Journal of Healthcare Engineering

Machine Learning for Physiological Data Analytics

Lead Guest Editor: Yatao Zhang
Guest Editors: Peng Li and Jingyu Shao

Machine Learning for Physiological Data Analytics

Journal of Healthcare Engineering

Machine Learning for Physiological Data Analytics

Lead Guest Editor: Yatao Zhang
Guest Editors: Peng Li and Jingyu Shao

Contents

Hindawi
Journal of Healthcare Engineering
Volume 2023, Article ID 4994668, 12 pages
https://doi.org/10.1155/2023/4994668

Research Article

Tidal Volume Level Estimation Using Respiratory Sounds

Lurui Wang and Zhongwei Jiang

Graduate School of Science and Engineering, Yamaguchi University, Yamaguchi, Japan

Correspondence should be addressed to Zhongwei Jiang; jiang@yamaguchi-u.ac.jp

Received 20 August 2022; Revised 19 October 2022; Accepted 24 November 2022; Published 16 February 2023

Academic Editor: Yatao Zhang

Respiratory sounds have been used as a noninvasive and convenient method to estimate respiratory flow and tidal volume. However, current methods need calibration, making them difficult to use in a home environment. A respiratory sound analysis method is proposed to estimate tidal volume levels during sleep qualitatively. Respiratory sounds are filtered and segmented into one-minute clips, all clips are clustered into three categories: normal breathing/snoring/uncertain with agglomerative hierarchical clustering (AHC). Formant parameters are extracted to classify snoring clips into simple snoring and obstructive snoring with the *K*-means algorithm. For simple snoring clips, the tidal volume level is calculated based on snoring last time. For obstructive snoring clips, the tidal volume level is calculated by the maximum breathing pause interval. The performance of the proposed method is evaluated on an open dataset, PSG-Audio, in which full-night polysomnography (PSG) and tracheal sound were recorded simultaneously. The calculated tidal volume levels are compared with the corresponding lowest nocturnal oxygen saturation (LoO_2) data. Experiments show that the proposed method calculates tidal volume levels with high accuracy and robustness.

1. Introduction

Sleep quality and sleep time are both important for human health. Sleep quality is the measurement of how restful and restorative the sleep process proceeds. Enough sleep hours do not necessarily guarantee to get the most restful type of sleep. More than 80 sleep disorders are known to affect sleep quality. Among all these factors that cause poor sleep quality, sleep-related breathing disorders (SRBD) is the second one of all sleep-related disorders (the first one is insomnia) [1]. SRBD is the condition of abnormal and difficult respiration during sleep, which has effects on the balance of oxygen and carbon dioxide in the blood. Tidal volume is one of the parameters for monitoring respiratory ventilation and pulmonary function. Tidal volume is the amount of air that moves in or out of the lungs with each respiratory cycle. The normal tidal volume is around 500 mL in an average healthy adult male and approximately 400 mL in a healthy female. The tidal volume during sleep can be measured by many methods, such as polysomnography (PSG) and inductance plethysmography [2]. However, these methods are expensive, require a specialized operation, and cause uncomfortable sleeping. Therefore, there is a need for a nonintrusive, easy-operating method that can be used in a home environment. The acoustic method is getting popular in respiration monitoring as it only involves acquiring and processing respiratory sound signals to estimate tidal volume. The development of smartphones and wearable devices also made it possible to monitor respiration and tidal volume during sleep. Monitoring respiratory quality using respiratory sound is becoming a hotspot in recent years.

Many researchers have focused on analyzing the correlation between respiratory sound and respiratory airflow due to its potential for assessing snoring risk and estimating tidal volume. Various models or algorithms are proposed to estimate respiratory flow through respiratory sounds. Gavriely and Cugell proposed that the breath-sound amplitude (BAS) and flow (F) generally follow a 1.75-power relationship [3]. Yap and Moussavi proposed a method to use average power and an exponential model to estimate respiratory flow through tracheal sound, which reached an estimation error of $5.8 \pm 3.0\%$ [4]. Reljin et al. used the

