Archives of Acoustics Vol. 50, No. 1, pp. 69–79 (2025), https://doi.org/10.24425/aoa.2025.153650

Research Paper

Snoring Sounds Classification of OSAHS Patients Based on Model Fusion

Yexin LUO(1), Jianxin PENG(1)*, Li DING(2)*, Yikai ZHANG(1), Lijuan SONG(3), Qianfan ZHANG(1), Houpeng CHEN(1)

(1) *School of Physics and Optoelectronics, South China University of Technology* Guangzhou, China

(2) *School of Advanced Manufacturing Engineering, Hefei University* Hefei, China

(3) *State Key Laboratory of Respiratory Disease, Department of Otolaryngology-Head and Neck Surgery, Laboratory of ENT-HNS Disease, First Affiliated Hospital, Guangzhou Medical University* Guangzhou, China

*Corresponding Authors e-mails: phjxpeng@163.com (Jianxin Peng); gtxydingli@163.com (Li Ding)

(*received April 19, 2024; accepted November 25, 2024; published online February 19, 2025*)

Obstructive sleep apnea hypopnea syndrome (OSAHS) is a prevalent and detrimental chronic condition. The conventional diagnostic approach for OSAHS is intricate and costly. Snoring is one of the most typical and easily obtained symptom of OSAHS patients. In this study, a series of acoustic features are extracted from snoring sounds. A fused model that integrates a deep neural network, K-nearest neighbors (KNN), and a random under sampling boost algorithm is proposed to classify snoring sounds of simple snorers (SSSS), simple snoring sounds of OSAHS patients (SSSP), and apnea-hypopnea snoring sounds of OSAHS patients (APSP). The ReliefF algorithm is employed to select features with high relevance in each classification model. A hard voting strategy is implemented to obtain an optimal fused model. Results show that the proposed fused model achieves commendable performance with an accuracy rate of 85.76 %. It demonstrates the effectiveness and validity of assisting in diagnosing OSAHS patients based on the analysis of snoring sounds.

Keywords: obstructive sleep apnea hypopnea syndrome; snoring sounds; deep neural network; model fusion.

1. Introduction

Obstructive sleep apnea hypopnea syndrome (OSAHS) is a chronic sleep-related disease with the high incidence and great harm that is characterized by partial or complete collapse of the upper airway during sleep (Eckert *et al.*, 2007; Friedman *et al.*, 2004; Izci, Douglas, 2012; Osman *et al.*, 2018). There are many contributors to the collapse, including an ineffective pharyngeal dilator muscle function during sleep, a low threshold for arousal to airway narrowing during sleep, and unstable control of breathing, which may be caused by a narrow, crowded, or collapsible upper airway of OSAHS patients (Osman *et al.*, 2018). OSAHS not only adversely influences the sleep quality of patients, but also leads to hypertension, coronary heart disease, diabetes, cerebrovascular disease, other complications, and even causes sudden death at night (Redline *et al.*, 2010; White, 2005). The recent epidemiological survey has found that the prevalence of OSAHS among the global population ranged from 9 % to 38 % (Caron *et al.*, 2017). The elderly are the high incidence group that the prevalence rate of OSAHS is as high as 90 % for older males and 78 % for older females (Castillo-Escario *et al.*, 2019). Polysomnography (PSG) is the gold standard for diagnosing OSAHS by detecting respiratory disturbance events that mainly include apnea and hypopnea events (Minaritzoglou *et al.*, 2008). The apnea-hypopnea index (AHI) is obtained by PSG to measure the average number of respiratory disturbance events per hour during sleep. According to the American Academy of Sleep Medicine (AASM), subjects can be diagnosed as a simple snorer, mild, moderate, and severe OSAHS patient based on AHI ≤ 5, $5 <$ AHI ≤ 15, $15 <$ AHI ≤ 30, and AHI > 30, respectively (Berry *et al.*, 2012). The PSG requires more than 15 sensors connected to the patients that needs to be operated and checked by professional doctors in the hospital to monitor multiple biological signals of the test subject during sleep. The expensive cost, inconvenient device, and complex process limit the wide use of PSG that cause OSAHS to be a serious disease with a low diagnostic rate (Gottlieb, Punjabi, 2020; Osman *et al.*, 2018). The high prevalence and low diagnostic rate make the OSAHS be a public health problem that greatly influences the life quality of patients. With the increasing concern about sleep problems, researchers have been focused on studying various physiological signals during sleeping to assist in monitoring apnea and hypopnea events. The AASM indicates that one or more physiological signals, including oxygen, nasal airflow, electrocardiogram, electroencephalography, and snoring sound can be applied to detect apnea and hypopnea events to diagnose OSAHS (Berry *et al.*, 2012).

Snoring is the most prominent symptom of OSAHS patients that caused by the vibration of the upper airway (Gislason, Benediktsdottir, 1995; Pevernagie *et al.*, 2010; Sowho *et al.*, 2020; Ulualp, 2010). The acoustic features of snoring sounds can reflect the specific structure of the upper airway (Lugaresi *et al.*, 1988). Studies have indicated that there are obviously anatomical and non-anatomical structural differences of the upper airway between simple snorers and OSAHS patients (Azarbarzin, Moussavi, 2013; Fiz *et al.*, 1996; Markandeya *et al.*, 2018; Herzog *et al.*, 2008). Early studies have indicated that palatal snoring mainly occurs in simple snorers without any obstruction of the upper airways, while non-palatal snoring can be an indicator for OSAHS patients (Qian *et al.*, 2021). Recent work by Sun *et al.* (2023) has revealed that snoring sounds of OSAHS patients exhibit higher formant frequencies. Perez-Padilla *et al.* (1993) found that there was different energy distribution around 800 Hz of snoring sounds between simple snoring and those of OSAHS patients. Based on this condition, studies have been focused on identifying simple snorers and OSAHS patients. Solà-Soler *et al.* (2007) classified simple snorers and OSAHS patients based on AHI = 10, which yielded 93 % precision. Sun *et al.* (2023) applied two Gaussian mixture models to explore the acoustic characteristics of snoring sounds throughout the whole night to classify simple snorers and OSAHS patients with 90.0 % accuracy. Ding *et al.* (2024) applied a fused model obtained from different domain to classify snoring sounds during the whole night of simple snorers and OSAHS patients, which could exactly identify OSAHS patients. Furthermore, researchers (Lee, Ellis, 2012; Hou *et al.*, 2019; Alshaer *et al.*, 2019; Cheng *et al.*, 2022; Ding *et al.*, 2023) have explored the characteristics of snoring sounds obtained by different sleep stages during the whole sleep to diagnose the severity of OSAHS patients. Lee *et al.* (2012) showed that there was different energy distribution of snoring sounds during apnea-hypopnea events and simple sleeping. Ding *et al.* (2023) proposed VGG19-LSTM model to classify snoring sounds of simple snores and OSAHS patients with 99.31 % accuracy and 99.13 % sensitivity. A long short-term memory (LSTM) neural network was proposed to classify three-category snoring sounds related to the severity of OSAHS with 81.6 % accuracy (Cheng *et al.*, 2022). These studies have demonstrated the effectiveness and convenience of diagnosing OSAHS patients based on analysis of snoring sound.

