algorithm. The SDR, PESQ and STOI values of DAE-NMF periodic clustering algorithm are 1.500000, 0.186008 and 0.187283 higher than those of DAE-*K*-means algorithm, respectively.

The comparison of the above algorithms proves the effectiveness of DAE-NMF algorithm in separating heart–lung sounds. After adding VMD algorithm to DAE-NMF, the values of SDR, PESQ and STOI separated from heart sounds are 3.535230, 0.552162 and 0.171111 higher than the original ones, respectively. It is proved that the DAE–NMF–VMD algorithm proposed in this paper can not only effectively separate cardiopulmonary sounds, but also have good quality of separated heart sounds.

The same as the evaluation index of heart sound, we found that under the same clustering method, the SDR, PESQ and STOI values of the separation method using DAE to extract features are higher than those of the separation method using NMF to extract features. The SDR, PESQ and STOI values of DAE-*K*-means algorithm are 0.931442, 0.220753 and 0.206424 higher than those of NMF-*K*-means algorithm, respectively. Under the same feature extraction algorithm, the separation method using NMF periodic clustering algorithm has higher SDR, PESQ and STOI values than the separation method using *K*-Means clustering algorithm. The SDR, PESQ and STOI values of DAE-NMF algorithm are 0.673863, 0.096222 and 0.010937 higher than those of DAE-*K*-means algorithm, respectively.

After adding VMD algorithm to DAE-NMF, the values of SDR, PESQ and STOI of lung sounds were 2.972652, 1.9426 and 0.05758 higher than the original ones, respectively. It is proved that the DAE–NMF–VMD algorithm proposed in this paper can not only effectively separate cardiopulmonary sounds, but also have good quality of separated lung sounds.

By observing and comparing the values of SDR, PESQ and STOI, it can be concluded that compared to other methods for heart and lung sound separation, the proposed DAE–NMF–VMD algorithm has achieved improvements in all three evaluation metrics. This demonstrates that the quality of heart and lung sound separation using the DAE–NMF–VMD algorithm is significantly higher than other methods and indicates the effectiveness of using this approach for heart and lung sound separation.

### 3.6 Result verification

To further validate the effectiveness of the DAE–NMF–VMD model in separating heart and lung sounds, real heart and lung sound mixed signals were used for the separation experiments. Figures 7 and 8 show the spectrograms and spectrograms of the separated heart and lung sound signals obtained by the DAE–NMF–VMD model from the audio files 113_1306244002866_A.wav and 101_1305030823364_B.wav, respectively.

Comparing the Waveform of the mixed signal with the separated heart and lung sounds in Fig. 10a, it can be observed that the heart and lung sound signals are effectively separated. In Fig. 10b, the spectrogram of the separated heart and lung sound signals shows that the frequencies are primarily concentrated in the range of 20 ~ 150 Hz and 50 ~ 1000 Hz, which aligns with the frequency ranges of heart and lung sound signals.
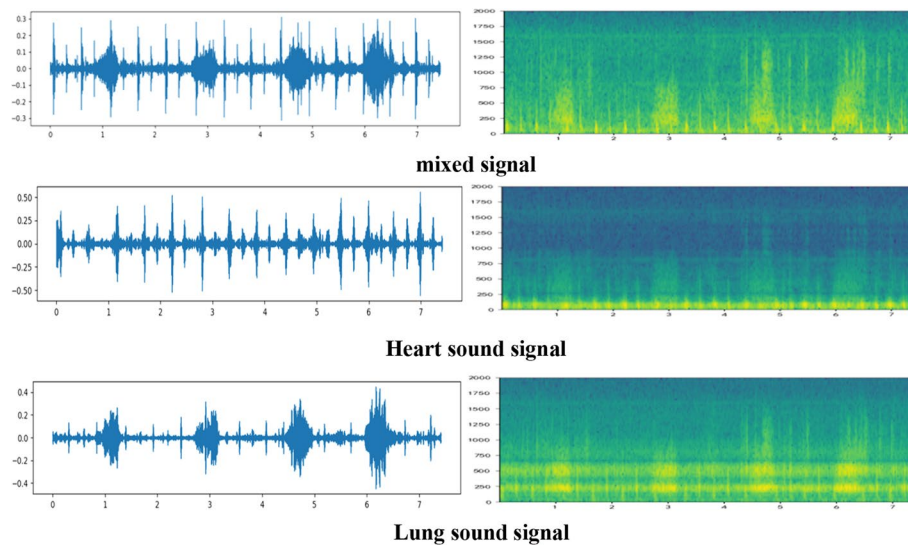
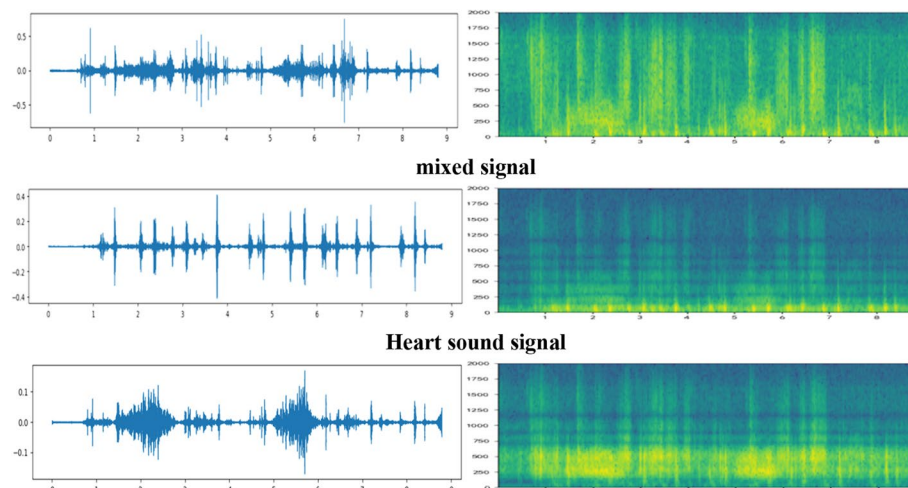**Fig. 10**  113_1306244002866_A.wav heart and lung sound separation effect