blanket fractal dimension (BFD) as the parameter for estimating the tidal volume from tracheal sounds recorded by an Android smartphone, the smallest normalized root-mean-squared error of 15.877% ± 9.246% was obtained with the BFD and exponential model [5]. Yadollahi and Moussavi extracted the average power, the logarithm of the variance, and the logarithm of the envelope of tracheal sound as a feature, they compared the ability of these features to fit the flow-sound relationship, suggesting that the logarithm of the variance is the best feature to describe the flow-sound relationship with a linear model [6]. Other studies indicated that the Shannon entropy and sound variance also have an exponential relationship with the respiratory flow [7, 8]. Most of these papers indicate that the flow rate and respiratory sound amplitude follow a power law. This relationship used to estimate the respiratory flow rate can be presented in the following equation:

$$\log F_{\text{est}} = C_1 \log(E) + C_2. \quad (1)$$

F_{est} is the estimated flow rate (L/min), E is the respiratory sound amplitude, and C_1 and C_2 are the coefficients. C_1 and C_2 are determined by the human upper airway structure and can be calculated via a few breaths with a known flow rate for each participant, this procedure is called calibration. Current methods require calibration to determine the model coefficients C_1 and C_2. Yadollahi and Moussavi found that the parameters of the flow-sound relationship during sleep and wakefulness are different [9]. Therefore, for monitoring the tidal volume during sleep, the model parameters should be calibrated with sleep respiratory sounds.

However, these methods mentioned above are only applied to normal respiration, and calibration is needed for each case. Furthermore, these methods had not worked well for respiration during snoring. During snoring, the sound amplitude is higher than normal breathing, on contrary, the respiratory airflow is lower than normal breathing. The main reason is that the upper airway is usually collapsed or obstructed, and is highly variable during snoring. Respiration monitoring during snoring is important as it greatly affects sleep quality. During snoring, the upper airway is partially or completely blocked, and the respiratory airflow is limited or vanishes. Snoring usually leads to intermittent hypoxemia (IH), hypercapnia, arousal, hypertension, and sleep fragmentation. In this paper, a qualitative tidal volume estimation by a respiratory sound signal is proposed. It only used respiratory sound for analysis and does not need calibration. Therefore, the respiratory sound data could be easily collected by recording equipment and could be used in a home environment.

The proposed method consists of 4 main steps. First, the respiratory sounds are preprocessed into clips. Second, all clips are clustered into the normal breathing/snoring/uncertain categories with agglomerative hierarchical clustering (AHC). Third, the snoring clips are classified into simple snoring and apneic snoring with the K-means algorithm based on formant parameters and time domain parameters. Finally, the maximum breathing pause interval (MBPI) is calculated for apneic snoring clips to set the tidal volume to a medium or low level. The last time is calculated for simple snoring to set the tidal volume to a high- or medium-level. All the predictions are compared with LoO_2 (lowest nocturnal oxygen saturation) to evaluate the performance. All steps are unsupervised and do not need any calibration. The flow of the proposed method is shown in Figure 1.

2. Materials and Methods

The tracheal sounds are extracted from the PSG-Audio dataset. The dataset comprises 212 polysomnograms along with synchronized tracheal sound. The dataset contains edf files comprising polysomnogram signals and rml files containing all annotations by the medical team [10]. The edf files contain 20 channels, the SpO_2 (blood oxygen saturation level, in channel 15) and tracheal sound (in channel 19) data are extracted from the edf files for analysis. The SpO_2 measures the amount of oxygen in the blood. The corresponding respiratory events (obstructive apnea/mixed apnea/hypopnea) are extracted from the rml files. The sampling frequency of SpO_2 and tracheal sound is 1 Hz and 48000 Hz, respectively. A five minutes data clip is shown in Figure 2.

2.1. Agglomerative Hierarchical Clustering

2.1.1. Processing. The first step of preprocessing is filtering and denoising. As the respiratory sound energy of healthy people is usually concentrated in the low-frequency range of [50, 2500] Hz, a 50–2500 Hz Butterworth bandpass filter is used to filter noise. The sampling rate of recording files is downsampled to 5000 Hz. The second step of preprocessing is segmentation. The duration of the clip length is settled by considering the micro and the macro aspect. One clip should be short enough to separate each breathing stage; therefore, the audio signal in one clip is stable. The length of the clip is better to be cut with 5 to 10 breath periods for analysis. The usual breath period during sleep is 3 to 6 seconds. The length of 30 seconds to 60 seconds is considerable. Furthermore, considering the time of apnea in a serious case, it usually takes more than 30 seconds. In this paper, the length of segmentation is set at 60 seconds.

