Biomedical Signal Processing and Control 104 (2025) 107592

Contents lists available at ScienceDirect

Biomedical Signal Processing and Control

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

Parkinson's disease detection from voice signals using adaptive frequency attribute topology

Tao Zhang [a,b,*], Jing Tian [a,b], Zaifa Xue [a,b], Xiaonan Guo [a,b]

[a] School of Information Science and Engineering, Yanshan University, Qinhuangdao 066004, China
[b] Hebei Key Laboratory of Information Transmission and Signal Processing, Qinhuangdao 066004, China

ARTICLE INFO

Keywords:
Voice disorder
Parkinson's disease
Adaptive frequency attribute topology
Connected structural feature

ABSTRACT

Dysphonia is one of the early symptoms of Parkinson's disease (PD). The voice features extracted from voice signals can be effectively used for PD detection. Among them, the energy features extracted based on attribute topology have achieved preliminary results by combining time–frequency domain and structure representation. However, these energy features are susceptible to temporal fluctuation and are sensitive to weak coupling, which reduces their stability and representativeness. To solve these problems, this paper proposes the connected structural feature based on adaptive frequency attribute topology (CS-AFAT). Firstly, the energy statistic windows are established according to the distribution of spectral lines in the spectrogram for the aggregated statistics of energy information in time domain, which can effectively eliminate the influence of time disturbances. Secondly, the attribute topology is established to preform visualized statistics on the energy variation information of each point in the window. Finally, to reduce the influence of weak coupling in the topology, the adaptive threshold mechanism is designed to remove the edge with low coupling strength in the topology, so as to obtain more stable and representative features for PD classification. The results on two different language datasets show that the highest classification accuracy of CS-AFAT is 92.41% and 96.67%. The advantage of the proposed CS-AFAT is that it can obtain classification performance superior to or comparable to advanced energy features while having low-dimensional feature representation, which verifies the effectiveness and advancement of the CS-AFAT feature extracted by combining global energy information statistics and adaptive threshold mechanism.

1. Introduction

Parkinson's disease (PD) is a prevalent and incurable neurodegenerative disease, which is characterized by clinical manifestations including rigidity, resting tremor, and postural instability [1,2]. The diagnosis of PD in clinical settings primarily relies on neurologists evaluating motor dysfunction to obtain the Unified Parkinson's Disease Rating Scale [3,4]. However, when motor dysfunction becomes apparent, it often indicates that upwards of 50 % of dopaminergic neurons have already been damaged and are irreparable. [5]. Therefore, it is crucial to detect PD during the prodromal stages. It is reported that approximately 90 % of patients with PD (PWP) experience hypokinetic dysarthria [6,7]. The manifestation of voice disorders can occur as early as five years prior to the onset of noticeable motor dysfunctions [8], including a trembly and unstable voice quality as well as difficulties of articulation and breathing [9,10]. Therefore, utilizing voice data can help to develop an invasion, fast, and low-cost method for early detecting PD, which has important scientific value and practical significance [11,12].

In the past few years, many studies have proposed using models obtained by signal processing and machine learning algorithms to automatically detect PD [13,14]. In the studies related to PD detection by voice signals, acoustic measurement features have achieved preliminary results and have been widely used. Voice recordings are usually quantified based on various acoustic measurement parameters such as jitter, shimmer, HNR and NHR, and transmitted to multiple classifiers to distinguish PWP and healthy controls (HC) [15,16,17]. The acoustic measurement parameters can measure the instability of voice signals and have a shared characteristic or value between normal and pathological voice [18]. Despite having the advantage of interpretability, the acoustic measurement features may be insufficient for performing advanced analysis [19]. Scholars proposed the Mel Frequency Cepstrum Coefficients to describe the voiceprint features of voice signals. Compared to acoustic measurement features, there has been an

* Corresponding author at: School of Information Science and Engineering, Yanshan University, Qinhuangdao 066004, China.
E-mail addresses: zhtao@ysu.edu.cn (T. Zhang), tianjing@stumail.ysu.edu.cn (J. Tian), xuezf@stumail.ysu.edu.cn (Z. Xue), guoxiaonan@ysu.edu.cn (X. Guo).

https://doi.org/10.1016/j.bspc.2025.107592
Received 5 September 2023; Received in revised form 6 January 2025; Accepted 19 January 2025
Available online 29 January 2025

improvement in classification accuracy by approximately 3 % [20]. However, the temporal information of the voice signal is lost as a result of applying the discrete Fourier transform within the window [16]. By combining variational mode decomposition and Hilbert spectrum, Veetil et al. [21] extracted the Hilbert cepstral coefficient from voice signals and classified them with a variety of machine learning classifiers. The results showed that superior classification performance is obtained on different datasets. Escalona et al. [22] extracted time-length features, MFCC based features, and spectral moment-based features from voice signals, and built a robust model based on these features. The classical features are extracted independently from either the time domain or the frequency domain, so capturing the discontinuity and mutation of voice signals is challenging, which leads to limited classification performance [23].

Time-frequency representation can model temporal and frequency changes jointly, capturing discriminate information in voice signals [23,24]. Aiming at addressing the constraints posed by classical features, researchers began to extract features by integrating the mixed information available in the time–frequency domain. Rios-Urrego et al. [25] proposed a transfer learning method based on Convolution Neural Network (CNN) to discriminate the spectrograms represented in the Mel-domain with the accuracy of 82 %. Zhang et al. [26] used CNN to classify the spectrograms of PWP and HC and achieved an accuracy of 91 %. Khaskhoussy et al. [10] carried out deep feature extraction of voice signals based on CNN and classified them into Support Vector Machine (SVM). The results implied the importance and potential of this model in Parkinson's disease detection. Hireš et al. [27] proposed a CNN-based ensemble method to analyze the spectrograms of vowels, which can effectively distinguish PWP and HC. Further, Hireš et al. [28] constructed log-frequency power spectrograms as the input to deep CNNs and evaluated the performance of deep CNNs for PD classification using the Xception neural network architecture. In order to reduce the training time of the CNN model, Guatelli et al. [29] proposed a method based on Extreme Learning Machine (ELM) to analyze the spectrogram of voice signals for detecting PD. Wodzinski et al. [30] utilized the modified ResNet algorithm to analyze the spectrograms of sustained vowel /a/ for detecting PD. The reported accuracy of this algorithm is above 90 %. The existing studies demonstrate that the deep features extracted from spectrograms have good classification potential in PD detection. However, deep learning models are driven by data and necessitate extensive amounts of data to be effectively trained. Additionally, the opaque nature of deep learning poses challenges for users to comprehend the underlying rationale behind the network's decision-making process in certain detections [27].

Researchers have directed their attention towards extracting more intricate and effective handcrafted features within time–frequency domain to address the limitations of deep learning. In order to alleviate the problem of small amount of trained PD data, Javanmardi et al. [31] generated new data by manipulating the time–frequency domain representation of voice signals. This data augmentation method effectively improved the accuracy of voice detection. Tuncer et al. [32] extracted effective texture features and statistical features from speech signals based on factor lattice pattern and statistical feature extractors, and achieved high classification performance by combining dynamic and self-organized feature engineering components. The statistical features related to energy variation show excellent classification performance. Zhang et al. extracted the local gradient statistical features of energy points from short-time Fourier transform (STFT) and Mel domain spectrogram respectively, and the results indicated that these features are capable of effectively discriminating individuals with PWP from HC [33,34]. Zhang et al. [35] extracted vocal cord information by performing empirical mode decomposition on the voice signals. They further obtained energy direction features from the decomposed sub-signals. The classification performance was improved compared with [33,34]. Zhang et al. [36] employed tunable Q-factor wavelet transform (TQWT) to conduct a multi-scale decomposition on the voice signals for extracting the oscillation information. And the instantaneous energy variation feature was derived from the sub-signals, and its differentiation accuracy is comparable to the study [35]. The above researches show that the energy features of the voice signals extracted from time-frequency hybrid domain have good discriminative ability in PD detection. However, the original energy direction features have the problem of high dimension, which may produce overfitting results on small datasets [37]. The energy variation trend information within each sub-region of the spectrogram is represented by the formal context, which can be further simplified by the co-occurrence direction attribute topology (CDAT). The results confirmed the generalization ability and effectiveness of the structural feature based on CDAT (SF-CDAT) on datasets of different languages [38].

However, the SF-CDAT feature proposed based on attribute topology has two limitations. On the one hand, the SF-CDAT feature is extracted based on short-time window of the spectrogram and is distributed dispersedly in the time domain, which indicates that the SF-CDAT feature is easily affected by the randomness of temporal fluctuation. To solve this problem, this paper constructs the frequency attribute topology, which establishes the frequency statistical window according to the spectral line distribution in the spectrogram, and uses the attribute topology to perform visual statistics on the energy information in the frequency statistics window, so as to obtain the global energy variation information in the spectrogram and eliminate the influence of time disturbance. On the other hand, there are weakly coupled topologies in SF-CDAT, which does not consider the sensitivity of weak coupling on the stability and representativeness of structural features. Aiming at this limitation, this paper sets the threshold for coupling to weaken the sensitivity of structural features to weak coupling. Further, the adaptive threshold selection method is proposed to improve the stability of features in different situations.

