



OPEN

Multi-level feature encoding algorithm based on FBPSI for heart sound classification

Yu Fang, Hongxia Leng, Weibo Wang & Dongbo Liu[✉]

Analysis of heart sound signals plays an essential role in preventing and diagnosing cardiac diseases. This study proposes a multi-level feature encoding algorithm based on frequency-balanced power spectral intensity for heart sound signal classification. Firstly, a wavelet threshold function is employed to denoise the heart sound signals. Then, the frequency-balanced power spectral intensity envelope is calculated, and an encoder is utilized to extract multi-level features based on the envelope. Finally, an ensemble bagging tree classifier is selected for classification. The experimental data includes binary classification data from the 2016 PhysioNet/CinC Challenge and ternary classification data from the self-collected hypertrophic cardiomyopathy dataset. Results demonstrate that the proposed algorithm performs well, achieving an average classification accuracy of 98.73% for normal and abnormal heart sounds, and 98.12% for normal and two types of hypertrophic cardiomyopathy heart sounds. The proposed method holds significant reference value for the early diagnosis of heart diseases.

Keywords Heart sound classification, Balanced power spectrum intensity, Multi-level feature encoding, Hypertrophic cardiomyopathy

Cardiovascular disease (CVD) is a global epidemic that is increasing in low- and middle-income countries¹. According to World Health Statistics 2023², the number of CVD deaths reached 17.9 million in 2019, and CVD will remain the leading cause of disease deaths globally in 2023³. The number of current CVD patients has reached 330 million in China. CVD morbidity and mortality are still a growing trend due to the unhealthy lifestyles of the population⁴. Hypertrophic cardiomyopathy (HCM) is one of the most common inherited heart diseases. According to hemodynamic characteristics, HCM can be classified into obstructive hypertrophic cardiomyopathy (OHCM) and non-obstructive hypertrophic cardiomyopathy (NOHCM)⁵. Echocardiography can determine the left ventricular outflow tract gradient (LVOTG) in patients with HCM, subdivided into 3 types according to the LVOTG value. A peak LVOTG of ≥ 30 mmHg (3.99 kPa) at quiet is considered resting obstruction; a peak LVOTG of < 30 mmHg at quiet but > 30 mmHg at load or during exercise is considered concealed obstruction; and a pressure step difference of < 30 mmHg at both quiet and load is considered non-obstruction⁶.

Heart sound (HS) signals contain rich cardiac physiological information. Analyzing HS signals can find early pathological features of the heart⁷, which is crucial for early diagnosis and treatment of diseases. HCM pathology HS is mainly associated with the fourth heart sound (S4) and systolic murmur. The S4 is produced at end-diastole due to decreased ventricular diastolic function, while outflow tract obstruction produces a jet murmur during cardiac systole⁸ as shown in the left plots of Fig. 1. From the spectrogram shown in the right of Fig. 1, it can be found that the frequency range of the normal signal is mainly within 200 Hz, while the HCM heart murmur is still present at 600 Hz with spectral energy.

HS analysis has been widely applied in the auxiliary diagnosis of diseases by transforming the original HS signal into more informative feature vectors. Among them, feature extraction in the frequency domain has been widely studied in HS classification.

Power spectral density (PSD) is a classical feature analysis method in the frequency domain. Xu et al.⁹ constructed a PSD feature matrix and classified normal and abnormal HSs with an accuracy of 84.72% for the 2016 PhysioNet/CinC Heart Sound Challenge dataset (PCCD). Wang et al.¹⁰ combined Mel-frequency cepstral coefficients (MFCCs) and PSD with an accuracy of 96.5% in 5000 HS recordings. Pathak et al.¹¹ applied the imaginary

School of Electrical and Electronic Information, Xihua University, Chengdu 610039, Sichuan, China. [✉]email: liudb@mail.xhu.edu.cn

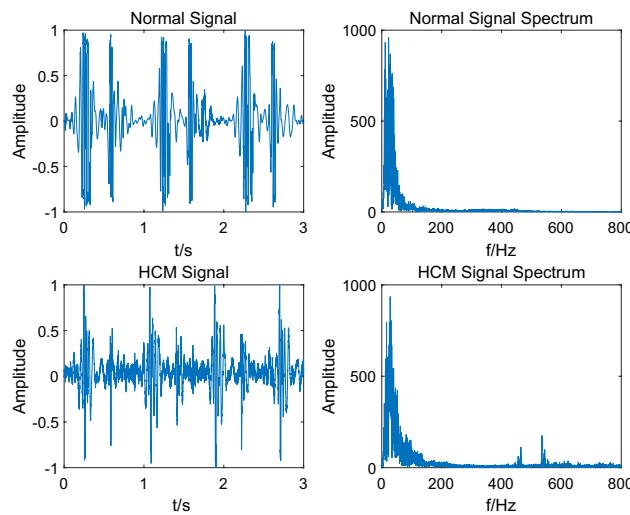


Figure 1. Time-domain and frequency-domain plots of normal and HCM HS.

part of the cross PSD to extract spectral features of HSs and achieved the classification accuracy of 74.98% for the normal and coronary artery disease HSs. Wang et al.¹² selected the best HS signals based on the PSD similarity quality factor and attained an automatic selection success rate of 95.83% in 144 signals, 25 normal signals, and 119 congenital heart diseases. Luo et al.¹³ proposed an enhanced PSD feature extraction method using color, texture, similarity, and other features of the image, with a difference significance value of 0.0013 in the SEED electroencephalography (EEG) dataset, which solved that the power spectrum could not effectively represent the differences between frequencies in the analysis of EEG signals. Moreover, Yang et al.¹⁴ proposed the balanced power spectrum intensity (BPSI) with an average accuracy of 88.31% for emotion recognition based on the DEAP EEG dataset, which improves the low recognition accuracy due to the large differences in EEG frequency bands.

In this study, to improve the effectiveness of frequency domain feature extraction. We improve the BPSI and propose the frequency-balanced power spectral intensity (FBPSI).