The aforementioned classification results of snoring sounds have clearly demonstrated that the structure of the upper airway of OSAHS patients is obviously different from that of simple snorer. The abnormal structure could cause the occurrence of apnea and hypopnea respiratory events, as well as abnormal snoring sounds, which provided a strong basis for the diagnosis of OSAHS based on snoring sounds. Few studies (Cheng *et al.*, 2022; Song *et al.*, 2023; Sun *et al.*, 2023) focused on whether the abnormal upper airway may influence the normal sleep process of OSAHS patients. Since the characteristic of snoring sounds could reflect the structure of the upper airway, intuitively classifying snoring sounds of simple snorers (SSSS), apnea-hypopnea snoring sounds of OSAHS patients (APSP), and simple snoring sounds of OSAHS patients (SSSP) could explore the characteristics of the upper airway in the different stages of sleep for simple snorers and OSAHS patients, respectively. The classification results could indicate that whether the abnormal upper airway can be reflected by snoring sounds and whether the abnormal upper airway influence the normal sleep for OSAHS patients. The existing studies about snoring sound classification are based on a single classification model, which had limited classification accuracy and robustness. On this condition, the snoring sound classification tasks based on a fusion strategy might help to diagnose OSAHS patients more accurately.

In this study, a fused model is proposed to classify three kinds of snoring sounds, including SSSS, APSP, and SSSP. A series of acoustic features were extracted from snoring sounds. Three classifiers were first used to classify these three kinds of snoring sounds based on extracted acoustic features. Then a hard voting model fusion strategy was applied to integrate these basic models to obtain a model with relatively better classification performance and higher robustness.

2. Material and methods

2.1. Dataset

The 46 subjects selected from the PSG-Audio dataset are applied to validate the proposed method, including 8 simple snorers and 38 OASHS patients with different severities (Korompili *et al.*, 2021). All snoring sounds are collected clinically. When a subject undergoes the PSG (Alice 6), an ultra-linear measurement condenser microphone (Berringer ECM800) is placed approximately 1 m above the subject's bed to record snoring sounds during the whole night. Sound signals are sampled at 48 kHz with 24-bit resolution and saved as WAV. All recorded signals are enhanced and segmented by the noise reduction algorithm (Wang *et al.*, 2017). These enhanced snoring segments are labeled by ear-nose-throat (ENT) experts as SSSS, SSSP, and APSP. In the experiment, there are 73 373 effective snoring segments extracted from all 46 subjects, including 12 967 SSSS, 44 748 simple SSSP, and 15 658 APSP. These snoring sounds are divided into a training set and a validation set by the ratio of 4:1.

2.2. Proposed fused model

In the work, a fused model is proposed to classify SSSS, SSSP, and ASSP to explore structures of the upper airway of simple snorers and OSAHS patients during sleeping. The overall structure of the proposed model is shown in Fig. 1. A series of acoustic features are firstly extracted to express snoring sounds. Three basic classifiers, including the deep neural network – DNN (Janiesch *et al.*, 2021), K-nearest neighbors – KNN (Zhang *et al.*, 2017), and random under sampling boost algorithm – RUSBoost (Seiffert *et al.*, 2010) are applied to classify these three types of snoring sounds. To adequately integrate these basic classifiers from different domains, a model fusion strategy based on hard voting is used to fuse these classifiers. That is to say, the final classification results of snoring sounds were obtained by averaging the probability of these three basic models.

The three basic classifiers used in this work are DNN, KNN, RUSBoost. KNN is one of the most mature and simplest machine learning classification algorithms with relatively high performance in different domains. The basic idea of KNN is to calculate the distance between the test sample and all training samples to obtain its nearest neighbors and then conduct KNN classification. Choosing the proper K-value is an important part for training the KNN model. The RUSBoost algorithm is an effective ensemble method for the classification task with the unbalanced sample distribution. RUSBoost incorporates random under-sampling technology to remove samples from the majority class at each boosting iteration of the Adaboost.M2 algorithm. Based on this strategy, RUSBoost could adequately apply samples of the majority class and solve the problem of unbalanced sample distribution. The parameters that RUSBoost needs to be trained are mainly concentrated on Adaboost.M2, including a base estimator, the learning rate, n-estimators, and so on. DNN is a useful technology for the classification task with large samples. In this work, a DNN structure with two hidden layers is constructed to classify snoring sounds. There are 100 neurons in the first hidden layer and 5 neurons in the second hidden layer. The loss function and the activation function used in the DNN are the logistic cost regression function and the sigmoid function, respectively. The optimizer used for training is Adam. The batch size and the learning rate are set as 64 and 0.05, respectively.

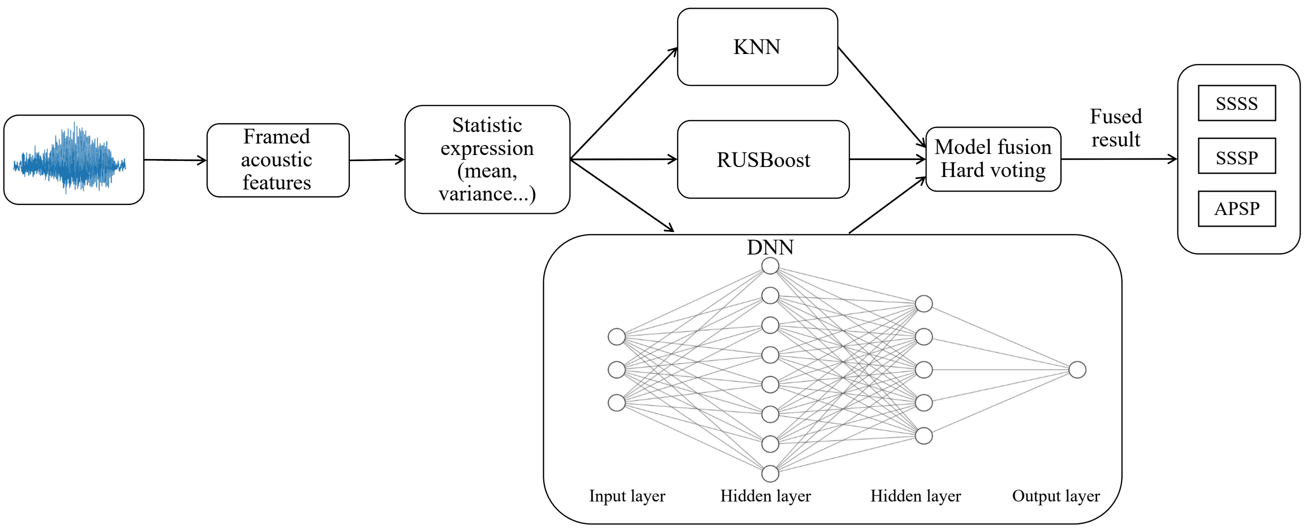

Fig. 1. Overall structure of the proposed system.