To sum up, in order to eliminate the randomness of time perturbations and reduce the sensitivity of SF-CDAT features to weak coupling, the adaptive frequency attribute topology (AFAT) is proposed to describe the variation trend of energy during the overall pronunciation process. Firstly, the energy statistic windows based on frequency are established based on the distribution of spectral lines. Secondly, attribute topology (AT) is established for visual statistics of the gradient information in the energy statistic window. Finally, the adaptive threshold is set to reduce the influence of weak coupling on the structure of AT and to establish a more stable and representative AFAT. And the effectiveness of the connected structural feature based on AFAT (CS-AFAT) is validated through the utilization of multiple classifiers.

The main contributions of the proposed method are summarized as follows:

(1) A novel connected structural feature based on the adaptive frequency attribute topology is proposed for detecting PD. This method utilizes frequency attribute topology to statistically analyze the global energy variation information in the spectrogram, and improves the topology structure corresponding to the energy statistical window by designing the adaptive threshold, thereby enhancing the stability of feature representation and the prediction performance of the model.
(2) To prevent random interference caused by the temporal attributes, the energy statistic windows based on frequency are established to aggregately count the energy information throughout the entire pronunciation time. These energy statistical windows are established based on the distribution of spectral lines in the spectrogram, which can effectively describe the characteristics of energy variation in the frequency domain.
(3) To enhance the stability and representativeness of the topology structure, the adaptive threshold is set to reduce the sensitivity of topology structure to weak coupling. Among them, the threshold of coupling strength is adaptively adjusted through feedback on

the classification accuracy of the model, which is conducive to improving the universality of the model on different datasets.

(4) The experiment results on two different language datasets demonstrate that the proposed CS-AFAT feature with lower dimension have strong generalization ability to different classifiers and achieve excellent classification performance. These results validate the effectiveness of the proposed CS-AFAT feature in PD detection.

2. Methods

In this paper, the AFAT is proposed to characterize the energy variation trend in the spectrogram, and on this basis, the CS-AFAT features are extracted to distinguish PD and HC. The overall architecture of the extraction method of CS-AFAT features is presented in Fig. 1. Firstly, the spectrograms can be obtained by applying STFT to the voice signals, and the frequency-based energy statistic windows are established. Then, the gradient information is calculated for each energy point in the energy statistic window, and the energy fluctuation information is statistically analyzed by the AT. Finally, the weak coupling edges of AT are eliminated by the adaptive threshold to obtain a more stable and structurally representational AFAT. The count of connected components of AFAT is sent to the different classifiers for performance verification.

2.1. Representation of energy variation information

The vocal structures of PWP such as the diaphragm, and the muscles connecting the throat and vocal cords are damaged by PD, resulting in abnormal pronunciation [13]. The vocal cords show an approximate periodic vibration pattern during the pronunciation process of HC, which is distorted in the pronunciation of PWP [16]. The time–frequency representations of patients and controls are respectively illustrated in Fig. 2(a) and Fig. 2(d). From the perspective of spectrograms, there are obviously irregular energy fluctuations in the spectrogram of PWP, while the energy in the spectrogram of HC is more well-distributed and regular. The energy distribution information in the spectrogram can be used to distinguish PWP from HC.

The STFT is employed to obtain the spectrogram of the voice signals, considering the disparity in energy distribution observed in the spectrogram between PWP and HC. The time–frequency domain representation is obtained by performing the STFT on the voice signal $x(t)$, as shown in Eq. (1):

$$X(t,f) = \int_{-\infty}^{\infty} x(\tau) w(\tau - t) e^{-j2\pi f\tau} d\tau \tag{1}$$

where $w(\tau)$ is the hamming window. The time–frequency representation $X(t,f)$ is further converted into the spectrogram $E(t,f)$ to describe the temporal and spectral energy variations in the time–frequency domain:

$$E(t,f) = |X(t,f)|^2 \tag{2}$$

For each energy point in $E_n(t,f)$, the energy variation is most obvious in the gradient direction of this point. Therefore, energy fluctuations in the spectrogram can be described by calculating the gradient of each energy point. The gradient of the $E(t,f)$ at point (t_p,f_p) is:

$$\mathrm{grad}E(t_p,f_p) = E'_t(t_p,f_p)\boldsymbol{i} + E'_f(t_p,f_p)\boldsymbol{j} \tag{3}$$

where $\boldsymbol{i}$ and $\boldsymbol{j}$ represent unit vectors in the time direction and frequency direction respectively, $E'_t(t_p,f_p)$ and $E'_f(t_p,f_p)$ represent the partial derivatives of $E(t_p,f_p)$ in the direction of time and frequency:

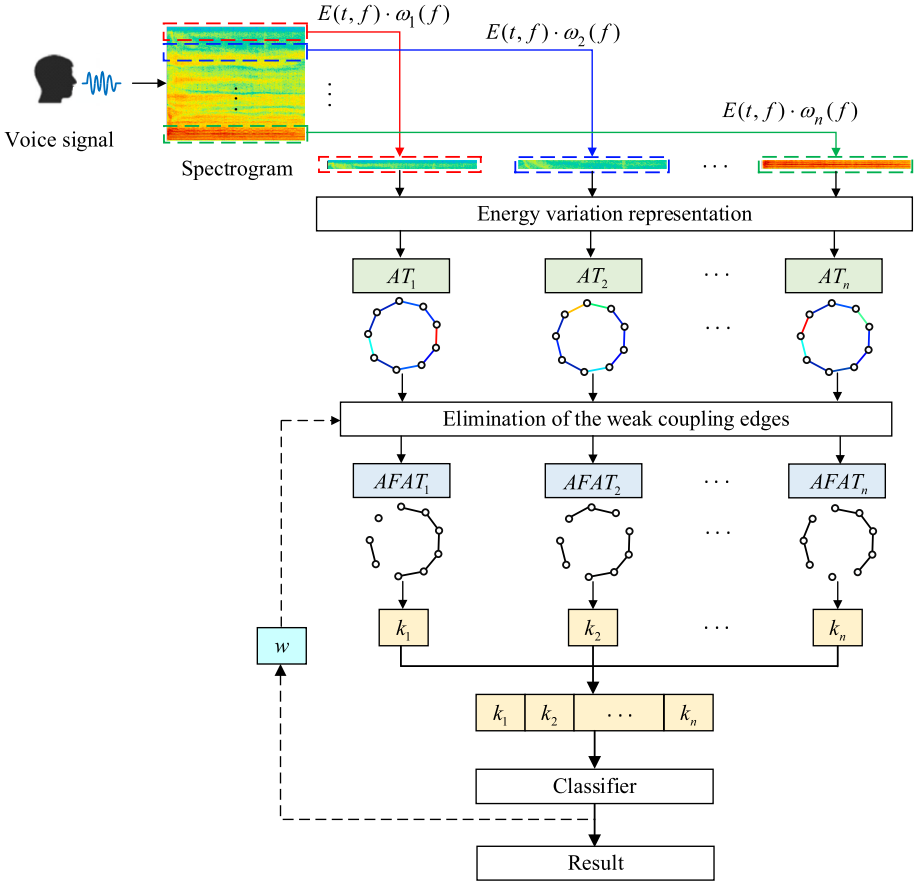

Fig. 1. The overview of the extraction method of CS-AFAT feature.

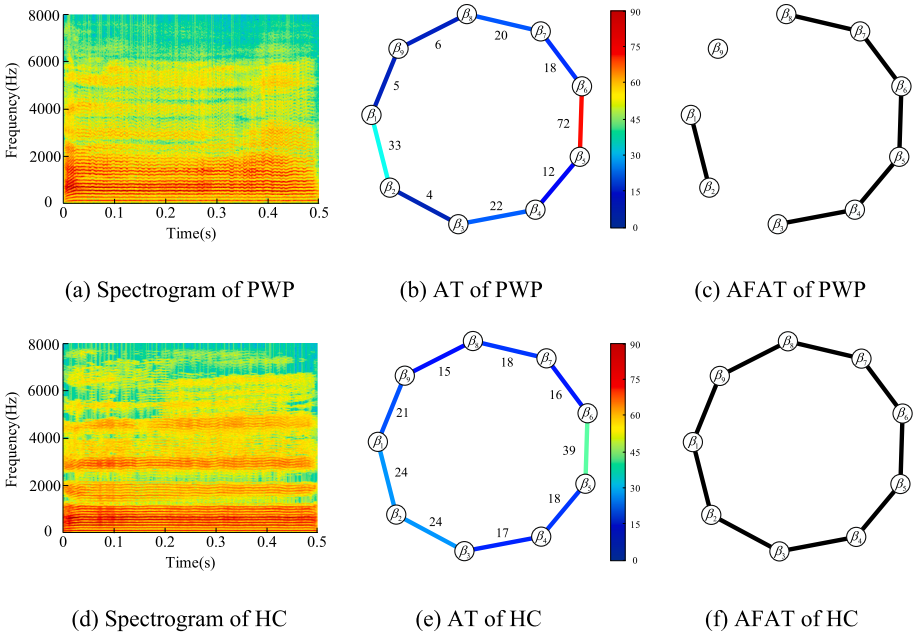

Fig. 2. The comparison of the spectrogram, AT and AFAT between PWP and HC. The voice sample of PD and HC respectively belongs to a 68-year-old woman and a 65-year-old woman in CPPDD dataset. The value of the edge weight in AT is visually displayed in the form of a colormap.