After obtaining effective features, the choice of classification model is an important factor, which affects the accuracy of HS analysis. Lee et al.¹⁵ applied different wavelet analysis methods combined with a one-dimensional and a two-dimensional convolutional neural network respectively, and the fusion of both of them on PCCD gave the best classification result with an accuracy of 98.57%. Hu et al.¹⁶ proposed a deep separable convolutional neural network and introduced a channel attention mechanism to improve the feature representation of the model, with 97.7% accuracy based on PCCD. The deep learning model greatly improves signal classification accuracy. However, numerous hyperparameters including learning rate, network structure, regularization parameters, etc., still need to be tuned and more experiments are required to find the optimal parameters.

Yunendah et al.¹⁷ extracted the features of MFCCs and validated the algorithm on PCCD using various classifiers where an optimal 95.78% accuracy was obtained using K-nearest neighbors (KNN). Yadav et al.¹⁸ based on the PCCD to split the signal into frames, extracted a total of 12 features such as energy entropy, over-zero domain, the spectral center of mass, etc., from each frame, and classified them using a support vector machine (SVM) with an accuracy of 97.78%.

In the above studies, scholars have classified HSs by extracting multiple spectral features combined with appropriate classifier models. However, for early prevention and daily monitoring of heart disease, improving computational efficiency and reducing the cost of training and prediction of classification models are still topics that need to be addressed.

Enoding, which is one of the effective methods to solve the above challenge can extract the original signal's abstract features¹⁹. Tsai et al.²⁰ proposed the cycle coding depth self-encoder to separate HS and lung sounds, in 126 mixed heart and lung sound signals, with an average accuracy of 85.67%. Cheng et al.²¹ proposed an HS classification model constructed based on a one-dimensional convolution and a Transformer encoder, with an accuracy of 96.4% on PCCD. Wang et al.²² utilized an improved one-dimensional neural network combined with an encoder and the classification accuracy was 97.9% on PCCD. Riccio et al.²³ used partitioned iteration function systems (PIFS) for image coding of 1D HS signals to generate 2D color images, which achieved 85% classification accuracy on PCCD. Anjum et al.²⁴ used linear predictive coding coefficient vectors to encode the entire EEG time-series data, these coefficients represent the relevant features of the EEG, which can better differentiate Parkinson's disease, and the classification accuracy was improved from 78.3 to 93.8% in 82 EEG recordings. Tuncer et al.²⁵ proposed a local dual octal pattern algorithm, which utilizes discrete wavelets to stratify the signal, extract comprehensive feature information of the signal, and classify and recognize it in the MPSSC snoring database, the unweighted average recall was improved from 72.6 to 94.65%.

To improve the problem of insufficient information and low accuracy of feature extraction, especially when classifying normal and pathological HS signals as well as the challenges of different pathological HS signals, a novel classification method based on multi-level feature encoding (MLFE) algorithm and FBPSI has been proposed in our study. It is utilized to capture subtle differences in signals more accurately by generating a more

expressive feature matrix and classifying the normal and abnormal signals by the ensemble bagging tree classification (EBTC). Through the tests of different parameters of FBPSI and MLFE algorithm, the best model is developed for kinds of HS classification. It is validated by multiple HS datasets with five evaluation indices and shows excellent classification performance compared with other HS classification methods. It has great potential to extract pathologic information of abnormal HS signals and can distinguish the different kinds of HS signals effectively, which will extend to more fine classification of HS signals.

Methods

The algorithmic sequence is demonstrated in Fig. 2, which will be detailed in the following subsections.

Experimental data

The following two HS datasets are used for validation in our study.

- (1) PhysioNet/CinC Challenge Dataset (PCCD) 2016 consists of 6 databases (a-f) totaling 3240 HS recordings, 2575 normal from healthy subjects, and 665 abnormal from cardiac patients. The sampling frequency was 2000 Hz and saved in *.wav format. The details of the dataset are shown in Table 1.
- (2) The Hypertrophic Cardiomyopathy Dataset (HCMD) comprised 270 HS recordings. 110 normal signals were from healthy subjects who passed a medical examination. The sampling frequency of normal HSs was set at 8 kHz or 44.1 kHz. In addition, 160 abnormal signals from patients suffering from HCM were acquired at 44.1 kHz in the cooperative hospital of our laboratory. Details of the HCMD are shown in Table 2. This study was performed in line with the principles of the Declaration of Helsinki. Approval was granted by the ethics committee of Chengdu Region General Hospital (No. 2015 research 01).

Preprocessing

To optimize the data processing efficiency, the signal with a sampling frequency of 44.1 kHz in the HCMD is resampled to 8 kHz, while the sampling frequency of the PCCD is retained at 2 kHz. To reduce noise interference, the wavelet threshold noise reduction function, which performs better threshold quantization of each level's

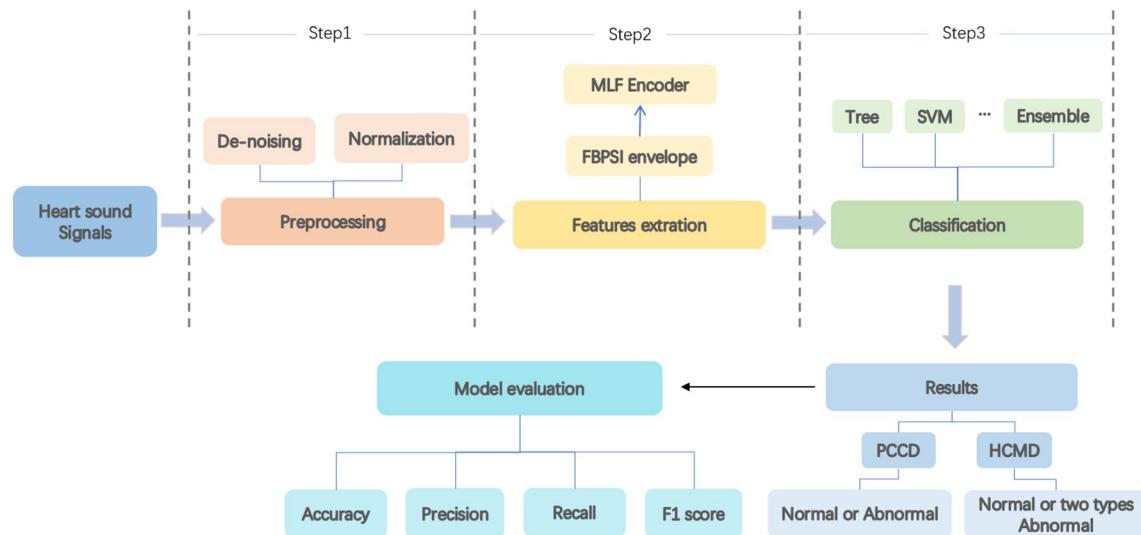


Figure 2. The workflow of the Multi-level feature encoder algorithm based on FBPSI for HS classification.