2.1.2. Feature Extraction. According to research about the human hearing mechanism, the human ear has different hearing sensitivity to sound waves of different frequencies. The human ear has a higher resolution of low-frequency sounds than high-frequency sounds. The Mel scale is a mapping from the human auditory perceived frequency to the actual frequency of the sound. By converting the frequencies to the Mel scale, features can better match the human auditory perception [6]. The Mel scale describes the nonlinear characteristics of the human ear frequency, and its relationship with frequency can be approximated by the following equation.

$$\text{Mel}(\text{f}) = 2595 * \log_{10}\left(\frac{\text{f}}{700} + 1\right), \quad (2)$$

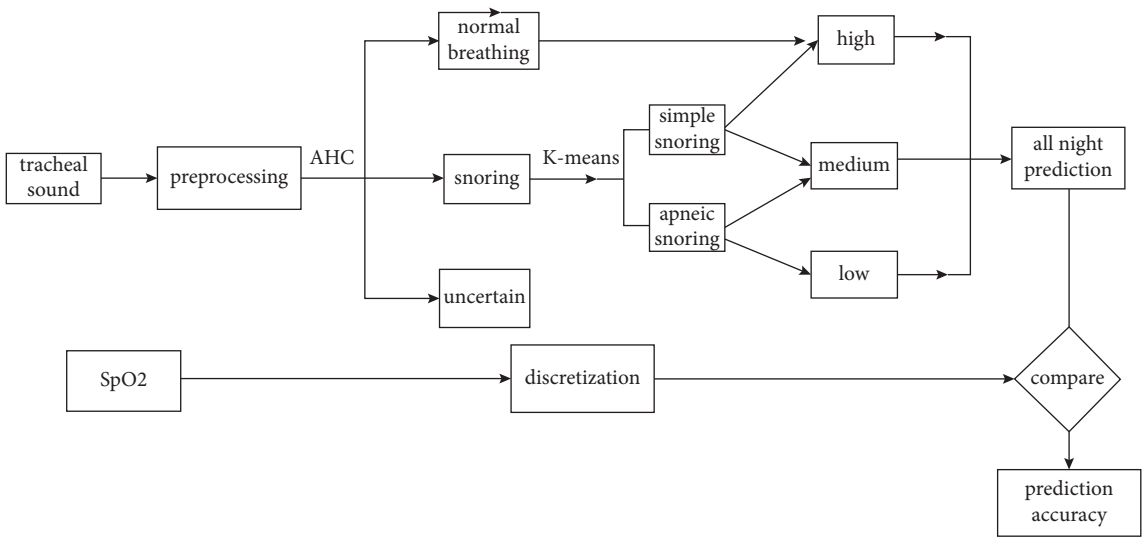

Figure 1: The flow of the proposed method.