$$E_t'(t_p,f_p)=\lim_{\Delta t\to 0}\frac{E(t_p+\Delta t,f_p)-E(t_p,f_p)}{\Delta t} \tag{4}$$

$$E_f'(t_p,f_p)=\lim_{\Delta f\to 0}\frac{E(t_p,f_p+\Delta f)-E(t_p,f_p)}{\Delta f} \tag{5}$$

The magnitude and angle of gradient of the energy point (t_p,f_p) describe the energy fluctuation trend. The magnitude and angle $\vartheta(t_p,f_p)$ of gradient can be calculated as:

$$\left|\mathrm{grad}E(t_p,f_p)\right|=\sqrt{\left|E_t'(t_p,f_p)\right|^2+\left|E_f'(t_p,f_p)\right|^2} \tag{6}$$

$$\vartheta(t_p,f_p)=\arctan\left[\frac{E_f'(t_p,f_p)}{E_t'(t_p,f_p)}\right] \tag{7}$$

2.2. Establishment of attribute topology

To describe the fluctuation of energy in the spectrogram, it is necessary to conduct a statistical analysis of the variation trend of energy for all energy points. It can be found from the spectrogram that the energy is concentratedly distributed in different spectral lines in the time–frequency domain. The frequency-based energy statistics window $\omega_n(f)$ is established according to the distribution of spectral lines, which is used to describes the variation characteristics of energy in the spectrogram:

$$\omega_n(f)=\begin{cases}1, & (n-1)\cdot B\le f<n\cdot B\\ 0, & \text{otherwise}\end{cases} \tag{8}$$

where $\omega_n(f)$ represents the window function of the n-th energy statistic window, and B is the bandwidth of the window.

The energy representation in the n-th time–frequency domain window can be obtained by applying window $\omega_n(f)$ to the spectrogram $E(t,f)$:

$$E_n(t,f)=E(t,f)\cdot\omega_n(f) \tag{9}$$

After applying window $\omega_n(f)$ to the energy spectrum using the frequency statistical window, it can be represented as:

$$E(t,f)=\begin{bmatrix}E_1(t,f)\\ E_2(t,f)\\ \vdots\\ E_n(t,f)\end{bmatrix} \tag{10}$$

where $E_n(t,f)$ represents the energy information in the n-th energy statistic window, as shown in Fig. 3. The direction and length of each arrow correspond to the gradient angle and magnitude of each energy point.

Divide the gradient angle variation interval of energy points into D sub-intervals equally, and the angle label β_d of the d-th angle subinterval can be represented as:

$$\beta_d=\frac{(d-1)\bullet\pi}{D} \tag{11}$$

The gradient information of each energy point in $E_n(t,f)$ can be statistically analyzed using attribute topology:

$$AT_n=(V,Edge) \tag{12}$$

where $V=\{\beta_1,\beta_2,\cdots,\beta_D\}$ is the set of the topological vertices, and $Edge$ is the coupling relation matrix between attributes in AT_n. For $\forall\beta_i,\beta_j\in V$, the calculation of $Edge$ between β_i and β_j is as:

$$Edge(\beta_i,\beta_j)=\begin{cases}\#\left\{(t_p,f_p)\middle|\left|\mathbf{grad}E(t_p,f_p)\right|\neq 0,\vartheta(t_p,f_p)\in \mathrm{I}\right\}, & |i-j|=1\\ 0, & otherwise\end{cases} \tag{13}$$

where $\mathrm{I}=\left[\min(\beta_i,\beta_j),\max(\beta_i,\beta_j)\right)$. # represents the number of elements in the set, and (t_p,f_p) represents the energy point in $E_n(t,f)$.

Fig. 2(b) and Fig. 2(e) show the ATs of PWP and HC in a frequency-based energy statistic window. It can be found that although there is a

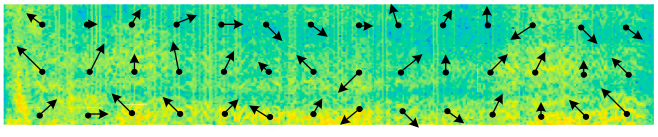

Fig. 3. The visual representation of energy variation trend in $E_n(t,f)$.

coupling relationship between the attributes in various directions, there are significant differences in the coupling strength between the attributes in the ATs of PWP. The weak coupling between attributes affects the description of the overall gradient statistical information by the topology structure within the window.

2.3. Extraction of CS-AFAT features

To enhance the stability of the topology and the representativeness of statistical information regarding energy variation, the $\mathrm{AFAT}_n = (V, Edge_{th})$ is established based on the $AT_n = (V, Edge)$. The coupling strength between attributes in AFAT can be represented as:

$$Edge_{th}(\beta_i, \beta_j) = \begin{cases} 0, & Edge(\beta_i, \beta_j) < w \\ 1, & \text{otherwise} \end{cases} \tag{14}$$

where w is the adaptive threshold, which can be obtain as Fig. 4 and Eq. (15).

$$w = \underset{w}{\operatorname{argmax}}\{F(ATs, w)\} \tag{15}$$

where $F(\cdot)$ represents the classification accuracy function of the model, and its input variables are AT and the adaptive threshold w.

Fig. 2(c) and Fig. 2(f) show the AFAT obtained after performing adaptive edge elimination on AT with the adaptive threshold $w = 12$. The connected structural characteristics of AFAT are used to statistically represent the energy variation in the energy statistics windows. Therefore, this study extracts the connected structural features based on AFAT, and the detailed algorithm is shown in Algorithm 1. k denotes the count of connected components in the topology structure, and it is initialized to 0.

Algorithm1: Connected structural feature extraction based on AFAT

Input: Adaptive frequency attribute topology $\mathrm{AFAT}_n = (V, Edge_{th})$
Output: Connected structural feature k_n
Process:
initialize a stack and the value of k_n
for all β in V **do**
  if β is not visited **then**
    $k_n = k_n + 1$, push β into the stack and mark β as visited
    while the stack is not empty **do**
      pop the top node and marked it as u
      mark u as visited
      for neighbor of u **do**
    if there is a neighbor node g not be visited **then**
    push g into the stack
    end if
  end for
      else
        return k_n
      end if
end for

The connected structural feature k_n extracted from the AFAT_n only describes energy variation trend over time within the specific frequency range. To comprehensively represent the fluctuation trend of energy in the spectrograms, the AFATs in energy statistics windows are cascaded from the frequency domain as shown in Fig. 5.

Extracting the connected structural features corresponding to the cascaded AFATs can obtain CS-AFAT features for classification:

$$\text{CS - AFAT} = \{k_1, k_2, \cdots, k_n\} \tag{16}$$

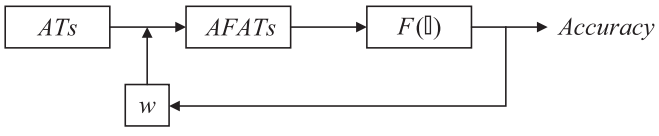

Fig. 4. The acquisition of the adaptive thresholds w.

where k_n represents the connected structural feature obtained in the AFAT_n.

3. Experiments

In this section, we introduce two datasets of PD voice that are utilized in the experiment, along with the experimental settings and results. The effectiveness of the proposed method is validated by the classification results of CS-AFAT on these datasets and comparing them with the results obtained by other existing methods.

3.1. Datasets

In order to make a fair comparison with the SF-CDAT features proposed in reference [38], this work conduct experiments on two different language datasets, including a Chinese PD voice dataset named CPPDD [39] and a Turkish PD voice dataset named SPDD [40], and strictly adheres to ethical standards.

The CPPDD dataset is a Chinese Parkinson's Disease voice dataset consisting of sustained Chinese vowels: a/o/e/i/u/ü. It comprises 38 PWP and 40 HC. The average age of the PWP is 61.35 years with a variance of 10.63, while the average age of the control group is 60.97 years with a variance of 9.63. The duration of Parkinson's disease in patients ranges from 0 to 6 years. The CPPDD dataset includes 40 male participants and 38 female participants. The speech samples were collected using the HYUNDAI HY-M11 microphone, positioned approximately 10 cm away from the participants' mouths. The sampling frequency of the microphone is set at 44.1 kHz. In this work, we utilized the sustained Chinese vowels "a" and "o".

The SPDD dataset is a Turkish Parkinson's Disease speech dataset that focuses on sustained vowels, namely the vowels "a" and "o". It consists of speech samples obtained from 20 HC and 20 PWP. The PWP have suffered from PD for 0 to 6 years. Among the participants, there are 24 male subjects and 16 female subjects. The average age of the patients is 64.86, with a variance of 8.97, while the average age of the healthy subjects is 62.55, with a variance of 10.79. The speech samples were collected using Trust MC-1500 microphones, with a sampling frequency of 44.1 kHz.

The comparative information of the CPPDD and the SPDD is shown in Table 1.

3.2. Experimental setup

Specific parameter settings for extracting features and classification during the experimental stages are given in this section. Furthermore, this section also includes the introduction of cross-validation methods and evaluation metrics for assessing the performance of features.

3.2.1. Parameter setting

During the extraction stage of the CS-AFAT features, the frequency bandwidth of the energy statistics window $\omega_n(t,f)$ is set to $B = 125$ Hz. There are 64 frequency-based energy statistic windows in each spectrogram. The number of gradient direction variation intervals of energy points is set to $D = 9$. The range of threshold for adaptive edge elimination on AT is set to $2 \leq w \leq 40$.