Dataset	Subject number	Normal recording	Abnormal recording	Total
Training-a	121	117	292	409
Training-b	106	386	104	490
Training-c	31	7	24	31
Training-d	38	27	28	55
Training-e	356	1958	183	2141
Training-f	112	80	34	114
Total		2575	665	3240

Table 1. 2016 PhysioNet/CinC challenge dataset.

Recording	Age	Subject number	Total
Normal	20–24	28	110
OHCM	14–70	30	80
NOHCM			80
Total		58	270

Table 2. Hypertrophic cardiomyopathy dataset.

high-frequency coefficients in the wavelet decomposition of the noisy signal²⁶, is applied to obtain the denoised HS signal, denoted as x . Then, the amplitude of x is normalized using the maximum-minimum method shown in Eq. (1).

$$x_{\text{normalized}} = \frac{x - x_{\min}}{x_{\max} - x_{\min}} \quad (1)$$

where x is the denoised HS signal, x_{\min} is the minimum value of x , x_{\max} is the maximum value, and $x_{\text{normalized}}$ is the normalized HS signal.

Frequency Balanced Power Spectral Intensity Envelope

For an N -point normalized HS signal, the frequency-balanced power spectral intensity (FBPSI) envelope is extracted according to the following steps.

1. The spectrum $X(k)$ is obtained by the N -point Fast Fourier Transform of the HS signal $x_{\text{normalized}}$.
2. To facilitate the local detail analysis and processing of the spectrum, the spectrum $X(k)$ is windowed. A shorter window can provide higher frequency resolution, and a longer can provide more accurate low-frequency information, but will reduce the frequency resolution.

The window length of $L_{\text{win}} = 100$ sampling points is set by experimental analysis and comparison. To suppress excessive data variation between windows, the window overlap is taken as half of the length. The selected window function is multiplied point by point with the amplitude of the $X(k)$, and more effective frequency domain information can be obtained. The hamming window with a narrower sidelobe is selected to reduce the spectral leakage.

3. The amplitude points of the HS signal in each window are summed and logarithmized separately to obtain the FBPSI envelope. FBPSI envelope as shown in Eq. (2).

$$Q_{\text{FBPSI}}(l) = \sum_{l=1}^M \left(\log \sum_{m=1}^{L_{\text{win}}} |X(m, l)| \right) \quad (2)$$

where $1 \leq m \leq L_{\text{win}}$, $1 \leq l \leq M$, m is the number of spectral sampling points, l denotes the l th window, M is the total number of windows, win is the hamming window, and $X(m, l)$ is the window function.

Multi-level feature encoding

Multi-level Feature Encoding is proposed as the key algorithm in the study. First, the input signal is decomposed at multiple levels using the maximum overlapping discrete wavelet transform²⁷ (MODWT), then 128-dimensional features are extracted at each level by a fixed window length and composed into a feature vector, which is used as the input to the classifier finally.

1. The input signal is decomposed in $J=4$ layers by MODWT and using the “sym4” wavelet. Compared with the traditional wavelet transform, both wavelet and scale functions of MODWT are convolved with the maximum overlap. This overlap can make the wavelet have better capture time–frequency localization characteristics. The transform process is shown in Eq. (3)

$$W_j(l) = \sum_{k=0}^{N-1} Q_{\text{FBPSI}}(l) \cdot h(l - k) \quad (3)$$

where $Q_{\text{FBPSI}}(l)$ is the input signal, $h(l - k)$ is the wavelet basis function, N is the length of $h(l - k)$, and $W_j(l)$ is the wavelet coefficients at the j th layer.

2. Multiple overlapping windows of length W_{length} are applied to each layer of signals.

$$win_a = Q_{\text{FBPSI}}(i : i + 8), a \subseteq [1, M - 8], i \subseteq [1, M], a \in \mathbb{Z}^+, i \in \mathbb{Z}^+ \quad (4)$$

where win_a represents the a th overlapping window of size W_{length} , a and i represent index.

3. The reference point Ref is specified (by default, the center of each window is taken as the reference point), and the left and right feature areas are determined.

$$Ref = \text{win}_a(\lceil W_{length}/2 \rceil) \quad (5)$$

where $\lceil \cdot \rceil$ denotes upward rounding.

$$area_L = \text{win}_a(1 : \lceil W_{length}/2 \rceil - 1) \quad (6)$$

$$area_R = \text{win}_a(\lceil W_{length}/2 \rceil + 1 : W_{length}) \quad (7)$$

4. The difference values between the two sides of the feature regions and the ref are calculated using the β as the weight. The formula is shown in Eq. (8).

$$Value_{weight-diff}(Q_{FBPSI}, Ref, \text{win}_a) = \sum_{i=1}^{\lceil W_{length}/2 \rceil - 1} [Q_{FBPSI}(\text{win}_a(i)) - Ref] * \beta^{i-1} \quad (8)$$

5. The histogram is used to divide the left and right counting regions to obtain region features. The signal is encoded using β bits, so the histogram length is calculated as 2^β . The left and right region features G are obtained from Eqs. (9) and (10)

$$G_L = G_L(Value_L) + 1 \quad (9)$$

$$G_R = G_R(Value_R) + 1 \quad (10)$$

6. Combining and normalizing the left and right region features will obtain a vector of size $2 \times 2^\beta$, so a feature vector of $2 \times 2^\beta \times J$ can be extracted for each recording.

The whole encoding process is shown in Fig. 3. After wavelet 4-level decomposition applies multiple windows of overlapping ($W_{length}=9$, $Ref=\text{win}(5)$), the left and right region features G_L and G_R are computed separately (weight $scale=4$), resulting in a 128-dimensional feature vector.