2.3. Feature extraction

In the work, a series of acoustic features from the time and frequency domains are extracted to express snoring sounds. There are 16 features with 45 dimensions, including the Mel-frequency cepstral coefficient (MFCC), the linear prediction coefficient (LPC), 800 Hz power ratio (PR800), the crest factor (CF), the fundamental frequency ($F0$), the pitch, formants, and a series of spectrum related features. Since the generation process of snoring sounds has a significant effect on its high frequency band and a smaller effect on the low frequency band, all snoring sounds are conducted pre-emphasizing that aims to compensate for the loss of high frequency components before the feature extraction. These pre-emphasized snoring sounds are framed by a hamming window with length of 20 ms and 50 % overlap. All features are firstly extracted for each frame. Statistic functions, including mean, minimum, maximum, and variance, are calculated by frames for each snoring segment to describe the feature distribution for each snoring segment.

2.3.1. Mel-frequency cepstral coefficient

The extraction of MFCC can be divided into five parts (Zheng *et al.*, 2001). Firstly, preprocessing, including pre-emphasis, and framing aims to compensate for the loss of high-value components. Then, performing fast Fourier transform on each frame signal to transform the time-domain signal into a frequency-domain signal. The spectral energy of each frame is calculated. Finally, the Mel filter is applied to transform frequency-domain signal into Mel-frequency scale to describe the human ear perception of frequency. The Mel-frequency (f_{mel}) could be obtained from the real liner frequency (f_{real}) by the equation:

$$f_{\text{mel}} = 2595 \cdot \log\left(1 + \frac{f_{\text{mel}}}{700}\right). \tag{1}$$

In this study, the average of all frames of an audio segment are taken as features. MFCCs with dimension of 13 were extracted.

2.3.2. Linear prediction coefficient

The basic concept of a linear prediction is that the current sampling value of audio can be approximately replaced by a linear combination of several past sampling values (Sun *et al.*, 2022). A unique set of prediction coefficients can be obtained by approximating the minimum mean square error of the actual audio sampling value and the linear prediction sampling value. LPC have the advantages of fast calculation and effective prediction. The 12-element LPC parameters of each sound segment were extracted, and the average value for each frame of every segment is calculated as the feature vector.

2.3.3. Power ratio

The PR is the ratio of power below and above a certain frequency f_0. It can roughly reflect the power distribution of audio signals divided by a certain frequency (Sun *et al.*, 2023). The PR can be expressed by:

$$\text{PR}_{f_0} = \log\left(\frac{\sum\limits_{f_i=0}^{f_0} (Y_i)^2}{\sum\limits_{f_i=f_0}^{f_C} (Y_i)^2}\right), \tag{2}$$

where f_C and Y are the cutoff frequency and spectrum of the audio signal, respectively. In this work, f_0 is set as 800 Hz. Four statistic features, including PR_{mean}, PR_{min}, PR_{max}, PR_{var} are calculated to express PR.

2.3.4. Fundamental frequency

The definition of $F0$ is the lowest oscillation frequency in a free oscillation system or the lowest frequency in a composite wave. It can reflect the opening and closing time of the vocal cords. In this work, the normalized autocorrelation function is applied to calculate $F0$ values for each frame audio signal. The average of all frames of an audio segment are taken as features.

2.3.5. Pitch

The tone is related to the fundamental frequency of the sound, reflecting the information of pitch. The average, minimum, maximum, and variance of all frames of an audio segment are taken as features, which are expressed as $\text{Pitch}_{\text{mean}}$, $\text{Pitch}_{\text{min}}$, $\text{Pitch}_{\text{max}}$, and $\text{Pitch}_{\text{var}}$, respectively.

2.3.6. Crest factor

The CF is defined as the ratio of the waveform peak to the effective value (Qian *et al.*, 2016):

$$\text{CF} = \frac{V_m}{V_e}, \tag{3}$$

where V_m is the maximum absolute value of an audio signal amplitude, and V_e is the root mean square value of the audio signal amplitude absolute value. It reflects the amplitude of changes in the audio signal in the time domain. The mean value of the peak factor of each frame of the signal is taken as a feature.

2.3.7. Spectrum related features

Spectrum related features are widely used in the analysis of snoring sounds. It can reflect important details of snoring sounds with different types. In this work, spectral cut-off frequency, spectral skewness, spectral slope, spectral variance, spectral kurtosis, spectral entropy, and spectral flux are extracted for further analysis (Sun *et al.*, 2023).

Spectral skewness is a measure of the direction and degree of skewness in the distribution of statistical data, which is a numerical characteristic of the degree of asymmetry in the distribution of statistical data. It is defined as the third-order standard moment of the sample, and the calculation formula is as follows:

$$\text{Skewness}(X) = E\left[\left(\frac{X-\mu}{\sigma}\right)^3\right] = \frac{k_3}{\sigma^3} = \frac{k_3}{k_2^{3/2}}, \tag{4}$$

where k_2 and k_3 represent the second- and third-order central moment, respectively.

Spectral slope is a measure of the speed at which the spectrum of an audio signal tilts towards high frequencies, typically calculated using linear regression. Spectral variance is used to measure the degree of dispersion of a sound signal. This can be expressed as follows:

$$\text{Var} = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2. \tag{5}$$

For spectral variance, which can reflect the interference of noise on data, this paper uses a noise power function of the carrier frequency, and the spectral variance of the signal can be obtained by the Fourier transform of its autocorrelation function:

$$V(\Omega) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{i2\pi\Omega\tau}\langle y(t)y(t+\tau)\rangle \,\mathrm{d}\tau. \tag{6}$$

Spectral kurtosis can be used to measure the steepness of the probability distribution of random variables. Take the average of the obtained results to obtain the average kurtosis in this work. In this work, the sample entropy is calculated for the entire effective snoring signal. Spectral traffic records the sum of squares of the normalized amplitude differences between two frames, which can describe the changes in adjacent frames. Its definition is

$$Fl_{i,i-1} = \sum_{k=1}^{Wl}\left(E_i(k) - E_{i-1}(k)\right)^2, \tag{7}$$

$$E_i(k) = \frac{x_i(k)}{\sum_{n=1}^{Wl} x_i(n)}, \tag{8}$$

where $E_i(k)$ is the normalized amplitude, and Wl is the sampling window length.