In the stage of classification, the popular classifiers in PD detection such as Logistic Regression (LR), Naïve Bayes (NB), SVM, Random Forest (RF), Multi-Layer Perception (MLP), and K-Nearest Neighbor (KNN) are used for classification [41,42,43]. Meanwhile, these classifiers are consistent with that used in [38] to facilitate performance comparison. The RBF kernel is utilized as the kernel function in the SVM, with gamma and C values set as 0.001 and 10, respectively. For the RF, the n_estimators parameter is set to 120, and the maximum depth is set at 30. The activation function employed in the MLP is ReLU, and the weight optimization is performed using the Adam algorithm. MLP adopts two

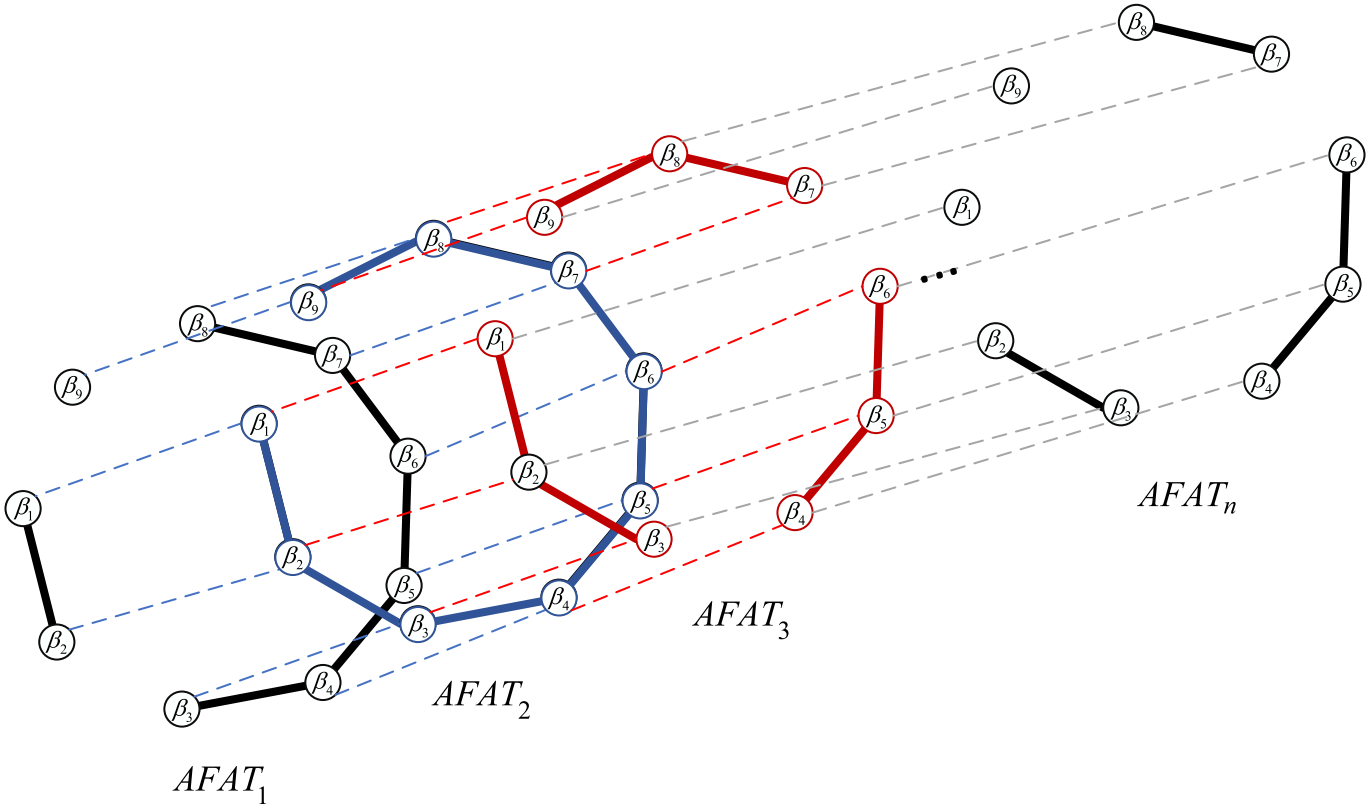

Fig. 5. Cascade the AFATs in the frequency domain.

Table 1
The comparison of CPPDD and SPDD.

Dataset	CPPDD	SPDD
Number of subjects (HC/PWP)	40/38	20/20
Gender distribution (Male/Female)	40/38	24/16
Average age (HC/PWP)	60.97/61.35	62.55/64.86
Age variance (HC/PWP)	9.63/10.63	10.79/8.97
Language type	Chinese	Turkish
Sustained vowels	"a", "o"	"a", "o"
Microphone model	HYUNDAI HY-M11	Trust MC-1500
Sampling frequency	44.1 kHz	44.1 kHz

hidden layers with both 20 nodes. And the parameter K in KNN classifier is set to 4.

3.2.2. Performance metrics

The performance metrics, including accuracy, recall, precision, and F1 scores, are used to assess the classification performance of the CS-AFAT feature. These performance metrics are derived from the analysis of the confusion matrix, which captures the counts of correctly and wrongly classified instances including HC and PWP. True positive (TP) and true negative (TN) represent the accurate classification of subjects, while false positive (FP) and false negative (FN) represent the misclassification instances. The confusion matrix serves as the basis for calculating these evaluation metrics.

The overall classification accuracy is represented by accuracy, which is the proportion of correctly classified samples among the total counts of evaluation samples:

$$Acc = \frac{TP + TN}{TP + TN + FP + FN} \tag{17}$$

Recall is derived by calculating the proportion of correctly classified samples TP among all actual positive samples:

$$Rec = \frac{TP}{TP + FN} \tag{18}$$

Precision is a metric used to measure the accuracy of predicted positive sample results. It represents the proportion of TP among all samples that are predicted to be positive:

$$Pre = \frac{TP}{TP + FP} \tag{19}$$

The F1 score is computed as the harmonic mean of recall and precision, representing the comprehensive classification performance of the classification model:

$$F1 = \frac{2 \cdot Rec \cdot Pre}{Rec + Pre} \tag{20}$$

3.2.3. Validation scheme

The sizes of CPPDD and SPDD datasets used in this paper are relatively small, which may lead to the risk of overfitting the classification method. To avoid this problem, the Stratified K-fold cross-validation and leave-one-object-out (LOSO) validation schemes are adopted in this paper.

Stratified K-fold cross-validation entails partitioning the dataset into K parts while preserving the proportional representation of diverse classes in each partition. The model is subsequently tested on one partition and trained on the remaining parts, repeating this procedure K times. The values of K in Stratified K-fold cross-validation are 5 and 10, which are set to determine the number of partitions or folds used in the evaluation process. By maintaining an equal proportion for each category as in the complete dataset, stratified K-fold cross-validation ensures that the testing set closely resembles the entire dataset as much as possible [44].

LOSO involves partitioning the dataset into several subsets, with each subset containing samples from all subjects except one. This

methodology guarantees that the data of each subject is utilized for testing precisely once, while the data of the remaining subjects are employed for training. Through the iterative process of systematically excluding one subject at a time, LOSO facilitates a comprehensive evaluation of the performance. The training set and testing set while using LOSO will not include samples from the identical subject, avoiding overly optimistic results [45].

In these two validation schemes, the Stratified K-Fold cross-validation is a technique that divides the dataset into K folds based on record distribution, while LOSO cross-validation is a subject-based approach ensuring that no data from the same speaker is present in both the training and testing sets, thus preventing overly optimistic evaluation outcomes. Therefore, the reliability and robustness of experiment results can be enhanced by using both subject-wise and record-wise folds [38].

3.3. Results and discussion

Feature extraction and classification are performed on separate datasets with different languages to validate the classification performance of the CS-AFAT features. This section provides the generalization validation of CS-AFAT to different classifiers, ablation experiments of important modules within CS-AFAT, comparative analysis of results with existing feature extraction methods, and the limitations of this study.

3.3.1. Generalization of CS-AFAT to different classifiers

The experimental results of CS-AFAT features based on various classifiers are presented in this section. The experimental results for the CPPDD and SPDD are displayed in Table 2 and Table 3, respectively. To prevent overfitting, different cross-validation methods are employed in this study. Among them, the LOSO approach ensures that the training and testing sets do not include voice samples from the same subject. Therefore, we evaluate the performance of CS-AFAT features based on the experimental results under LOSO validation as the foundation.

Table 2
The classification performance of CS-AFAT feature on CPPDD.