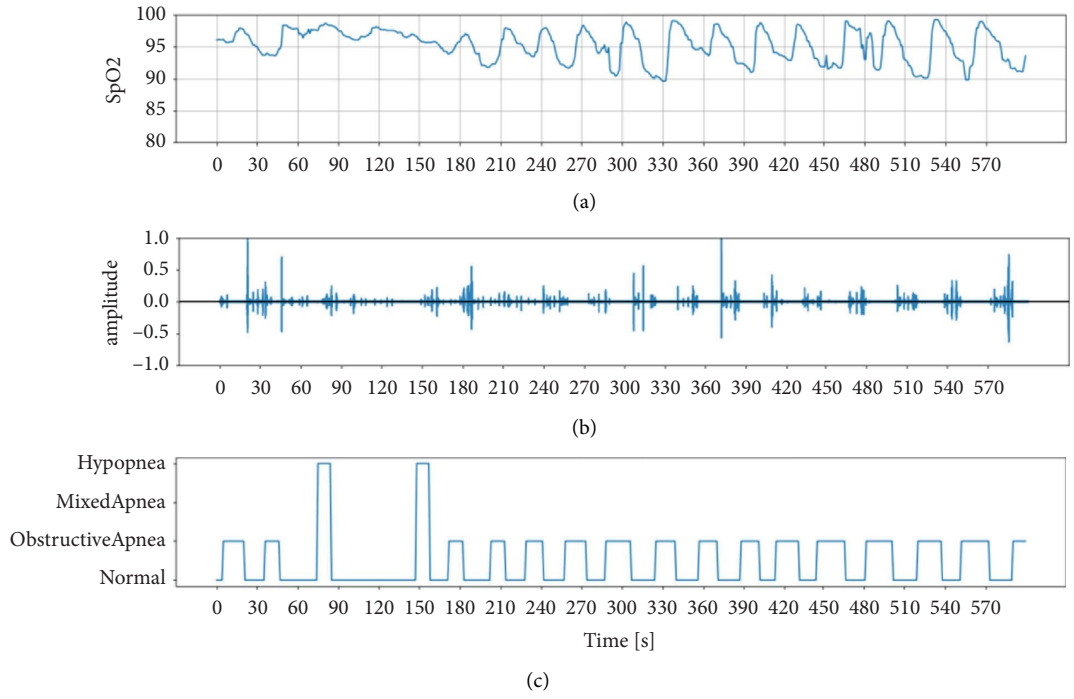

Figure 2: Data extracted from PSG-audio: (a) SpO_2; (b) Tracheal sound; (c) Respiratory events.

f is the frequency in Hertz.

The Mel-scale Frequency Cepstral Coefficients (MFCC) is a cepstral parameter extracted in the Mel-scale frequency domain [11]. MFCC were extracted from each clip file as the feature. The MFCC extraction algorithm usually includes windowing the signal into frames, and applying the fast Fourier transform (FFT) on frames to get the short-time Fourier transform spectrum (STFT). Then, the STFT spectrum was filtered with Mel-filter banks to get the Mel-spectrum, the Mel-spectrum was transformed into Mel-frequency cepstrum by taking the logarithm and then followed by applying the discrete cosine transform (DCT) to get MFCC coefficients. The MFCC feature vector describes the power spectral envelope of a single frame. Figure 3 shows the waveform, the Mel-spectrum, and the MFCC of a snoring sound clip with a duration of 60 seconds.

2.1.3. Similarity Calculation. The MFCC of each clip is a two-dimensional matrix, each column presents for a frame, and each row in the matrix corresponds to the Mel-frequency cepstral coefficients for the corresponding frame. As the respiratory sound signal is quasiperiodic, the MFCC matrix can be averaged by each row to get a one-