2.3.8. Formants

Formants are areas in the spectrum of audio signals where energy is relatively concentrated. It reflects the physical characteristics of the vocal tract, namely the degree of contraction of the throat. The first three formant frequencies of snoring sounds are extracted in this work, including the first formant ($F1$), the second formant ($F2$), and the third formant ($F3$). The average value of all frames is applied to express a piece of snoring sound.

2.4. Feature selection

Studies have indicated that extracted features not only determine the performance of a classification model, but also determine the complexity of the model and influence its computation cost (Kursa, Rudnicki, 2010; Li *et al.*, 2017). Selecting effective features with high discriminability and low complexity is an important step for machine learning. It can reduce the dimension of features and the complexity of the proposed classification model. In this work, the ReliefF algorithm is applied to select features by calculating the contribution of each feature to the classification task (Wu *et al.*, 2020).

The idea of ReliefF algorithm can be simply expressed as: if a feature has the same category to its nearest neighbor (with similar numerical values), the feature weight will be reduced; if the feature is different from its nearest neighbor category, increase its weight. The specific calculation method for the weight W is as follows. Firstly, setting the weights of all features W to 0. When calculating the weight of the j-th feature, an observation value x_o is randomly selected from the feature and the k-observation values are found in the dataset of each category of the feature that are closest in value to the observation value. Updating the weight of the feature parameter by the relationship between each nearest neighbor (x_n) and the observed value (x_o). Then repeating the iterative calculation until all parameters of the feature are traversed. The specific calculation formula is as follows:

1) when the observed value x_o is of the same category as the nearest neighbor x_n:

$$W_j^i = W_j^{i-1} - \frac{\Delta j(x_o, x_n)}{m}\cdot don; \tag{9}$$

2) when the observed value x_o is different from the category of the nearest neighbor x_n:

$$W_j^i = W_j^{i-1} + \frac{p_{y_n}}{1-p_{y_o}} * \frac{\Delta j(x_o, x_n)}{m}\cdot don, \tag{10}$$

where W_j^i is the weight of the i-th iteration of the j-th feature; $\Delta j(x_o, x_n)$ is the relative difference between x_o and x_n, where F_j represents the set of the j-th feature parameter, then the expression for $\Delta j(x_o, x_n)$ is

$$\Delta j(x_o, x_n) = \frac{|x_o - x_n|}{\max(F_j) - \min(F_j)}, \tag{11}$$

where don is the formal distance function between x_o and x_n:

$$don = \frac{\widetilde{d}_{on}}{\sum_{r=1}^{k} \widetilde{d}_{or}}, \tag{12}$$

$$\widetilde{d}_{on} = \exp\left[-\left(\frac{\text{rank}(o,n)}{\text{sigma}}\right)^2\right], \tag{13}$$

where $\text{rank}(o,n)$ is the corresponding position of a certain nearest neighbor x_n in the total nearest neighbor sorting table of x_o after sorting KNN by distance. Calculate sigma in Eq. (13) to change the scaling ratio, p_{y_o} is the prior probability of the category to which the observed value x_o belongs, p_{y_n} is the prior probability of the category to which the nearest neighbor x_n belongs.

3. Result

3.1. Feature selection

A strategy of feature selection based on the ReliefF algorithm is applied to select features with high robustness and low redundancy. Figure 2 displays the normalized weight value of each feature to the related label. Most features make significant contributions to this classification task, especially for MFCC and pitch features. The importance weights of MFCC1 to MFCC13 are higher than 0.1, which indicates that there is evidently different energy distribution on each frequency band divided by the Mel filter. The MFCC5 to MFCC8 yield the highest weights more than 0.14. These results show that the differences of snoring sounds of simple snorers, normal snoring sounds of OSAHS patients, and abnormal snoring sounds of OSAHS patients mainly concentrated on the low and middle frequency bands. $\text{Pitch}_{\text{var}}$ also has relatively high weight values, which means that the three kinds of snoring sound have different pitches.

Furthermore, the relationship between the dimension of selected features and the classification results is explored to select optimal features. Figure 3 shows the relationship between the dimension of selected features and the accuracy based on the KNN, RUSBoost, and DNN classifiers. The dimension of features has great influence on the classification results for all classification models. With the increase of the dimension of selecting features, the accuracy of classifiers gradually increases and tends to be stable. When the feature dimension exceeds the optimal one, the classification result will not increase with the increase of the feature dimension. The redundant features not only cannot improve the model classification performance, but also increase the computational complexity of the model. For different classification models, there are significant differences in the degree of influence of features and the dimension of optimal features. The optimal feature dimension is 16, 18, and 37 for KNN, RUSBoost, and DNN classifier, respectively. The related accuracy of KNN, RUSBoost, and DNN model with optimal features are 85.44 %, 84.45 %, and 83.91 %, respectively.