Classifier	Validation method	Acc(%)	Sen(%)	Pre(%)	F1(%)
LR	5-fold	92.93 ± 3.02	94.30 ± 3.62	94.03 ± 4.46	94.08 ± 2.50
	10-fold	92.91 ± 3.06	94.72 ± 3.98	93.77 ± 5.12	94.10 ± 2.49
	LOSO	90.31	92.95	90.95	91.94
NB	5-fold	89.27 ± 3.52	93.85 ± 4.50	88.93 ± 4.36	91.22 ± 2.85
	10-fold	90.30 ± 4.15	95.18 ± 4.32	89.55 ± 5.48	92.15 ± 3.29
	LOSO	89.27	94.27	88.43	91.26
SVM	5-fold	95.29 ± 1.47	96.90 ± 3.37	95.48 ± 3.76	96.08 ± 1.17
	10-fold	95.55 ± 2.16	97.83 ± 3.07	95.03 ± 3.68	96.33 ± 1.78
	LOSO	91.62	91.63	94.12	92.86
RF	5-fold	95.03 ± 3.26	97.34 ± 4.82	94.66 ± 4.10	95.87 ± 2.77
	10-fold	93.97 ± 5.42	96.03 ± 7.60	94.27 ± 5.60	94.90 ± 4.73
	LOSO	92.41	96.04	91.60	93.76
MLP	5-fold	90.31 ± 4.43	95.16 ± 2.88	89.33 ± 4.41	92.13 ± 3.50
	10-fold	90.55 ± 4.35	95.59 ± 4.15	89.41 ± 4.23	92.34 ± 3.48
	LOSO	87.70	93.83	86.59	90.06
KNN	5-fold	93.71 ± 3.28	96.01 ± 5.54	93.75 ± 3.70	94.74 ± 2.85
	10-fold	94.24 ± 4.08	96.92 ± 5.45	93.78 ± 4.30	95.20 ± 3.52
	LOSO	92.41	94.71	92.67	93.68

Table 3
The classification performance of CS-AFAT feature on SPDD.

Classifier	Validation method	Acc(%)	Sen(%)	Pre(%)	F1(%)
LR	5-fold	97.18 ± 1.40	98.39 ± 0.90	97.24 ± 2.23	97.80 ± 1.09
	10-fold	97.44 ± 2.42	98.40 ± 2.07	97.66 ± 2.67	98.01 ± 1.87
	LOSO	96.41	98.39	96.06	97.21
NB	5-fold	92.56 ± 1.46	94.70 ± 2.61	93.58 ± 1.60	93.90 ± 1.15
	10-fold	91.03 ± 2.03	95.98 ± 2.81	90.65 ± 3.78	93.16 ± 1.47
	LOSO	89.23	93.15	90.23	91.67
SVM	5-fold	97.69 ± 1.40	98.38 ± 1.71	98.00 ± 1.42	98.18 ± 1.13
	10-fold	97.69 ± 2.55	98.77 ± 2.79	97.63 ± 2.05	98.18 ± 2.08
	LOSO	96.67	96.77	97.95	97.35
RF	5-fold	96.15 ± 2.56	98.38 ± 1.68	95.83 ± 4.03	97.04 ± 1.93
	10-fold	96.41 ± 3.46	98.38 ± 2.81	96.23 ± 4.29	97.24 ± 2.66
	LOSO	94.36	97.98	93.46	95.67
MLP	5-fold	97.69 ± 1.90	98.79 ± 1.10	97.64 ± 2.09	98.21 ± 1.45
	10-fold	97.44 ± 2.09	98.80 ± 1.93	97.29 ± 2.56	98.02 ± 1.61
	LOSO	95.64	98.79	94.59	96.65
KNN	5-fold	96.15 ± 2.40	97.58 ± 2.62	96.47 ± 2.52	97.00 ± 1.90
	10-fold	95.90 ± 3.46	96.80 ± 4.13	96.82 ± 2.43	96.77 ± 2.78
	LOSO	93.08	94.76	94.38	94.57

On CPPDD, the highest classification accuracy of 92.41 % is achieved when using RF and KNN. On SPDD, the SVM obtains the highest classification accuracy of 96.67 %. The effectiveness of CS-AFAT in PD detection and its applicability to various datasets are confirmed by the classification results. In addition, excellent classification performance was achieved when using different classifiers, indicating that CS-AFAT features have strong generalization ability for different classifiers.

3.3.2. Ablation experiments

In this section, two sets of ablation experiments are conducted to examine the effects of both adaptive threshold and frequency-based energy statistical window on the performance.

3.3.2.1. Temporal aggregation statistics. Table 4 provides the accuracy of the ablation experiment of temporal aggregation statistics of energy information on CPPDD and SPDD. The validation scheme is LOSO. Symbol × represents obtaining the statistical information of energy variation based on a short-time energy statistics window, and symbol √ represents obtaining the temporal aggregation statistics of energy variation by establishing a frequency-based energy statistics window. To intuitively compare the advantages of temporal aggregation statistics, the accuracy histogram of multiple classifiers under LOSO is shown in Fig. 6.

It can be found in Fig. 6 that the performance of CS-AFAT obtained by using the frequency-based energy statistic window for temporal aggregation statistics is notably superior compared to that based on the short-time energy statistic window.

The irregular energy fluctuation in the spectrogram is random in the time domain, which is the underlying cause. Therefore, the feature extracted from the short-time window focuses on the statistical distribution of energy within the local time–frequency domain, which is easily affected by time randomness and cannot reflect the overall energy distribution differences. While the frequency-based energy statistic window is based on the spectral line distribution, which performs aggregately statistics on the gradient direction in the whole

Table 4
The accuracy of ablation experiments of performing temporal aggregation statistics with energy statistic window.

Dataset	Temporal aggregation statistics	LR	NB	SVM	RF	MLP	KNN
CPPDD	×	83.77	80.63	86.65	87.96	83.77	80.37
	√	90.31	89.27	91.62	92.41	87.70	92.41
SPDD	×	90.51	77.18	89.74	92.05	92.82	86.15
	√	96.41	89.23	96.67	94.36	95.64	93.08

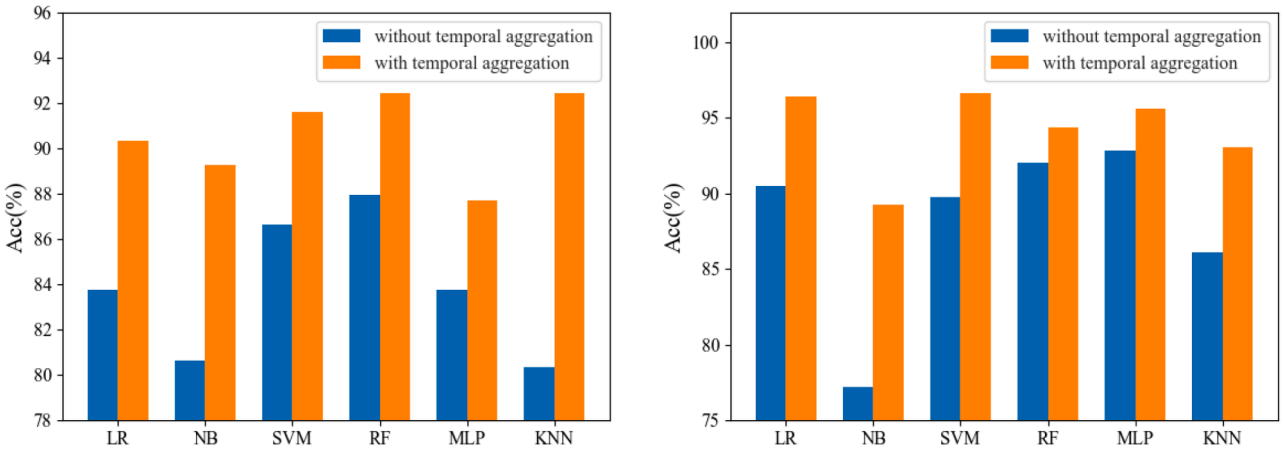

Fig. 6. The effect of temporal aggregation statistics on the classification performance.

pronunciation time to prevent the time attribute carried in the features from affecting the classification. In addition, the frequency-based energy statistic window has a large range and can achieve feature dimension reduction. Therefore, the frequency-based energy statistic window not only preserves complete energy information but also utilizes lower-dimensional features to achieve better classification performance.

3.3.2.2. Adaptive threshold. Table 5 shows the accuracy of ablation experiments on adaptive thresholds on CPPDD and SPDD datasets, and LOSO is used for cross-verification. Symbol × represents that the adaptive threshold processing is not performed on the attribute topology, and symbol √ indicates extracting connected structure features from the AFAT. To intuitively compare the advantages of adaptive threshold in feature extraction, the accuracy histogram of multiple classifiers based on LOSO is shown in Fig. 7.

It can be found that the classification performance of CS-AFAT after adaptive threshold processing is significantly improved. This is because AT counts the distribution of energy information in a certain frequency range of the spectrogram. The randomness of the energy variation leads to the coupling relationship between the attributes in all directions, resulting in the connectivity of the structure of the AT, as illustrated in Fig. 2(b) and Fig. 2(e). Therefore, AT is unable to structurally reflect the energy distribution differences in the spectrogram between PWP and HC. Adaptive thresholds can be used to eliminate weak coupling in AT to reduce the sensitivity of topological structures to weak coupling, and then improve the stability of topology structures and the representativeness of energy distribution information.

Table 5
The accuracy of ablation experiments about the adaptive threshold.

Dataset	Adaptive threshold	LR	NB	SVM	RF	MLP	KNN
CPPDD	×	84.82	84.03	82.72	84.03	83.77	81.15
	√	90.31	89.27	91.62	92.41	87.70	92.41
SPDD	×	70.77	48.46	70.77	68.97	67.44	66.92
	√	96.41	89.23	96.67	94.36	95.64	93.08

The AFATs obtained after adaptive threshold processing are significantly different between PWP and HC. Based on the information in Fig. 2 (c) and Fig. 2(f), it can be observed that the AFAT of HC is connected, while there is an obvious disconnection between attributes in the AFAT of PWP. The difference in connected structure reflects the difference of energy fluctuation in pronunciation between PWP and HC. HC have a strong pronunciation control ability, corresponding to the stable and regular energy distribution in the spectrogram [38]. Therefore, there is a strong coupling relationship between attribute nodes in the AT of HC, which reflects a connected topology structure in AFAT. However, the pronunciation control ability of PWP is poor, and the energy fluctuation of the spectrogram is irregular. Therefore, there exists a weak coupling correlation among certain directional attributes in AT, which leads to disconnection between attribute nodes in AFAT after adaptive threshold processing.