**Fig. 11**  101_1305030823364_B.wav heart and lung sound separation effect

Similarly, the mixed heart–lung sound separation results shown in Fig. 11 also confirm this, demonstrating the effectiveness of the DAE–NMF–VMD algorithm for separating heart–lung sound signals.

## 4 Conclusion

This paper presents a heart–lung sound separation method based on DAE–NMF–VMD, which, in addition to separating mixed heart–lung sounds based on DAE, applies VMD for denoising and enhancement of the separated signals. Unlike traditional heart–lung sound separation methods, DAE–NMF–VMD does not require supervised training data and leverages the periodic characteristics of heart and lung sound signals for separation. The research results indicate that this method yields satisfactory separation outcomes compared to other methods. SDR, PESQ and STOI

Sun *et al. EURASIP Journal on Advances in Signal Processing*    (2024) 2024:59

Page 16 of 17

test results demonstrate that the signal quality of the separated heart and lung sound signals using DAE–NMF–VMD is significantly improved compared to traditional methods. In the current study, only the separation and denoising of mixed heart–lung sound signals have been performed. In future research, attempts can be made to combine this algorithm with other advanced sound processing techniques to improve signal separation performance. Next research could focus on developing classification algorithms for heart and lung sound signals to assess their health status and analyze medical conditions.

**Abbreviations**
DAE      Deep autoencoder
NMF      Nonnegative matrix factorization
VMD      Variational mode decomposition
ICA       Independent component analysis
SDR      Signal distortion rate
PESQ     Perceptual evaluation of speech quality
STOI      Short-term objective intelligibility
STFT      Short-time Fourier transform
LPS       Logarithmic power spectra
MSE      Mean squared error
DFT      Discrete Fourier transform
ISTFT     Inverse short-time Fourier transform
IMFs      Intrinsic mode functions
EMD      Empirical mode decomposition

## Declarations

**Competing interests**
The authors declare that they have no competing interests.

## References
1.  H. Ren, H. Jin, C. Chen et al., A novel cardiac auscultation monitoring system based on wireless sensing for healthcare. IEEE J. Transl. Eng. Health Med. **6**, 1–12 (2018)
2.  A. Gurung, C.G. Scrafford, TielschJM, et al., Computerized lung sound analysis as diagnostic aid for the detection of abnormal lung sounds: a systematic review and meta-analysis. Respir. Med. **105**(9), 1396–1403 (2011)
3.  F.G. Nabi, K. Sundaraj, C.K. Lam, R. Palaniappan, Identification of asthma severity levels through wheeze sound characterization and classification using integrated power features. Biomed. Signal Process. Control **52**, 302–311 (2019)
4.  F. Pancaldi, M. Sebastiani, G. Cassone et al., Analysis of pulmonary sounds for the diagnosis of interstitial lung diseases secondary to rheumatoid arthritis. Comput. Biol. Med. **96**, 91–97 (2018)
5.  N.S. Haider, B.K. Singh, R. Periyasamy et al., Respiratory sound based classification of chronic obstructive pulmonary disease:a risk stratification approach in machine learning paradigm. J. Med. Syst. **43**, 255 (2019)
6.  S. Don, Random subset feature selection and classification of lung sound. Procedia Comput. Sci. **167**, 313–322 (2020)