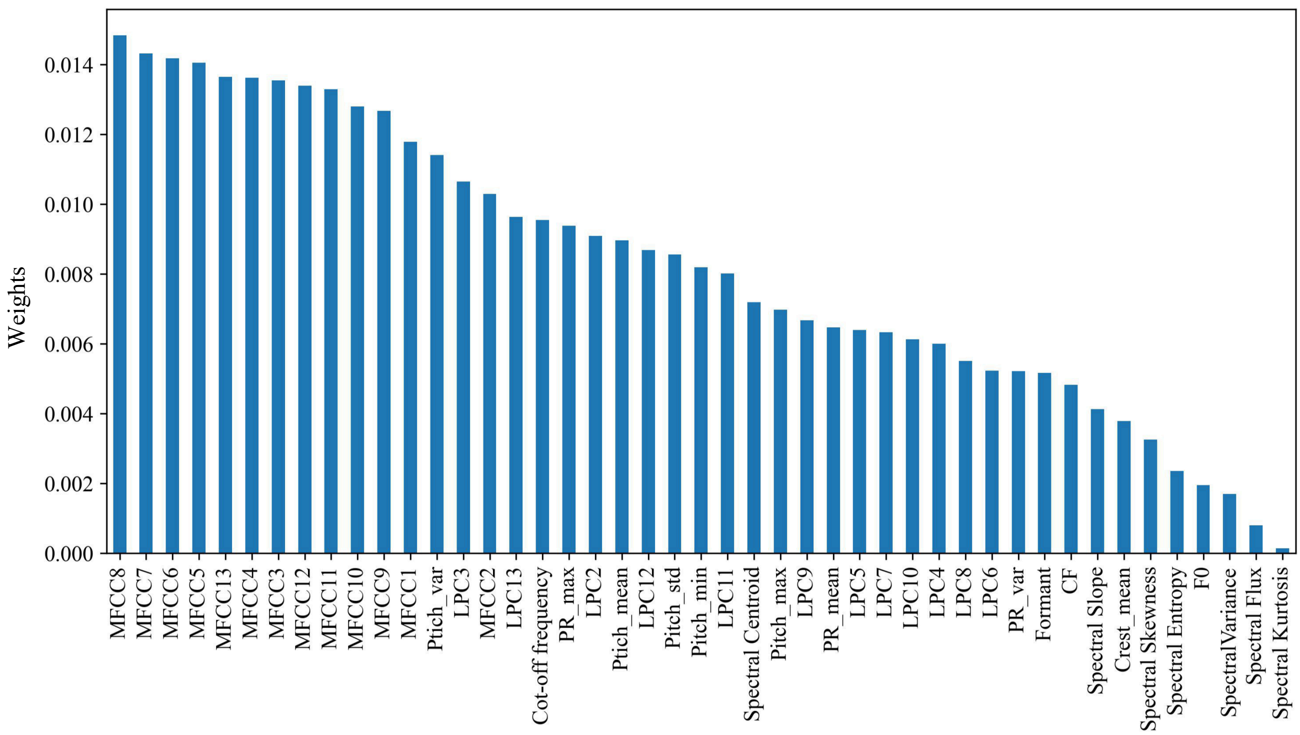

Fig. 2. Normalized weights of each feature obtained by ReliefF algorithm.

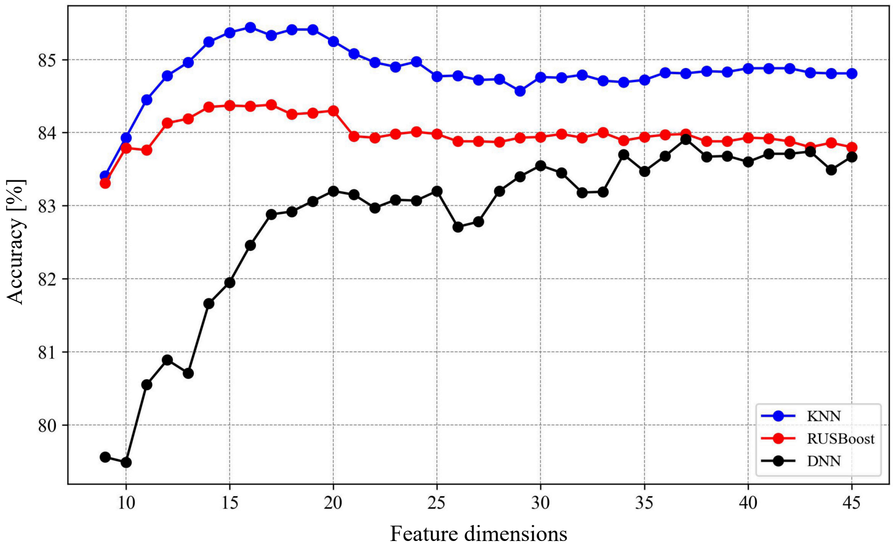

Fig. 3. Relationship of cross-validation average accuracy of KNN, RUSBoost, and DNN with the selected feature dimensions.

3.2. Classification results

Table 1 shows the classification results of SSS, SSSP, and ASSP based on KNN, RUSBoost, DNN classifiers under the original feature set. The accuracy obtained by KNN, RUSBoost, and DNN are 84.81 %, 83.80 %, and 83.67 %, respectively. Under the same feature set, different classifiers may have different emphases. KNN and DNN achieve much higher recall for SSSS and SSSP and lower recall for ASSP than RUSBoost. Specifically, the recall of SSSS, SSSP, and ASSP obtained by DNN are 97.53 %, 91.94 %, and 48.54 %, respectively. The recall of SSSS, SSSP, and ASSP obtained by KNN are 99.14 %, 91.26 %, and 54.50 %, respectively. The recall of SSSS, SSSP, and ASSP obtained by KNN are 97.99 %, 84.60 %, and 69.75 %, respectively.

Table 1. Classification results of SSS, SSSP, and ASSP based on different classifiers under the original feature set.

Snoring type	Evaluation	KNN	RUSBoost	DNN	Fused model
SSSS	Accuracy	0.8481	0.8380	0.8367	0.8556
	Recall	0.9914	0.9799	0.9753	0.9954
	Precision	0.9799	0.9704	0.9900	0.9847
	$F1$	0.9856	0.9751	0.9826	0.9900
SSSP	Recall	0.9126	0.8460	0.9194	0.9083
	Precision	0.8515	0.8886	0.8314	0.8655
	$F1$	0.8810	0.8667	0.8732	0.8864
ASSP	Recall	0.5450	0.6975	0.4854	0.5938
	Precision	0.6939	0.6179	0.6839	0.6987
	$F1$	0.6105	0.6553	0.5678	0.6420

To obtain classification results with higher robustness and stableness, the three basic models KNN, RUSBoost, and DNN are further fused by the voting strategy. The fused model adequately fuses the advantage of the three basic models. It achieves 85.56 % accuracy, which increases nearly 2 % compared with RUSBoost and DNN. The fused model not only maintains the relatively high recall for SSSS, but also significantly increases the recall of ASSP. The recalls obtained by the fused model are 10.84 % and 4.88 % higher than DNN and KNN, respectively. The recalls of SSSS and SSSP of fused model are 99.57 % and 90.21 %, respectively, which indicates that there are evident differences between SSSS and SSSP. The classification results imply that the upper airway structure of OSAHS patients on the normal sleep is different from that of simple snorers.

To obtain model with lower complexity and high performance, the feature selection strategy is applied in the model. Table 2 shows the classification results of SSS, SSSP, and ASSP based on KNN, RUSBoost, and DNN classifiers under the selected feature set with

Table 2. Classification results of SSS, SSSP, and ASSP based on different classifiers under the selected feature set.

Snoring type	Evaluation	KNN	RUSBoost	DNN	Fused model
SSSS	Accuracy	0.8544	0.8445	0.8391	0.8576
	Recall	0.9926	0.9840	0.9775	0.9957
	Precision	0.9859	0.9830	0.9909	0.9856
	$F1$	0.9892	0.9835	0.9842	0.9906
SSSP	Recall	0.9090	0.8521	0.9127	0.9021
	Precision	0.8617	0.8930	0.8385	0.8709
	$F1$	0.8847	0.8721	0.8740	0.8862
ASSP	Recall	0.5841	0.7072	0.5143	0.6162
	Precision	0.6974	0.6257	0.6782	0.6931
	$F1$	0.6357	0.6639	0.5850	0.6524

the dimension of 16, 18, and 37, respectively. Comparing Tables 1 and 2, the progress of feature selection not only reduces the complexity of the proposed fused model, but also improves the classification of SSSS, SSSP, and ASSP. Compared with the original feature set, the recall of ASSP obtained by the fused model conducting the feature selection improves value of 2.24 %.