3.3.3. Evaluating the proposed method in contrast to existing works

The CS-AFAT is compared with the existing works to make further assessment of the performance. To enhance the credibility of the comparison results, the datasets used in the existing works selected are the same as in this study. Firstly, the classical acoustic measurement features in PD detection are compared with CS-AFAT features. Secondly, the classification results of the proposed method are compared with those obtained by using CNN for spectrogram classification. In addition, EMD-EDF [35], IEV-TQWT [36], SFLG-FT [33], SFLG-Mel [34], and SF-CDAT [38] all extract energy features based on spectrograms, which are of high comparative value to the CS-AFAT features. Based on the findings from Table 6 and Table 7, it is evident that the CS-AFAT feature exhibits

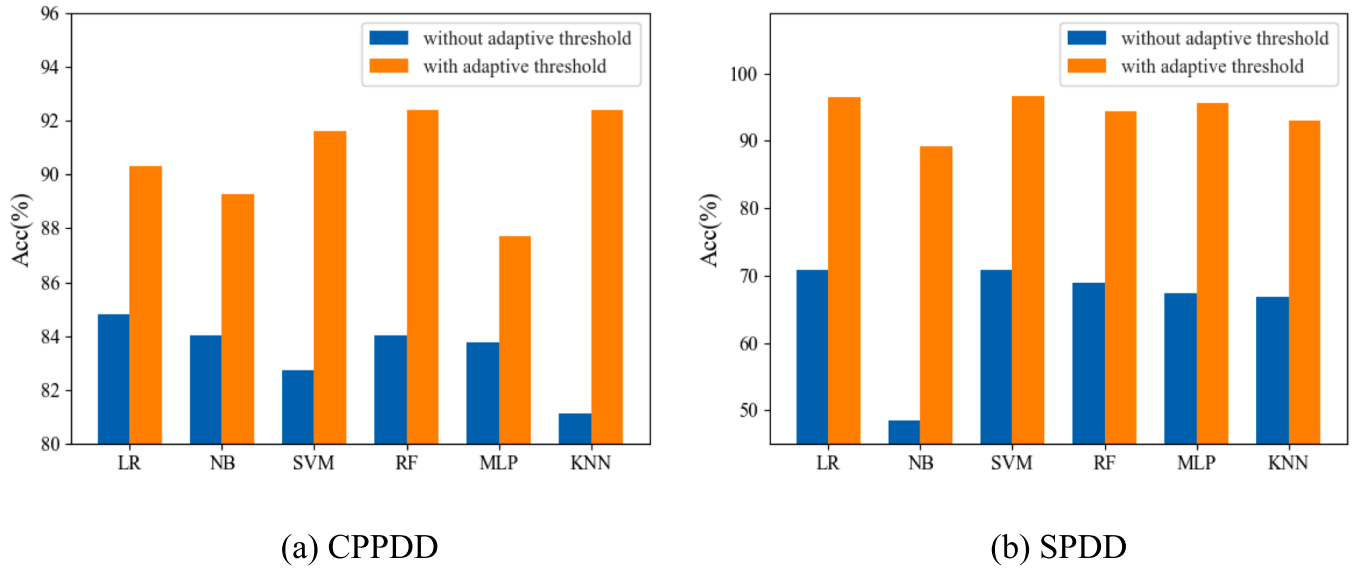

Fig. 7. Impact of the adaptive threshold on the accuracy.

Table 6
Evaluating the proposed method against existing feature extraction techniques on CPPDD.

Method	Featuredimension	Validationmethod	Acc(%)	Rec(%)	Pre(%)	F1(%)
Acoustic Measurements	21	LOSO	83.39	77.76	81.51	79.60
CNN [26]	9216	LOSO	91.00	85.21	92.12	88.53
EMD-EDF [35]	2880	LOSO	92.59	90.99	94.67	92.28
SFLG-Mel [34]	2880	LOSO	91.39	91.10	–	–
SFLG-FT [33]	2880	10-fold	90.81	82.28	–	–
IEV-TQWT [36]	2880	LOSO	**94.06**	94.53	**95.65**	**95.09**
SF-CDAT	320	LOSO	87.96	92.07	88.19	90.09
CS-AFAT (proposed)	64	LOSO	92.41	**96.04**	91.60	93.76

Table 7
Evaluating the proposed method against existing feature extraction techniques on SPDD.

Method	Featuredimension	Validationmethod	Acc(%)	Rec(%)	Pre(%)	F1(%)
Acoustic Measurements [16]	21	LOSO	79.00	–	–	75.00
CNN [26]	9216	LOSO	93.40	90.01	95.01	92.44
EMD-EDF [35]	2880	LOSO	96.54	92.01	**99.45**	96.20
SFLG-Mel [34]	2880	LOSO	95.33	95.45	–	–
SFLG-FT [33]	2880	10-fold	**97.27**	**97.11**	–	–
IEV-TQWT [36]	2880	LOSO	90.37	92.19	91.30	91.75
SF-CDAT	320	LOSO	92.82	96.37	92.64	94.47
CS-AFAT (proposed)	64	LOSO	96.67	96.77	97.95	

superior classification performance compared to most other features, and CS-AFAT has a lower feature dimension. In addition, the proposed CS-AFAT feature has the highest recall on CPPDD, indicating that CS-AFAT has strong discrimination ability for PWP. On SPDD, the F1 score of CS-AFAT is the highest among the existing features.

In the comparative analysis of existing methods, acoustic measurement features are widely used to quantify voice signals as biomarkers to detect PD. However, acoustic measurement features, as extracted in either the time domain or frequency domain, do not provide a comprehensive and detailed representation of the underlying information [37]. The acoustic measurement features are roughly extracted from the entire duration of the voice, which leads to limited accuracy of the models trained on these acoustic measurement features [46]. According to the comparison results of Table 6 and Table 7, the acoustic measurement features exhibit the poorest classification performance. In contrast, the proposed CS-AFAT feature demonstrates an enhancement of approximately 10 % in accuracy across two datasets, which validates the effectiveness of CS-AFAT features in PD detection.

CNN features are obtained by learning the energy difference of spectrograms between PWP and HC through CNN [26]. Both CNN features and CS-AFAT features are extracted based on spectrograms. Upon comparing their performance, it is evident that the differentiation performance of CNN features is lower than the proposed CS-AFAT features. This is because CNN is a data-driven model with poor learning ability for small-scale datasets, which may lead to overfitting and poor generalization [15]. While CS-AFAT feature extraction is carried out from the

perspective of voice signal processing and visualized through AFAT, making the physical meaning clearer.

EMD-EDF, IEV-TQWT, SFLG-FT, SFLG-Mel, and SF-CDAT all extract energy features from the time–frequency point of view. SFLG-FT [33] and SFLG-Mel [34] respectively extract energy features from spectrograms in the time–frequency domain and Mel domain, while both have the problem of high feature dimension. Under 10-fold cross-validation, the CS-AFAT outperforms SFLG-Mel in terms of classification performance. Conversely, SFLG-FT exhibits higher accuracy compared to the CS-AFAT. However, it is worth noting that overly optimistic results may arise because the training sets and testing sets may contain voice samples of the same subject while using 10-fold cross-validation. EMD-EDF [35] is extracted from the spectrogram of the IMFs component after EMD, which achieves high classification performance on CPPDD and SPDD. The IMF component obtained by EMD can provide resonance information of the vocal tract and vocal cords and then obtains a more effective feature representation [47]. By utilizing the TQWT [48,49], the IEV-TQWT feature [36] acquires the oscillation characteristics of voice signals to achieve a time–frequency representation that is more effective. On this basis, the spectrograms are employed to extract the features related to instantaneous energy variation. The accuracy of EMD-EDF and IEV-TQWT is higher than that of the proposed CS-AFAT features on CPPDD. However, the EMD-EDF and IEV-TQWT features have higher dimensions and are easy to overfit on small-scale datasets. In addition, the accuracy and operational efficiency of the model can be impacted by the high-dimensional voice features [46].

SF-CDAT [38] is an improvement of SFLG-FT, which employs attribute topology for structural representation and dimensionality reduction of features. Despite being low-dimensional, the SF-CDAT feature achieves high classification accuracy on two datasets. The dimension of CS-AFAT is further reduced by the frequency-based statistical window. The dimension of CS-AFAT is only one-fifth of that of SF-CDAT, but its classification accuracy has been improved by about 4 % compared with SF-CDAT, and other performance metrics are all higher than those of SF-CDAT. SF-CDAT is extracted based on the short-time energy statistic window, carrying temporal attributes. The irregular fluctuations of energy occur randomly in the time domain. Therefore, the SF-CDAT feature is susceptible to temporal randomness interference and cannot reflect the overall energy distribution characteristics. The CS-AFAT establishes a frequency statistical window based on the distribution of spectral lines, which performs global statistical analysis on the energy gradient direction over the entire voice duration, thus better describing the time–frequency characteristics of voice signals. In addition, the calculation of formal context is not required in the proposed CS-AFAT feature, simplifying the process of feature extraction. The code for extracting proposed CS-AFAT feature is available at https://github.com/ysutaoteam/CS-AFAT.