Sun *et al. EURASIP Journal on Advances in Signal Processing*    (2024) 2024:59

Page 17 of 17

7.  A. Mondal, I. Saxena, A noise reduction technique based on nonlinear kernel function for heart sound analysis. IEEE J. Biomed. Health Inf. **22**, 775–784 (2018)

8.  D. Emmanouilidou, E.D. McCollum, D.E. Park et al., Computerized lung sound screening for pediatric auscultation in noisy field environments. IEEE Trans. Biomed. Eng. **65**, 1564–1574 (2018)

9.  M. T. Pourazad,Z. Moussavi, F. Farahmand, et al., Heart sounds separation from lung sounds using independent component analysis, in *2005 IEEE Engineering in Medicine and Biology 27th Annual Conference*, pp. 2736–2739 (2006)

10. J. C. Chien, M. C. Huang, Y. D. Lin, et al., A study of heart sound and lung sound separation by independe-nt component analysis technique, in *2005 IEEE Engineering in Medicine and Biology 27th Annual Conference, Shanghai, China*, pp. 5708–5711 (2005)

11. F.J. Cañadas-Quesada et al., A non-negative matrix factorization approach based on spectro-temporal clustering to extract heart sounds. Appl. Acoust. **125**, 7–19 (2017)

12. E. Grooby, J. He, D. Fattahi et al., Noisy neonatal chest sound separation for high-quality heart and lung sounds. Biomed. Health Inf. **27**(6), 2635–2646 (2023)

13. W. Weibo et al., Heart-lung sound separation by nonnegative matrix factorization and deep learning. Biomed. Signal Process. Control **79**, 104180 (2023)

14. R. Nersisson, M.M. Noel, Hybrid Nelder-Mead search based optimal least mean square algorithms for heart and lung sound separation. Eng. Sci. Technol. Int. J. **20**, 1054–1065 (2017)

15. N. Q. Al-Naggar, M. H. Al-Udyni, Performance of adaptive noise cancellation with normalized last-mean-square based on the signal-to-noise ratio of lung and heart sound separation, J. Healthc Eng. 9732762 (2018)

16. C. Yang, N. Hu, D. Xu, et al., Monaural cardiopulmonary sound separation via complex-valued deep autoencoder and cyclostationarity. Biomed. Phys. Eng. Express. 9 (2023)

17. J. Xingxing, S. Qiuyu, D. Guifu et al., Research and application review of variational mode decomposition methods. J. Instrum. Meast. **44**(01), 55–73 (2023)

18. K.H. Tsai, W.C. Wang, C.H. Cheng et al., Blind monaural source separation on heart and lung sounds based on periodic-coded deep autoencoder. IEEE J. Bio-med. Health Inf. **24**, 3203–3214 (2020)

19. F Chen, J Wang and C Li., 94 GHz asymmetric antenna radar for speech signal detection and enhancement via variational mode decomposition and improved threshold strategy. IEEE Access. 10,97930–97944(2022)gfd

20. P. Bentley, G. Nordehn, M. Coimbra, et al. Classifying heart sounds challenge. http://www.peterjbentley.com/heart challenge/

21. B.M. Rocha, D. Filos, L. Mendes et al., An open access database for the evaluation of respiratory sound classification algorithms. Physiol Measurement. **40**, 035001 (2019)

22. E. Vincent, R. Gribonval, C. Févotte, Performance measurement in blind audio source separation. IEEE Trans. Audio Speech Lang. Process. **14**(4), 1462–1469 (2006)

23. A. W. Rix, J. G. Beerends, M. P. Hollier,et al. Perceptual evaluation of speech quality (PESQ)-a new method for speech quality assessment of telephone networks and codecs, in *IEEE International Conference on Acoustics, Speech, and Signal Processing*, pp. 749–752 (2001)

24. C.H. Taal, R.C. Hendriks, R. Heusdens et al., An algorithm for intelligibility prediction of time-frequency weighted noisy speech. IEEE Trans. Audio Speech Lang. Process. **19**(7), 2125–2136 (2011)

## Publisher's Note