Ensemble bagged-trees classification

In this study, we use an EBTC to classify HS signals. EBTC uses the statistical method of bootstrap aggregation to solve the overfitting decision trees and improve the performance of the classification model^[28,29]. As shown in Fig. 4. First, multiple new sets are generated by randomly sampling multiple subsets of the feature set using a method with put-back. Then, the model is constructed by applying the decision tree algorithm. An integrated classifier is built based on the voting results of each model, and it is used for prediction.

Assuming that there is an integrated model containing A decision trees, each decision tree gives a result $Y_{bag}(j)$ for each input subset, and its final prediction is shown in Eq. (11).

$$Y_{bag} = \text{argmax}_c \sum_{i=1}^A 1(Y_{bag}(j) = c) \quad (11)$$

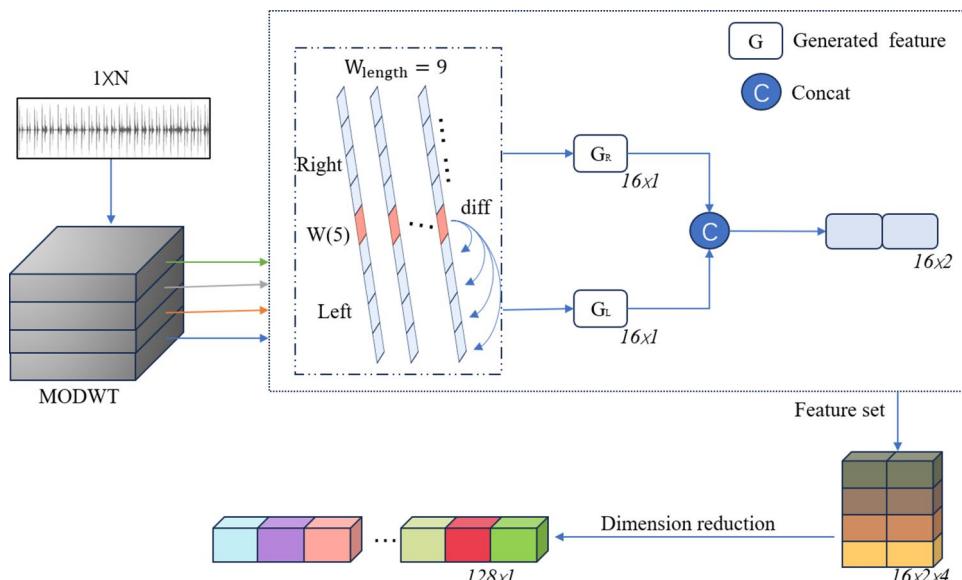


Figure 3. Flow chart of multi-level feature extraction coding algorithm.

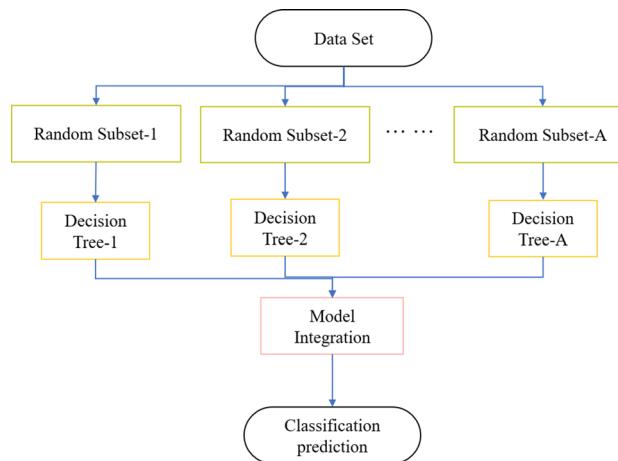


Figure 4. Ensemble Bagging Trees training procedure.

where $j=1,2,\dots,A$. c is the category label, $1(\cdot)$ is the indicator function, and the bracketed condition returns 1 if it holds, and 0 otherwise.

Evaluation metrics

The performance of the proposed method is evaluated using four factors, accuracy (Acc), specificity (Spe), precision (Pre), unweighted average recall (UAR)³⁰, and F1-score (F1) and defined in Eqs. (12)–(16), respectively.

$$Acc = \frac{TP + TN}{TP + TN + FP + FN} \quad (12)$$

$$Spe = \frac{TN}{TN + FP} \quad (13)$$

$$Pre = \frac{TP}{TP + FP} \quad (14)$$

$$UAR = \sum_{n=0}^k \frac{\left(\frac{TP_n}{TP_n + FN_n} \right)}{k} \quad (15)$$

$$F1 = \frac{2TP}{2TP + FP + FN} \quad (16)$$

where k represents the number of categories, TP_n represents the number of true positive samples in the n th category, FP_n represents the number of false negative samples in the n th category, and FN_n represents the number of false positive samples in the n th category. TN represents the number of true negative samples.

Informed consent

Informed consent was obtained from all individual participants included in the study.

Results

In this study, the PCCD and HCMD are applied to validate the effectiveness of the proposed method, respectively. Experiments with different parameter settings of MLFE are applied based on the PCCD dataset, and the best combination of MLFE is selected through the HS classification results to extract the most effective features.

1. Classification of HS signals with different weights

Table 3 shows the HS classification accuracies of the MLFE and FBPSI + MLFE algorithms based on *Ref* (taking the center point) corresponding to different window lengths and weights. β is taken as 2, 4, and 8, and *win* is taken as 5, 9, and 17. For the MLFE, a maximum accuracy of 95.70% can be achieved when $W_{length}=17$ and $\beta=8$ for classifying normal and abnormal HSs. For the FBPSI + MLFE algorithm, the accuracy is 98.73% when $W_{length}=9$, $\beta=4$.

β	Acc (%)					
	MLFE			FBPSI + MLFE		
	$W_{length} = 5$	$W_{length} = 9$	$W_{length} = 17$	$W_{length} = 5$	$W_{length} = 9$	$W_{length} = 17$
2	87.41	87.90	94.60	98.00	97.20	97.20
4	84.80	87.40	95.50	98.60	98.73	98.70
8	86.10	88.30	95.70	96.90	98.70	97.50

Table 3. Classification accuracy results of normal and abnormal of different W_{length} and β in PCCD. Significant values are in [bold].