3.3.4. Limitations

There are also some limitations in this study. Firstly, the CPPDD and SPDD datasets lack some clinical descriptions of patients, such as the Hoehn&Yahr stage or UPDRS score, and how many patients are in each stage. Secondly, to further validate the effectiveness and generalization ability of the CS-AFAT feature, more experiments should be carried out on more different datasets. Thirdly, the feature extraction in this study relies on the spectrograms obtained by STFT, without exploring the performance in other transform domains. In further work, the performance of features extracted from various transformation domains will be compared and evaluated.

4. Conclusion

This study introduces a novel CS-AFAT feature, which utilizes energy variation information from time-frequency domain and adaptive frequency attribute topology. The main objective of this study is to automatically differentiate individuals with PD from HC using the proposed CS-AFAT feature. The proposed method establishes a frequency-based energy statistic window based on the distribution of spectral lines, overcoming the interference caused by the temporal attribute in existing energy features. The energy information is structurally represented by the attribute topology, and the adaptive threshold is used to eliminate the weak coupling between attributes to improve the stability of the structural feature. The highest classification accuracy of CS-AFAT feature on the CPPDD and SPDD with LOSO is 92.41 % and 96.67 %. Compared with existing related work, the proposed CS-AFAT feature has achieved classification performance comparable to the advanced methods. At the same time, the CS-AFAT feature has a lower dimension, which effectively prevents the over-fitting results on a small-scale PD dataset. The results validate the progressiveness and effectiveness of the CS-AFAT feature in PD detection.

Regarding the future work, it will focus on expanding the dataset size and enriching the description of the clinical information of patients, so as to validate the generalization capability of the CS-AFAT feature and better understand the contribution of the work. This paper only considers datasets from two languages, but future work will extend the proposed CS-AFAT features to other languages, such as Spanish or English, and discuss the applicability of the proposed features contrastively, which will be an interesting research topic. In addition, combining the proposed method with other transform domains such as wavelet domain [50], fractional domain and Mel domain is also worth considering. Therefore, future research directions will focus on optimizing datasets and combining features with different transform domains.

CRediT authorship contribution statement

Tao Zhang: Supervision, Methodology, Funding acquisition, Formal analysis, Conceptualization. **Jing Tian:** Writing – original draft, Validation, Software, Methodology. **Zaifa Xue:** Writing – review & editing, Methodology, Formal analysis. **Xiaonan Guo:** Project administration, Investigation.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This work was funded by the National Natural Science Foundation of China (under Grant 62176229), the Natural Science Foundation of Hebei Province (under Grant F2024203014), and the Natural Science Foundation of Heilongjiang Province (under Grant LH2023H029).

Data availability

The authors do not have permission to share data.

References

[1] L. Ali, C. Zhu, M. Zhou, Y. Liu, Early diagnosis of Parkinson's disease from multiple voice recordings by simultaneous sample and feature selection, Expert Syst. Appl. 137 (2019) 22–28, https://doi.org/10.1016/j.eswa.2019.06.052.

[2] J. Jankovic, Parkinson's disease: clinical features and diagnosis, J. Neurol. Neurosurg. Psychiatry. 79 (4) (2008) 368–376, https://doi.org/10.1136/jnnp.2007.131045.

[3] C.G. Goetz, B.C. Tilley, S.R. Shaftman, G.T. Stebbins, S. Fahn, P. Martinez-Martin, W. Poewe, C. Sampaio, M.B. Stern, R. Dodel, et al., Movement Disorder Society-sponsored revision of the Unified Parkinson's Disease Rating Scale (MDS-UPDRS): scale presentation and clinimetric testing results, Mov. Disord. 23 (15) (2008) 2129–2170, https://doi.org/10.1002/mds.22340.

[4] B.E. Sakar, C.O. Sakar, G. Serbes, O. Kursun, Determination of the optimal threshold value that can be discriminated by dysphonia measurements for unified Parkinson's disease rating scale, in: In 2015 IEEE 15th International Conference on