Tables 3 and 4 illustrate the confusion matrices of KNN, DNN, RUSBoost, and its related fused model under the original feature set and selected feature set. There is a substantial distinction between snoring sounds of simple snorers and snoring sounds of OSAHS patients. For all classification models, recalls of SSSS are higher than 98 %. Under all test conditions, a certain amount of ASSP and SSSP are mislabeled, resulting in relatively lower recall and precision. The results of Tables 3d and 4d indicate that the proposed fused method could effectively merge the advantage of different classifiers and different features to relatively accurate SSSS, SSSP, and ASSP.

Table 3. Confusion matrices of SSS, SSSP, and ASSP based on different classifiers under the original feature set.

Real label	Predict label			Recall [%]
	SSSS	SSSP	ASSP	
a) KNN-under the original feature set				
SSSS	**3213**	23	6	99.1
SSSP	43	**10 209**	935	91.3
ASSP	23	1758	**2133**	54.5
Precision [%]	98	85.2	69.4	–
b) RUSBoost-under the original feature set				
SSSS	**3189**	24	28	98.4
SSSP	27	**9532**	1628	85.2
ASSP	28	1118	**2768**	70.7
Precision [%]	98.3	89.3	62.6	–
c) DNN-under the original feature set				
SSSS	**3161**	78	2	97.5
SSSP	26	**10 285**	876	91.9
ASSP	6	2008	**1900**	48.5
Precision [%]	99	83.1	68.4	–
d) Fusion-under the original feature set				
SSSS	**3226**	11	4	99.5
SSSP	28	**10 161**	998	90.8
ASSP	22	1568	**2324**	59.4
Precision [%]	98.5	86.6	69.9	–

Table 4. Confusion matrices of SSS, SSSP, and ASSP based on different classifiers under the selected feature set.

Real label	Predict label			Recall [%]
	SSSS	SSSP	ASSP	
a) KNN-under the selected feature set				
SSSS	**3217**	20	4	99.3
SSSP	30	**10 169**	988	90.9
ASSP	16	1612	**2286**	58.4
Precision [%]	98.6	86.2	69.7	–
b) RUSBoost-under the selected feature set				
SSSS	**3176**	37	28	98
SSSP	63	**9464**	1660	84.6
ASSP	34	1150	**2730**	69.8
Precision [%]	97	88.9	61.8	–
c) DNN-under the selected feature set				
SSSS	**3168**	70	3	97.8
SSSP	25	**10 210**	952	91.3
ASSP	4	1897	**2013**	51.4
Precision [%]	99.1	83.9	67.8	–
d) Fusion-under the selected feature set				
SSSS	**3227**	12	2	99.6
SSSP	29	**10 092**	1066	90.2
ASSP	18	1484	**2412**	61.6
Precision [%]	98.6	87.1	69.3	–

4. Discussion

In this study, a fused model based on KNN, RUSBoost, and DNN is proposed to classify SSSS, SSSP, and APSP. The ReliefF algorithm is applied to select optimal features in each basic model. The hard voting strategy is employed to fuse the three basic models. The feature selection and model fusion strategies evidently improve the classification performance of the proposed model. Experiment results show that the proposed model achieves 85.76 % accuracy. The recall and precision of SSSS are 99.57 % and 98.56 %, respectively. The recall and precision of SSSP are 90.21 % and 87.09 %, respectively. The recall and precision of ASSP are 61.62 % and 69.31 %, respectively.

Table 5 displays details of studies on the identification of APSP. Since there is no open snoring dataset with label, studies of analysis of snoring sounds are based on dataset collected and labeled by their own labs. The unavoidable situation makes it impossible to compare the performance of different classifica-

Table 5. Literature reviews about snoring sounds classification of OSAHS patients.

Author	Subjects	Feature	Validation method	Accuracy [%]
Cheng *et al.* (2022)	44	MFCC, LPC, Fbanks	LSTM	81.60
Ding *et al.* (2023)	50	Mel-spectrogram	VGG19+LSTM	85.21
Song *et al.* (2023)	40	Mel-spectrogram	CNN, ResNet, and XGBoost fused model	83.44
Shen *et al.* (2020)	32	MFCC, LPCC, and LPMFCC	LSTM	87.00
Hou *et al.* (2019)	120	MFCC	GMMs	80.00
This work	40	A series of acoustic features	KNN, RUSBoost, and DNN fused model	85.76

tion models directly. As Table 5 shows, these studies are capable of classifying snoring sounds with apnea-hypopnea events or without apnea-hypopnea events. Specifically, Cheng *et al.* (2022) extracted acoustic features including MFCC, LPC and used LSTM to classify SSSS, normal snoring sounds of OSAHS patients, and post-apnea snoring sounds of OSAHS patients with accuracy of 81.6 %. Their work had high recall for SSSS and normal snoring sounds of OSAHS patients and low recall for post-snoring sounds of OSAHS patients with value of 88.1 %, 93.4 %, and 63.5 %, respectively. The classification model proposed by this work achieved recall with values of 99.87 %, 90.21 %, and 61.26 % for SSSS, SSSP, and APSP, which are relatively better than the mentioned studies. The comparison demonstrates that the fused model yields higher classification result and better robustness. Since the snoring sound is generated by the vibration of the upper airway, the classification results of SSSS, SSSP, and APSP demonstrate that the structure of the upper airway is evidently different from that of OSAHS patients. Obesity, smoking, and other pathological reasons cause the upper airway of OSAHS patients gets narrow (Ghosh *et al.*, 2021). The narrow upper airway is the main reason for the occurrence of apnea and hypopnea events of OSAHS patients. The classification results of simple snoring sounds of OSAHS patients and apnea-hypopnea snoring sounds of OSAHS patients indicate that OSAHS patients snore continually throughout the whole night, which is caused by the narrow upper airway. It can be said that the narrow upper airway not only induces hypopnea and apnea events during sleep, but also negatively influences the normal sleep qualities and frequently causes snoring sounds. Furthermore, the high recall and precision of SSSS show solid experimental verification for identifying simple snorers and OSAHS patients based on analysis of snoring sounds. These studies mentioned in Table 5 are concentrated on distinguishing simple snoring sounds of OSAHS patients and apnea-hypopnea snoring sounds of OSAHS patients (Cheng *et al.*, 2022; Ding *et al.*, 2023; Shen *et al.*, 2020; Song *et al.*, 2023). The accuracies of all classification model are higher than 80 %. These results indicate that there are evident differences among snoring sounds occurred in different sleep stages for OSAHS patients.