2. Classification of HSs at different reference points

In encoding, the change of ref affects the extraction effect of MLFE, and after determining the window length and weight, different ref positions are set for experiments. From Table 4, the MLFE algorithm obtains a maximum accuracy of 96.57% when $ref = win(17)$, the endpoint of the window, and the FBPSI + MLFE obtains an accuracy of 98.73% when $ref = win(5)$, the center of the window.

3. Classification of HSs at different decomposition modes

As shown in Table 5, based on the above parameters, decomposition methods are selected for experiments, including Empirical Mode Decomposition (EMD), Variational Mode Decomposition (VMD), Tunable Q-factor Wavelet Transform (TQWT), and MODWT. The MLFE and the FBPSI + MLFE algorithm can obtain the best results when MODWT is selected, which for classifying normal and abnormal HSs is 96.57% and 98.73%, respectively. Table 6 compares the FBPSI, MLFE, and FBPSI + MLFE, and the proposed FBPSI + MLFE algorithm achieves better performance metrics on PCCD. Figure 5b shows a visual comparison of the experimental results of the five performance metrics of the above three algorithms for each fold in the tenfold cross-validation.

Then, the proposed method was validated on normal, OHCM, and NOHCM triple classification. The algorithm is set up with consistent parameters, the signal is decomposed with MODWT, $W_{length} = 9$, Ref is chosen to be the center of the window, $\beta = 4$, and finally, the 128-dimensional features are output. As shown in Fig. 6, a comparison of FBPSI, MLFE, and FBPSI + MLFE algorithms is presented for HCM. Again, the proposed algorithm achieves better results than the other two algorithms on all five metrics.

	Acc(%)	
Ref	MLFE	FBPSI + MLFE
Win(1)	95.80	97.90
Win(3)	95.00	98.70
Win(5)	95.20	98.73
Win(7)	94.40	98.65
Win(9)	95.70	98.10
Win(11)	95.60	–
Win(13)	95.00	–
Win(15)	95.40	–
Win(17)	96.57	–

Table 4. Classification accuracy results of normal and abnormal of different Ref based on optimal β and W_{length} in PCCD. Significant values are in [bold].

	Acc (%)	MLFE				Acc (%)	FBPSI + MLFE			
		Spe (%)	Pre (%)	UAR (%)	F1 (%)		Spe (%)	Pre (%)	UAR (%)	F1 (%)
EMD	90.28	81.95	86.87	81.95	84.07	85.00	81.13	86.46	96.19	91.07
VMD	96.51	93.68	95.49	93.68	94.55	98.46	95.56	98.65	99.42	99.03
TQWT	96.37	93.14	95.34	93.14	94.19	98.61	95.60	99.38	98.87	99.12
MODWT	96.57	93.88	95.76	93.89	94.79	98.73	95.65	98.88	99.53	99.21

Table 5. Comparison results of different decomposition methods based on optimal parameters of encoding in PCCD. Significant values are in [bold].

	Acc (%)	Spe (%)	Pre (%)	UAR (%)	F1 (%)
FBPSI	87.13	89.35	88.87	95.81	92.21
MLFE	96.57	93.88	95.76	93.89	94.79
FBPSI + MLFE	98.73 ± 0.53	95.65 ± 2.40	98.88 ± 0.60	99.53 ± 0.36	99.21 ± 0.33

Table 6. Comparison results of the FBPSI, MLFE, and FBPSI + MLFE in PCCD. Significant values are in [bold].

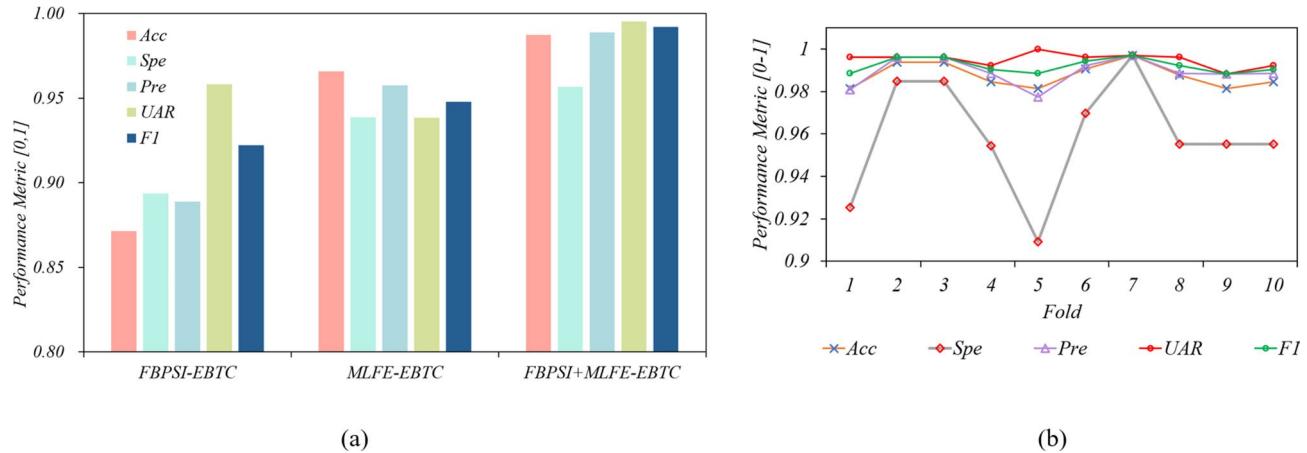


Figure 5. (a) The performance comparison of three algorithms in PCCD; (b) The results of the FBPSI + MLFE algorithm obtained from tenfold cross-validation in PCCD.