Bioinformatics and Bioengineering, 2015, pp. 1–4, https://doi.org/10.1109/BIBE.2015.7367737.
[5] M.C. Rodriguez-Oroz, M. Jahanshahi, P. Krack, I. Litvan, R. Macias, E. Bezard, J. A. Obeso, Initial clinical manifestations of Parkinson's disease: features and pathophysiological mechanisms, Lancet Neurol. 8 (12) (2009) 1128–1139, https://doi.org/10.1016/S1474-4422(09)70293-5.
[6] J. Rusz, T. Tykalová, M. Novotný, D. Zogala, E. Růžička, P. Dušek, Automated speech analysis in early untreated Parkinson's disease: Relation to gender and dopaminergic transporter imaging, Eur. J. Neurol. 29 (1) (2022) 81–90, https://doi.org/10.1111/ene.15099.
[7] O.M. El-Habbak, A.M. Abdelalim, N.H. Mohamed, H.M. Abd-Elaty, M. A. Hammouda, Y.Y. Mohamed, M.A. Taifor, A.W. Mohamed, Enhancing Parkinson's disease diagnosis accuracy through speech signal algorithm modeling, CMC-Comput. Mat. Contin 70 (2022) 2953–2969, https://doi.org/10.32604/CMC.2022.020109.
[8] B. Harel, M. Cannizzaro, P.J. Snyder, Variability in fundamental frequency during speech in prodromal and incipient Parkinson's disease: a longitudinal case study, Brain Cogn. 56 (1) (2004) 24–29, https://doi.org/10.1016/j.bandc.2004.05.002.
[9] F. Amato, G. Saggio, V. Cesarini, G. Olmo, G. Costantini, Machine learning- and statistical-based voice analysis of Parkinson's disease patients: A survey, Expert Syst. Appl. 219 (2023) 119651, https://doi.org/10.1016/j.eswa.2023.119651.
[10] R. Khaskhoussy, Y.B. Ayed, Improving Parkinson's disease recognition through voice analysis using deep learning, Pattern Recognit. Lett. 168 (2023) 64–70, https://doi.org/10.1016/j.patrec.2023.03.011.
[11] Y. Li, X. Zhang, P. Wang, X. Zhang, Y. Liu, Insight into an unsupervised two-step sparse transfer learning algorithm for speech diagnosis of Parkinson's disease, Neural Comput. Appl. 33 (15) (2021) 9733–9750, https://doi.org/10.1007/s00521-021-05741-0.
[12] H.C. Tunc, C.O. Sakar, H. Apaydin, G. Serbes, A. Gunduz, M. Tutuncu, F. Gurgen, Estimation of Parkinson's disease severity using speech features and extreme gradient boosting, Med. Biol. Eng. Comput. 58 (2020) 2757–2773, https://doi.org/10.1007/s11517-020-02250-5.
[13] L. Moro-Velázquez, J.A. García, J.D. Arias-Londoño, N. Dehak, J.I. Godino-Llorente, Advances in Parkinson's disease detection and assessment using voice and speech: A review of the articulatory and phonatory aspects, Biomed. Signal Process. Control. 66 (2021) 102418, https://doi.org/10.1016/j.bspc.2021.102418.
[14] R. Dubey, M. Kumar, A.B. Upadhyay, R.B. Pachori, Automated diagnosis of muscle diseases from EMG signals using empirical mode decomposition based method, Biomed. Signal Process. Control. 71 (2022) 103098, https://doi.org/10.1016/j.bspc.2021.103098.
[15] N.P. Narendra, B.W. Schuller, P. Alku, The detection of Parkinson's disease from speech using voice source information. IEEE/ACM Trans. Audio, Speech, Lang. Process. 29 (2021) 1925–1936, https://doi.org/10.1109/TASLP.2021.3078364.
[16] C.O. Sakar, G. Serbes, A. Günduz, H.C. Tunc, H. Nizam, B.E. Sakar, M. Tutuncu, T. Aydin, M.E. Isenkul, H. Apaydin, A comparative analysis of speech signal processing algorithms for Parkinson's disease classification and the use of the tunable Q-factor wavelet transform, Appl. Soft Comput. 74 (2019) 255–263, https://doi.org/10.1016/j.asoc.2018.10.022.
[17] M.A. Little, P.E. McSharry, E.J. Hunter, J. Spielman, L.O. Ramig, Suitability of dysphonia measurements for telemonitoring of Parkinson's disease, IEEE Trans. Biomed. Eng. 56 (4) (2009) 1015, https://doi.org/10.1109/TBME.2008.2005954.
[18] I. Hammami, L. Salhi, S. Labidi, Voice pathologies classification and detection using EMD-DWT analysis based on higher order statistic features, IRBM. 41 (2020) 161–171, https://doi.org/10.1016/j.irbm.2019.11.004.
[19] L. Brabenec, J. Mekyska, Z. Galaz, I. Rektorova, Speech disorders in Parkinson's disease: early diagnostics and effects of medication and brain stimulation, J. Neural Transm. 124 (2017) 303–334, https://doi.org/10.1007/s00702-017-1676-0.
[20] A. Benba, A. Jilbab, A. Hammouch, Analysis of multiple types of voice recordings in cepstral domain using MFCC for discriminating between patients with Parkinson's disease and healthy people, Int. J. Speech. Technol. 19 (2016) 449–456, https://doi.org/10.1007/s10772-016-9338-4.
[21] I.K. Veetil, V. Sowmya, J.R. Orozco-Arroyave, E.A. Gopalakrishnan, Robust language independent voice data driven Parkinson's disease detection, Eng. Appl. Artif. Intell. 129 (2024) 107494, https://doi.org/10.1016/j.engappai.2023.107494.
[22] M.M. Escalona, Y. Campos-Roca, C.J. Pérez Sánchez, Enhancing noise robustness of automatic Parkinson's disease detection in diadochokinesis tests using multicondition training, Expert Syst. Appl. 260 (2025) 125401, https://doi.org/10.1016/j.eswa.2024.125401.
[23] B. Karan, S.S. Sahu, J.R. Orozco-Arroyave, K. Mahto, Non-negative matrix factorization-based time-frequency feature extraction of voice signal for Parkinson's disease prediction, Comput. Speech Lang. 69 (2021) 101216, https://doi.org/10.1016/J.CSL.2021.101216.
[24] A. Kacha, F. Grenez, J.R. Orozco-Arroyave, J. Schoentgen, Principal component analysis of the spectrogram of the speech signal: Interpretation and application to dysarthric speech, Comput. Speech Lang. 59 (2020) 114–122, https://doi.org/10.1016/J.CSL.2019.07.001.
[25] C.D. Rios-Urrego, J.C. Vásquez-Correa, J.R. Orozco-Arroyave, E. Nöth, Transfer learning to detect Parkinson's disease from speech in different languages using convolutional neural networks with layer freezing, In International Conference on Text Speech and Dialogue, Springer. 12284 (2020) 331–339, https://doi.org/10.1007/978-3-030-58323-1_36.
[26] T. Zhang, Y. Zhang, Y. Cao, L. Li, L. Hao, Diagnosing Parkinson's disease with speech signal based on convolutional neural network, Int. J. Comput. Appl. Technol. 63 (2020) 348–353, https://doi.org/10.1504/ijcat.2020.10032598.
[27] M. Hireš, M. Gazda, P. Drotár, N.D. Pah, M.A. Motin, D.K. Kumar, Convolutional neural network ensemble for Parkinson's disease detection from voice recordings, Comput. Biol. Med. 141 (2022) 105021, https://doi.org/10.1016/j.compbiomed.2021.105021.
[28] M. Hireš, P. Drotár, N.D. Pah, Q.C. Ngo, D.K. Kumar, On the inter-dataset generalization of machine learning approaches to Parkinson's disease detection from voice, Int. J. Med. Inform. 179 (2023) 105237, https://doi.org/10.1016/j.ijmedinf.2023.105237.
[29] R. Guatelli, V. Aubin, M. Mora, J. Naranjo-Torres, A. Mora-Olivari, Detection of Parkinson's disease based on spectrograms of voice recordings and Extreme Learning Machine random weight neural networks, Eng. Appl. Artif. Intell. 125 (2023) 106700, https://doi.org/10.1016/j.engappai.2023.106700.
[30] M. Wodzinski, A. Skalski, D. Hemmerling, J.R. Orozco-Arroyave, E. Noth, Deep learning approach to Parkinson's disease detection using voice recordings and convolutional neural network dedicated to image classification, Annu Int Conf IEEE Eng Med Biol Soc. 717–720 (2019), https://doi.org/10.1109/EMBC.2019.8856972.
[31] F. Javanmardi, S.R. Kadiri, P. AlkuA, A comparison of data augmentation methods in voice pathology detection, Comput. Speech Lang. 83 (2024) 101552, https://doi.org/10.1016/j.csl. 202023.101552.
[32] T. Tuncer, S. Dogan, M. Baygin, P.D. Barua, E.E. Palmer, S. March, E.J. Ciaccio, R. Tan, U.R. Acharya, FLP: Factor lattice pattern-based automated detection of Parkinson's disease and specific language impairment using recorded speech, Comput. Biol. Med. 173 (2024) 108280, https://doi.org/10.1016/j.compbiomed.2024.108280.
[33] T. Zhang, P. Jiang, Y. Zhang, Y. Cao, Parkinson's disease diagnosis based on local statistics of speech signal in time-frequency domain, J. Biomed. Eng. 38 (1) (2021) 21–29, https://doi.org/10.7507/1001-5515.202001024.
[34] T. Zhang, L. Lin, Y. Zhang, X. Niu, Statistical analysis of local gradient in mel transform domain for Parkinson's dysphonia, J. Front. Comput. Sci. Technol. 16 (2021) 2345–2356. http://fcst.ceaj.org/CN/10.3778/j.issn.1673-9418.2102055.
[35] T. Zhang, Y. Zhang, H. Sun, H. Shan, Parkinson disease detection using energy direction features based on EMD from speech signal, Biocybern. Biomed. Eng. 41 (2021) 127–141, https://doi.org/10.1016/j.bbe.2020.12.009.
[36] T. Zhang, J. Tian, Z. Xue, B. Yin, Y. Wang, A novel feature extraction method based on TQWT and instantaneous energy variation for Parkinson's disease detection, Biomed. Signal Process. Control. 85 (2023) 105087, https://doi.org/10.1016/j.bspc.2023.105087.
[37] T. Zhang, L. Lin, Z. Xue, A voice feature extraction method based on fractional attribute topology for Parkinson's disease detection, Expert Syst. Appl. 219 (2023) 119650, https://doi.org/10.1016/j.eswa.2023.119650.
[38] T. Zhang, L. Lin, J. Tian, Z. Xue, X. Guo, Voice feature description of Parkinson's disease based on co-occurrence direction attribute topology, Eng. Appl. Artif. Intell. 122 (2023) 106097, https://doi.org/10.1016/j.engappai.2023.106097.
[39] T. Zhang, P. Jiang, L. Li, X. Zhang, Dysphonic analysis of Parkinson's disease based on partially ordered topological graph, J. Biomed. Eng. 38 (1) (2019) 59–69, https://doi.org/10.3969/j.issn.0258-8021.2019.01.008.
[40] B.E. Sakar, M.E. Isenkul, C.O. Sakar, A. Sertbas, F. Gurgen, S. Delil, H. Apaydin, O. Kursun, Collection and analysis of a Parkinson speech dataset with multiple types of sound recordings, IEEE J. Biomed. Health Inform. 17 (4) (2013) 828–834, https://doi.org/10.1109/JBHI.2013.2245674.
[41] G. Solana-Lavalle, J.C. Galan-Hernandez, R. Rosas-Romero, Automatic Parkinson disease detection at early stages as a pre-diagnosis tool by using classifiers and a small set of vocal features, Biocybern. Biomed. Eng. 40 (2020) 505–516, https://doi.org/10.1016/j.bbe.2020.01.003.
[42] G. Solana-Lavalle, R. Rosas-Romero, Analysis of voice as an assisting tool for detection of Parkinson's disease and its subsequent clinical interpretation, Biomed. Signal Process. Control. 66 (2021) 102415, https://doi.org/10.1016/j.bspc.2021.102415.
[43] B.E. Sakar, G. Serbes, N. Aydin, Emboli detection using a wrapper-based feature selection algorithm with multiple classifiers, Biomed. Signal Process. Control. 71 (2022) 103080, https://doi.org/10.1016/j.bspc.2021.103080.
[44] P. Guerra, M. Castelli, N. Côrte-Real, Machine learning for liquidity risk modelling: A supervisory perspective, Econ. Anal. Policy. 74 (2022) 175–187, https://doi.org/10.1016/j.eap.2022. 02.001.
[45] A.S. Ozbolt, L. Moro-Velazquez, I. Lina, A.A. Butala, N. Dehak, Things to consider when automatically detecting Parkinson's disease using the phonation of sustained vowels: Analysis of methodological issues, Appl. Sci. 12 (2022) 991, https://doi.org/10.3390/app12030991.
[46] Z. Xue, H. Lu, T. Zhang, J. Xu, X. Guo, A local dynamic feature selection fusion method for voice diagnosis of Parkinson's disease, Comput. Speech Lang. 101536 (2023), https://doi.org/10.1016/j.csl.2023.101536.
[47] R. Sharma, S.R. Prasanna, R.K. Bhukya, R. Das, Analysis of the Intrinsic Mode Functions for Speaker Information, Speech Commun. 91 (2017) 1–16, https://doi.org/10.1016/j.specom.2017.04.006.
[48] S. Ulukaya, G. Serbes, Y.P. Kahya, Resonance based separation and energy based classification of lung sounds using tunable wavelet transform, Comput. Biol. Med. 131 (2021) 104288, https://doi.org/10.1016/j.compbiomed.2021.104288.
[49] B. Cansiz, C.U. Kilinc, G. Serbes, Tunable Q-factor wavelet transform based lung signal decomposition and statistical feature extraction for effective lung disease classification, Comput. Biol. Med. 178 (2024) 108698, https://doi.org/10.1016/j.compbiomed.2024.108698.
[50] S. Ulukaya, G. Serbes, Y.P. Kahya, Wheeze type classification using non-dyadic wavelet transform based optimal energy ratio technique, Comput. Biol. Med. 104 (2019) 175–182, https://doi.org/10.1016/j.compbiomed.2018.11.004.