Song *et al.* (2023) proposed a CNN, ResNet, and XGBoost fused model to classify snoring sounds occurred in different sleep stages and achieved 83.44 % accuracy. The classification model may be only concentrated on differences at one latitude and achieve limited classification results. The model fusion strategy based on different fusion methods is proposed to fuse basic classification models that has been widely used in different kinds of classification tasks. In this work, a hard voting fusion strategy is applied to fuse KNN, RUSBoost, and DNN classifiers. This method significantly increases classification recall and precision of SSSS, SSSP, and ASSP. It also improves the effectiveness and robustness of the proposed model. Experiment results show promising foreground for diagnosing severities of OSAHS patients based on analysis of snoring sounds.

There are also some limitations of the proposed model. Firstly, validation experiments of this work are conducted based on subject dependence. It mainly focuses on exploring differences among these types of snoring sounds. Further subject independent experiments should be conducted to validate the generation error and robustness of the proposed model. Moreover, the proposed model just focuses on exploring differences among snoring sounds occurred in different sleep stages. The relationship between apnea-hypopnea snoring sounds and apnea-hypopnea events should be studied to identify apnea-hypopnea events and estimate AHI values of OSAHS patients.

5. Conclusion

In this work, a fused model based on KNN, RUSBoost, and DNN is proposed to classify SSSS, SSSP, and APSP. Firstly, a series of acoustic features are extracted to express snoring sounds. Three classifiers KNN, RUSBoost, and DNN are independently trained. The ReliefF algorithm is applied to select features in each classification model. A hard voting strategy is used to obtain an optimal fused model. Experiment results show that the proposed fused model achieves high performance with accuracy of 85.76 %. The recalls of SSSS, SSSP, and APSP obtained by the proposed model are 99.87 %, 90.21 %, and 61.26 %, respectively. It demonstrates the effectiveness and validity of assisting in diagnosing OSAHS patients based on analysis of snoring sounds.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (grant no. 11974121) and National Youth Foundation of China (grant no. 81900927).

Data availability statement

The data used to support the findings of this study are available from the corresponding author upon request.

Ethical approval

This study was approved by the Ethics Committee of Guangzhou Medical University and an informed consent was obtained from each participant.

Conflicts of 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.

References

1. Alshaer H., Hummel R., Mendelson M., Marshal T., Bradley T.D. (2019), Objective relationship between sleep apnea and frequency of snoring assessed by machine learning, *Journal of Clinical Sleep Medicine*, **15**(3): 463–470, https://doi.org/10.5664/jcsm.7676.
2. Azarbarzin A., Moussavi Z. (2013), Snoring sounds variability as a signature of obstructive sleep apnea, *Medical Engineering and Physics*, **35**(4): 479–485, https://doi.org/10.1016/j.medengphy.2012.06.013.
3. Berry R.B. *et al.* (2012), Rules for scoring respiratory events in sleep: Update of the 2007 AASM manual for the scoring of sleep and associated events, *Journal of Clinical Sleep Medicine*, **8**(5): 597–619, https://doi.org/10.5664/jcsm.2172.
4. Caron C.J.J.M. *et al.* (2017), Obstructive sleep apnoea in craniofacial microsomia: Analysis of 755 patients, *International Journal of Oral and Maxillofacial Surgery*, **46**(10): 1330–1337, https://doi.org/10.1016/j.ijom.2017.05.020.
5. Castillo-Escario Y., Ferrer-Lluis I., Montserrat J.M., Jane R. (2019), Entropy analysis of acoustic signals recorded with a smartphone for detecting apneas and hypopneas: A comparison with a commercial system for home sleep apnea diagnosis, *IEEE Access*, **7**: 128224–128241, https://doi.org/10.1109/ACCESS.2019.2939749.
6. Cheng S. *et al.* (2022), Automated sleep apnea detection in snoring signal using long short-term memory neural networks, *Biomedical Signal Processing and Control*, **71**(Part B): 103238, https://doi.org/10.1016/j.bspc.2021.103238.
7. Ding L., Peng J., Song L., Zhang X. (2023), Automatically detecting apnea-hypopnea snoring signal based on VGG19 + LSTM, *Biomedical Signal Processing and Control*, **80**(Part 2): 104351, https://doi.org/10.1016/j.bspc.2022.104351.
8. Ding L., Peng J., Song L., Zhang X. (2024), Automatically detecting OSAHS patients based on transfer learning and model fusion, *Physiological Measurement*, **45**(5): 055013, https://doi.org/10.1088/1361-6579/ad4953.
9. Eckert D.J., Jordan A.S., Merchia P., Malhotra A. (2007), Central sleep apnea: Pathophysiology and treatment, *Chest*, **131**(2): 595–607, https://doi.org/10.1378/chest.06.2287.
10. Fiz J.A. *et al.* (1996), Acoustic analysis of snoring sound in patients with simple snoring and obstructive sleep apnoea, *European Respiratory Journal*, **9**(11): 2365–2370, https://doi.org/10.1183/09031936.96.09112365.
11. Friedman M., Ibrahim H., Joseph N.J. (2004), Staging of obstructive sleep apnea/hypopnea syndrome: A guide to appropriate treatment, *Laryngoscope*, **114**(3): 454–459, https://doi.org/10.1097/00005537-200403000-00013.
12. Ghosh P., Varma N.K.S., Ajith V.V., Suresh A. (2021), Upper airway and its association with neck circumference and hyoid position in OSA subjects – A cephalometric study, *International Journal of Current Research and Review*, **13**(6): 167–171, https://doi.org/10.31782/IJCRR.2021.13610.
13. Gislason T., Benediktsdottir B. (1995), Snoring, apneic episodes, and nocturnal hypoxemia among children 6 months to 6 years old: An epidemiologic study of lower limit of prevalence, *Chest*, **107**(4): 963–966, https://doi.org/10.1378/chest.107.4.963.
14. Gottlieb D.J., Punjabi N.M. (2020), Diagnosis and management of obstructive sleep apnea: A review, *Journal of the American Medical Association*, **323**(14): 1389–1400, https://doi.org/10.1001/jama.2020.3514.
15. Herzog M., Schmidt A., Bremert T., Herzog B., Hosemann W., Kaftan H. (2008), Analysed snoring sounds correlate to obstructive sleep disordered breathing, *European Archives of Oto-Rhino-Laryngology*, **265**(1): 105–113, https://doi.org/10.1007/s00405-007-0408-8.
16. Hou L., Zhang W., Shi D., Liu H. (2019), Estimation of apnea hypopnea index based on acoustic features of snoring, *Journal of Shanghai University (Natural Science)*, **25**(4): 435–444, https://doi.org/10.12066/j.issn.1007-2861.1942.
17. Izci B., Douglas N.J. (2012), Obstructive sleep apnea-hypopnea syndrome, [in:] *Obstructive Sleep Apnea: Causes, Treatment and Health Implications*, Sacchetti L.M., Mangiardi P. [Eds.], pp. 129–182, Nova Science Publishers.
18. Janiesch C., Zschech P., Heinrich K. (2021), Machine learning and deep learning, *Electronic Markets*, **31**(3): 685–695, https://doi.org/10.1007/s12525-021-00475-2.
19. Karunajeewa A.S., Abeyratne U.R., Hukins C. (2011), Multi-feature snore sound analysis in obstructive sleep apnea–hypopnea syndrome, *Physiological Measurement*, **32**(1): 83, https://doi.org/10.1088/0967-3334/32/1/006.
20. Korompili G. *et al.* (2021), PSG-Audio, a scored polysomnography dataset with simultaneous audio recordings for sleep apnea studies, *Scientific Data*, **8**(1): 197, https://doi.org/10.1038/s41597-021-00977-w.
21. Kursa M.B., Rudnicki W.R. (2010), Feature selection with the Boruta package, *Journal of Statistical Software*, **36**(11): 1–13, https://doi.org/10.18637/jss.v036.i11.
22. Lee B.S., Ellis D.P.W. (2012), Noise robust pitch tracking by subband autocorrelation classification, [in:] *13th Annual Conference of the International Speech Communication Association 2012*, https://doi.org/10.21437/interspeech.2012-221.