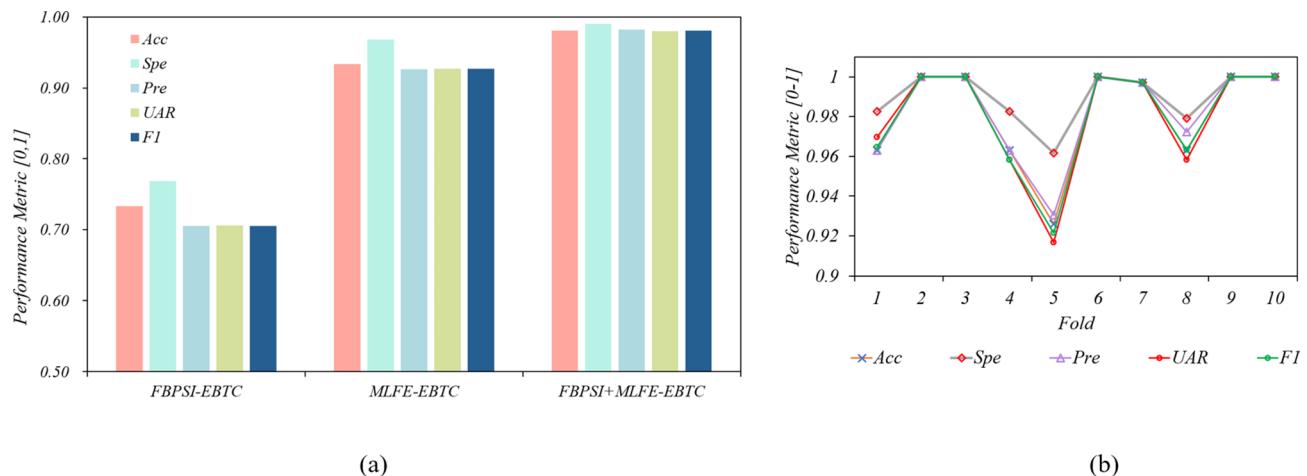


Figure 6. (a) The performance comparison of three algorithms in HCMD; (b) The results of the FBPSI + MLFE algorithm obtained from tenfold cross-validation in HCMD.

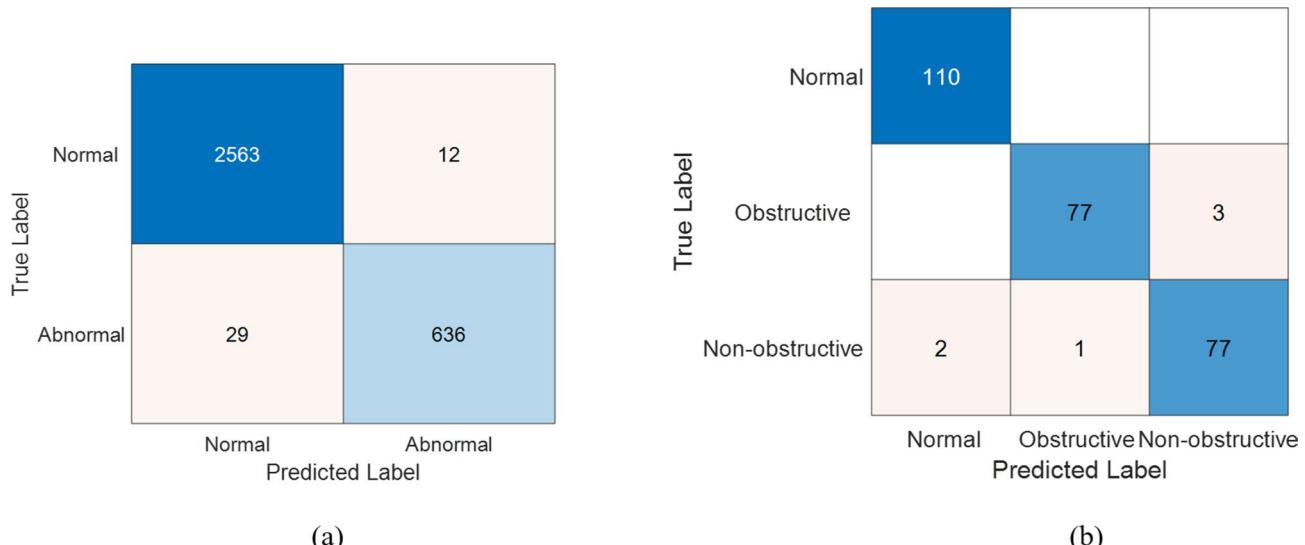
Finally, a comprehensive performance evaluation of decision trees (DT), SVM, KNN, multilayer neural network (MNN), and EBTC was conducted. Based on the results of cross-validation the classifier with the higher classification accuracy among them is selected as the final model, their respective hyperparameters are shown in Table 7. The classification algorithm workstation is configured with an Inter(R) Core (TM) i5-8300H CPU @ 2.3 GHz processor, 16 GB RAM, 64-bit operating system, and \times 64-based processor.

Table 8 shows the results of the normal and abnormal HS signals in PCCD and the normal, OHCM, and NOHCM HS signals in HCMD under different classifiers. The classification confusion matrixes are shown in Fig. 7. From Table 8, it is easy to find that the model training time for the EBTC is shorter, longer than DT and SVM in PCCD, and only longer than DT in HCMD. Meanwhile, the HS classification accuracy results of EBTC are the best both in PCCD and HCMD. Based on the experiment results, the EBTC with higher accuracy and lower classification time complexity is applied in continuous work.

Modules	Hyperparameters
DT	Fine Tree: A decision tree with detailed branching. Maximum Splits: 100. Splitting Criterion: Gini Diversity Index
SVM	Quadratic Kernel Function: The kernel used is a quadratic polynomial. Kernel Scale: Automatic. Box Constraint Level: 1. Multiclass Method: One-vs-One. Standardize Data: Yes
KNN	Weighted KNN. Number of Neighbors: 10. Distance Metric: Euclidean. Distance Weight: Inverse Distance Squared. Standardize Data: Yes
MNN	Wide Neural Network. Number of Fully Connected Layers: 1. First Layer Size: 100. Activation Function: ReLU. Iteration Limit: 1000. Regularization Strength (Lambda): 0. Standardize Data: Yes
EBTC	Learner Type: Decision Tree. Number of Learners: 10

Table 7. Model hyperparameters.

No.	Module	PCCD		HCMD	
		Acc(%)	Train time (s)	Acc (%)	Train time (s)
1	DT	98.30 ± 0.76	2.4645	96.67 ± 2.73	2.5808
2	SVM	98.54 ± 0.39	7.6301	96.30 ± 2.47	8.9542
3	KNN	87.00 ± 2.58	9.2105	86.74 ± 2.43	7.3553
4	MNN	97.90 ± 1.00	10.161	96.80 ± 2.65	10.796
5	EBTC	98.73 ± 0.53	7.9118	98.12 ± 1.79	6.9425

Table 8. The classification results of different modules.**Figure 7.** (a) Confusion matrix for the PCCD; (b) Confusion matrix for the HCMD.

Discussion

According to the experiment results above, we have analyzed the FBPSI + MLFE + EBTC method from the parameter set to the classifier selection. Now the excellence of the proposed method compared to other literature applying in PCCD is discussed and shown in Table 9. Unlike PSD⁹ or BPSI, the FBPSI proposed can minimize the differences between different frequency bands of the same HS signal, which is more effective for subsequent encoding.

The proposed method has also been compared with previous research for the encoding section. It encoded a one-dimensional into a two-dimensional image, but the algorithm requires the length of the signal to be fixed. Signals that do not meet this requirement are truncated or repeatedly spliced, resulting in a classification accuracy of only 85% for the dataset²³. Some researchers utilized the Transformer encoder to enhance feature extraction, leading to better classification results^{21,22}. However, the structure of the Transformer encoder in literature^{21,22} has a high computational experiment with different parameter settings and the performance is dependent on the positional coding information. Moreover, the encoder also consumes more time on the design. The proposed MLFE adopts a multi-level feature extraction approach, which enables the features to carry more comprehensive information.

Methods	Acc (%)	Spe (%)	Pre (%)	UAR (%)	F1 (%)
PSD-CNN ⁹	84.72	94.63	—	77.63	—
WST + CWT & 1D CNN + 2D CNN ¹⁵	98.57	98.00	98.00	98.00	98.00
DNN + stem_block + attention + DWconv ¹⁶	97.70	92.22	98.98	98.20	98.59
42 MFCCs-KNN ¹⁷	95.78	—	—	—	96.00
12 statistical features-SVM ¹⁸	97.78	97.78	—	97.78	—
CTENN ²¹	96.40	97.40	—	92.87	95.16
Conv-Encoder ²²	97.90	96.50	98.80	98.40	—
PIFS-ResNet18 ²³	85.00	—	—	—	86.00
Our method—FBPSI + MLFE-EBTC	98.73	95.65	98.88	99.53	99.21

Table 9. Classification results of normal and abnormal of the proposed methods and other literature in PCCD.

Classification-wise, CNNs are often applied to classify and recognize HS signals^{9,15}. Some researchers directly extract features and classify HS signals by building deep learning networks¹⁶. However, the network structure is still a complex learning model, which requires sufficient training data and computational resources to achieve optimal results. The machine learning model is a traditional and effective way of HS signal classification, and some researchers have focused on applying it to classify HS signals and achieved good results as well^{17,18}. The EBTC utilized in this work uses random sampling for each classifier, which is an advantage for unbalanced datasets, and combining the results of multiple classifiers makes the model more stable.

As shown in Table 9, on PCCD, compared with other literature, the specificity of our method is 95.65%, which is 2.35% lower than the best result of literature 15, and the precision is only 0.1% lower than that of literature 16. However, our method shows better performance on UAR, which means that our algorithm performs well on both normal and abnormal signal classification. In addition, the accuracy and F1 score of our method is 98.73% and 99.21%, respectively, which are the best results for detecting the disease more accurately and helping doctors to make more precise diagnoses in the early stage.

Again, as can be found in Fig. 7b, all 110 normal HS signals were classified correctly. Two NOHCM HS samples were classified as normal, and four samples were misclassified between the two pathologic HSs. This is because both OHCM and NOHCM exhibit abnormal myocardium hypertrophy, producing a similar heart murmur, thus leading to the classification error between the two. Second, the HS signal in the acquisition process is present in the background noise and other interference signals in the body, although the signal has gone in the denoising, there is still a subtle interference, this interference noise leads to the characteristics of the signal is blurred, thus generating classification errors. However, in general, good triple classification results were obtained in HCMD with accuracy, specificity, precision, UAR, and F1 scores of 98.12%, 99.03%, 98.26%, 98.00%, and 98.05%, respectively, which verified the effectiveness of the proposed method.

Conclusion

Aiming at the poor classification accuracy and feature redundancy of the previous research in the PCCD, this study proposes a multilevel feature encoding algorithm based on frequency-balanced power spectral intensity. After the FBPSI extraction, a series of experiments have been accomplished with different parameter settings including weights, window lengths, reference points, and decomposition methods to get the optimal encoder. Finally, 128-dimensional features are extracted based on the FBPSI envelope using multiple overlapping windows of length 9 and fed into the EBTC for classification. The proposed method is evaluated by applying five performance metrics and the results show that MLFE can improve the classification accuracy to 98.73%. The proposed method has better accuracy, UAR and F1 scores, and simple feature computation as compared to other literature algorithms. It also achieves good performance metrics on the self-collected HCM dataset with high accuracy of 98.12% which provides useful reference to help physicians in clinical diagnosis.

Future work: The algorithm still has some limitations, and the specificity of the algorithm needs to be improved. Explore the feasibility of unifying the sampling frequency of different datasets. In the future, we should add more kinds of HS data to our experimental dataset and apply more public datasets to validate our method. In addition, since it takes time to set the optimal parameters for different datasets, we will investigate an adaptive algorithm to automatically determine the optimal parameters for a more effective HS classification model.

Data availability

Some of the datasets generated and/or analyzed in this study are available at <https://archive.physionet.org/pn3/challenge/2016/>. please contact the corresponding author for additional questions about data access.

Received: 27 May 2024; Accepted: 14 August 2024

Published online: 25 November 2024

References

1. Baikai, M. J. O. *et al.* Self-reported cardiovascular disease risk factor screening among people living with HIV vs. members of the general population in Botswana: A community-based study. *J. BMC Public Health* **24**, 198. <https://doi.org/10.1186/s12889-024-17651-6> (2024).