23. Li J. *et al.* (2017), Feature selection: A data perspective, *ACM Computing Surveys*, **50**(6): 94, https://doi.org/10.1145/3136625.
24. Lugaresi E., Cirignotta F., Montagna P. (1988), Pathogenic aspects of snoring and obstructive apnea syndrome, *Schweizerische Medizinische Wochenschrift*, **118**(38).
25. Markandeya M.N., Abeyratne U.R., Hukins C. (2018), Characterisation of upper airway obstructions using wide-band snoring sounds, *Biomedical Signal Processing and Control*, **46**: 201–211, https://doi.org/10.1016/j.bspc.2018.07.013.
26. Minaritzoglou A., Vagiakis E. (2008), Polysomnography: Recent data on procedure and analysis, *Pneumon*, **21**(4).
27. Osman A.M., Carter S.G., Carberry J.C., Eckert D.J. (2018), Obstructive sleep apnea: Current perspectives, *Nature and Science of Sleep*, **10**: 21–34, https://doi.org/10.2147/NSS.S124657.
28. Perez-Padilla J.R., Slawinski E., Difrancesco L.M., Feige R.R., Remmers J.E., Whitelaw W.A. (1993), Characteristics of the snoring noise in patients with and without occlusive sleep apnea, *American Review of Respiratory Disease*, **147**(3), https://doi.org/10.1164/ajrccm/147.3.635.
29. Pevernagie D., Aarts R.M., De Meyer M. (2010), The acoustics of snoring, *Sleep Medicine Reviews*, **14**(2): 131–144, https://doi.org/10.1016/j.smrv.2009.06.002.
30. Qian K. *et al.* (2021), Can machine learning assist locating the excitation of snore sound? A review, *IEEE Journal of Biomedical and Health Informatics*, **25**(4): 1233–1246, https://doi.org/10.1109/JBHI.2020.3012666.
31. Qian K., Janott C., Zhang Z., Heiser C., Schuller B. (2016), Wavelet features for classification of vote snore sounds, [in:] *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 221–225, https://doi.org/10.1109/ICASSP.2016.7471669.
32. Redline S. *et al.* (2010), Obstructive sleep apnea–hypopnea and incident stroke, *American Journal of Respiratory and Critical Care Medicine*, **182**(2), https://doi.org/10.1164/rccm.200911-1746oc.
33. Seiffert C., Khoshgoftaar T.M., Van Hulse J., Napolitano A. (2010), RUSBoost: A hybrid approach to alleviating class imbalance, [in:] *IEEE Transactions on Systems, Man, and Cybernetics Part A: Systems and Humans*, **40**(1): 185–197, https://doi.org/10.1109/TSMCA.2009.2029559.
34. Shen F., Cheng S., Li Z., Yue K., Li W., Dai L. (2020), Detection of snore from OSAHS patients based on deep learning, *Journal of Healthcare Engineering*, https://doi.org/10.1155/2020/8864863.
35. Solà-Soler J., Jané R., Fiz J.A., Morera J. (2007), Automatic classification of subjects with and without Sleep Apnea through snoring analysis, [in:] *2007 29th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, https://doi.org/10.1109/IEMBS.2007.4353739.
36. Song Y., Sun X., Ding L., Peng J., Song L., Zhang X. (2023), AHI estimation of OSAHS patients based on snoring classification and fusion model, *American Journal of Otolaryngology*, **44**(5): 103964, https://doi.org/10.1016/j.amjoto.2023.103964.
37. Sowho M., Sgambati F., Guzman M., Schneider H., Schwartz A. (2020), Snoring: A source of noise pollution and sleep apnea predictor, *Sleep*, **43**(6), https://doi.org/10.1093/sleep/zsz305.
38. Sun X., Ding L., Song Y., Peng J., Song L., Zhang X. (2023), Automatic identifying OSAHS patients and simple snorers based on Gaussian mixture models, *Physiological Measurement*, **44**(4): 045003, https://doi.org/10.1088/1361-6579/accd43.
39. Sun X., Peng J., Zhang X., Song L. (2022), Effective feature selection based on Fisher Ratio for snoring recognition using different validation methods, *Applied Acoustics*, **185**: 108429, https://doi.org/10.1016/j.apacoust.2021.108429.
40. Ulualp S.O. (2010), Snoring and obstructive sleep apnea, *Medical Clinics of North America*, **94**(5): 1047–1055, https://doi.org/10.1016/j.mcna.2010.05.002.
41. Wang C., Peng J., Song L., Zhang X. (2017), Automatic snoring sounds detection from sleep sounds via multi-features analysis, *Australasian Physical and Engineering Sciences in Medicine*, **40**(1): 127–135, https://doi.org/10.1007/s13246-016-0507-1.
42. White D.P. (2005), Pathogenesis of obstructive and central sleep apnea, *American Journal of Respiratory and Critical Care Medicine*, **172**(11), https://doi.org/10.1164/rccm.200412-1631SO.
43. Wu Z., Wang X., Jiang B. (2020), Fault diagnosis for wind turbines based on ReliefF and eXtreme gradient boosting, *Applied Sciences*, **10**(9): 3258, https://doi.org/10.3390/app10093258.
44. Zhang S., Li X., Zong M., Zhu X., Cheng D. (2017), Learning k for kNN Classification, *ACM Transactions on Intelligent Systems and Technology*, **8**(3): 43, https://doi.org/10.1145/2990508.
45. Zheng F., Zhang G., Song Z. (2001), Comparison of different implementations of MFCC, *Journal of Computer Science and Technology*, **16**(6): 582–589, https://doi.org/10.1007/BF02943243.