2. World health statistics 2023. WHO. <https://www.who.int/publications/item/9789240074323> (2023)
3. Noncommunicable Diseases Fact Sheets. WHO. <https://www.who.int/news-room/fact-sheets/detail/noncommunicable-diseases> (2023)
4. Report on Cardiovascular Health and Diseases in China 2022: an Updated Summary. *Biomed Environ Sci.*, **36**(8), 669–701 (2023).
5. 2020 AHA/ACC Guideline for the Diagnosis and Treatment of Patients With Hypertrophic Cardiomyopathy. *Circulation* **142**(25), 558–631 (2020).
6. Maron, M. S. *et al.* Hypertrophic cardiomyopathy is predominantly a disease of left ventricular outflow tract obstruction. *Circulation* **114**, 2232–2239 (2006).
7. Zhang, J. F. & Zhang, G. Y. An improved residual neural network for heart sound classification. *Comput. Eng. Sci.* **44**, 1514–1520 (2022).
8. Zhang, X. L., Fang, Y., Liu, D. B., Wang, W. B. & Wang, H. B. Research on feature extraction algorithm of HCM heart murmur sound based on WER-PCA. *Space Med. Eng.* **33**, 59–65 (2020).
9. Xu, C. D., Xing, P. L., Zhou, J. & Ying, D. W. Classification of heart sounds using power spectral density and convolutional neural networks. *Comput. Eng. Appl.* **57**, 125–132 (2021).
10. Wang, Q. *et al.* Heart sound classification algorithm based on time-frequency combination feature and adaptive fuzzy neural network. *J. Biomed. Eng.* **40**, 1152–1159 (2023).
11. Pathak, A., Samanta, P., Mandana, K. & Saha, G. An improved method to detect coronary artery disease using phonocardiogram signals in noisy environment. *Appl. Acoust.* **164**, 107242. <https://doi.org/10.1016/j.apacoust.2020.107242> (2020).
12. Wang, J. M. *et al.* Optimum heart sound signal selection based on the similarity of power spectral density. *Chin. J. Med. Phys.* **39**, 1401–1406 (2022).
13. Luo, G. *et al.* Feature extraction method based on enhanced power spectral density for emotion analysis using EEG. *Chin. J. Med. Phys.* **39**, 349–356 (2022).
14. Yang, L. Y., Meng, T. H., Zhang, Q. Y. & Chao, S. Implementation of EEG emotion analysis via feature fusion. *J. Xidian Univ.* **49**, 95–102 (2022).
15. Lee, A. J. & Kwak, C. K. Heart sound classification using wavelet analysis approaches and ensemble of deep learning models. *Appl. Sci.* **13**, 11942. <https://doi.org/10.3390/app132111942> (2023).
16. Hu, Q. L., Hu, J. Q., Yu, X. Y. & Liu, Y. Automatic classification of heart sounds built on deep separable conversion. *Comput. Appl. Softw.* **40**, 154–159 (2023).
17. Nur, Y. F., Adnan, M. P. & Moo, K. L. An optimal approach for heart sound classification using grid search in hyperparameter optimization of machine learning. *Bioengineering* **10**, 45. <https://doi.org/10.3390/bioengineering10010045> (2023).
18. Yadav, A., Singh, A., Dutta, M. K. & Travieso, C. M. Machine learning-based classification of cardiac diseases from PCG recorded heart sounds. *Neural Comput. Appl.* **32**, 17843–17856 (2020).
19. Caracuel, M. M. *et al.* Systemic neurophysiological signals of auditory predictive coding. *Psychophysiology* **61**, 14544. <https://doi.org/10.1111/psyp.14544> (2024).
20. Tsai, K. H. *et al.* Blind monaural source separation on heart and lung sounds based on periodic-coded deep autoencoder. *IEEE J. Biomed. Health Inform.* **24**, 3203–3214 (2020).
21. Cheng, J. W. & Sun, K. X. Heart sound classification network based on convolution and transformer. *J. Sensors* **23**, 8168. <https://doi.org/10.3390/s23198168> (2023).
22. Wang, J. L. *et al.* Multiclassification for heart sound signals under multiple networks and multi-view feature. *J. Meas.* **225**, 114022. <https://doi.org/10.1016/j.measurement.2023.114022> (2024).
23. Daniel, R. *et al.* CNN-based classification of phonocardiograms using fractal techniques. *J. Biomed. Signal Process. Control.* **86**, 105186. <https://doi.org/10.1016/j.bspc.2023.105186> (2023).
24. Fahim, A. M. *et al.* Linear predictive coding distinguishes the spectral EEG features of Parkinson's disease. *J. Parkinsonism Relat. Disord.* **79**, 79–85 (2020).
25. Turker, T., Erhan, E. & Sengul, D. An automated snoring sound classification method based on local dual octal pattern and iterative hybrid feature selector. *J. Biomed. Signal Process. Control.* **63**, 102173. <https://doi.org/10.1016/j.bspc.2020.102173> (2021).
26. Wang, Y., Wang, H. B. & Liu, L. H. Noise reduction for heart sound based on wavelet transform. *J. Inform. Electron. Eng.* **08**, 303–307 (2010).
27. Abdelhakim, S., Amine, D. S. M. E. & Fadia, M. Maximal overlap discrete wavelet transform-based abrupt changes detection for heart sounds segmentation. *J. Mech. Med. Biol.* <https://doi.org/10.1142/S0219519423500173> (2023).
28. Nsaif, M. Y. *et al.* A new voltage based fault detection technique for distribution network connected to photovoltaic sources using variational mode decomposition integrated ensemble bagged trees approach. *J. Energies.* **15**, 7762. <https://doi.org/10.3390/en1507762> (2022).
29. Ajay, D. *et al.* A flexible analytic wavelet transform and ensemble bagged tree model for electroencephalogram-based meditative mind-wandering detection. *J. Healthcare Anal.* **05**, 100286. <https://doi.org/10.1016/j.health.2023.100286> (2024).
30. Shah, A. S., Daniel, S., Anup, K. & Elmaghriby, A. A deep convolutional neural network-XGB for direction and severity aware fall detection and activity recognition. *J. Sensors* **22**, 2547. <https://doi.org/10.3390/s22072547> (2022).

Acknowledgements

This study was supported by the National Natural Science Foundation of China under Grant No. 61901393 and the Ministry of Education "chunhui project" of China No. 202201500.

Author contributions

Conceptualization, Y.F. and D.L.; writing–original draft preparation, Y.F., D.L. and H.L.; methodology, Y.F., D.L. and H.L.; software, Y.F. and H.L.; writing–review and editing, Y.F., W.W. and H.L.; validation, Y.F. and H.L.. All authors read and approved the final manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to D.L.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2024