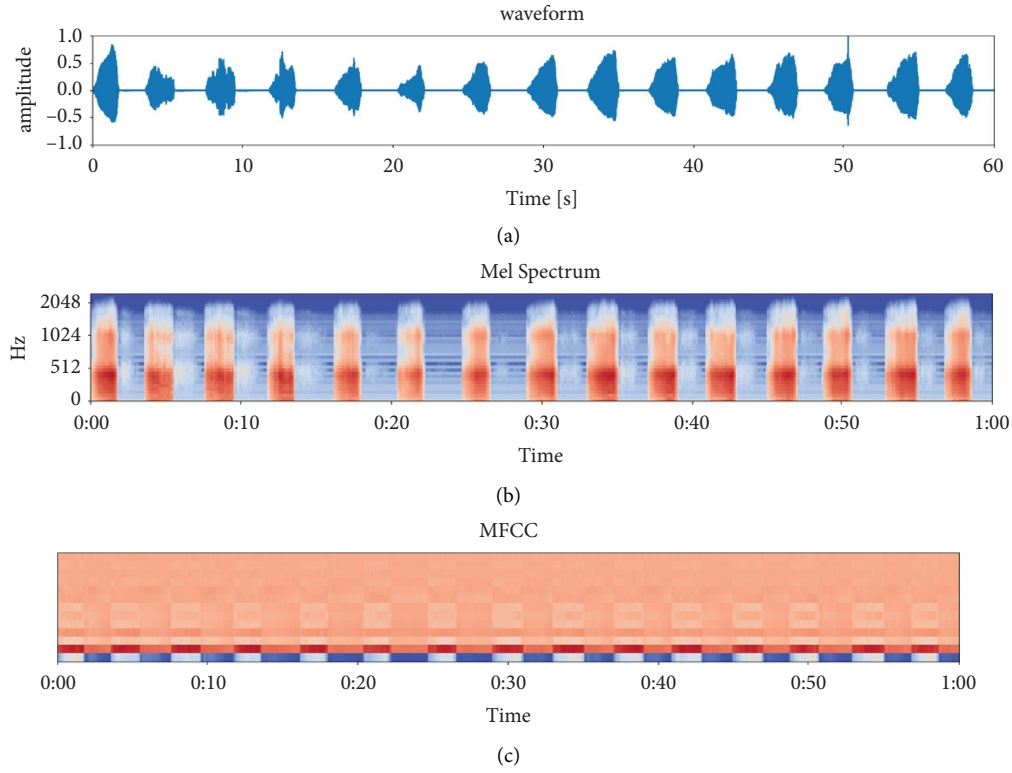

Figure 3: Spectrum of a snoring sound clip. (a) The waveform; (b) the Mel-spectrum; (c) MFCC.

dimension vector. As a vector can be presented as a point in a high-dimension space by its Cartesian coordinates, the MFCC matrix can be presented as points in a high-dimension space. The distance between the two clips can be measured by the distance between these two points. Based on our experiences, the Euclidean distance gave the most satisfying cluster result. The Euclidean distance between two points in Euclidean space is the length of a line segment between the two points. In general, if p and q are two points in n-dimensional Euclidean space, then the distance between them can be calculated by the following equation:

$$d(p,q) = \sqrt{(p_1 - q_1)^2 + (p_2 - q_1)^2 + \cdots + (p_n - q_n)^2}. \quad (3)$$

2.1.4. Agglomerative Hierarchical Clustering. Hierarchical clustering is a method of cluster analysis that can discover the structure of the dataset in an unsupervised way. It seeks iteratively merging nodes into bigger clusters (agglomerative), or divisive clustering nodes in the inverse (divisive) to build a hierarchy of all data. Agglomerative hierarchical clustering (AHC) is the most common type of hierarchical clustering [12, 13]. Pairs of clusters are successively merged until all clusters have been merged into one big cluster that contains all objects. At each iteration, two nodes or clusters, which have the minimum distance are merged. The result is a tree-based representation of all the objects, named a dendrogram. The number of clusters needs to be set before the algorithm begins.

A 120 minutes length file (2 hours) was selected from all the data and segmented into 60 seconds length clips for demonstration; therefore, 120 clips were used in the experiments. The STFT spectrum window length is 1000 ms with an overlap of 500 ms. The 40 Mel-scale filters were set in MFCC extraction. The distance matrix size is a symmetry matrix with a size of (120, 120). The dendrogram of the clustering result is shown in Figure 4. Based on the structure of the dendrogram, the dendrogram was divided into 3 clusters. Cluster 1, cluster 2, and cluster 3 are presented with cyan, magenta, and yellow, respectively. The dendrogram is shown in Figure 4, and the dendrogram is truncated for showing the main structure for the better visualization effect. The properties of each cluster are listed in Table 1.

One clip was chosen from each cluster as an example for analysis. The waveform and Mel-spectrum of examples present for each example are shown in Figure 5. Figure 5(a) is a spectrum of snoring. The snoring sounds are almost the same in amplitude and evenly spaced, the pitch of the snoring sound is in the low-frequency range and corresponds to a fundamental frequency with associated harmonics, and inspiratory is louder than expiratory. Figure 5(b) is a spectrum of normal respiration. It is characterized by a broader spectrum and is audible both during the inspiratory and expiratory phases. Figure 5(c) is a spectrum of uncertain types. The signal is very weak, and its spectrum has almost equal energy at frequencies below 2000 Hz. It is mixed with the weak breath, but the signal level is insufficient for analysis.

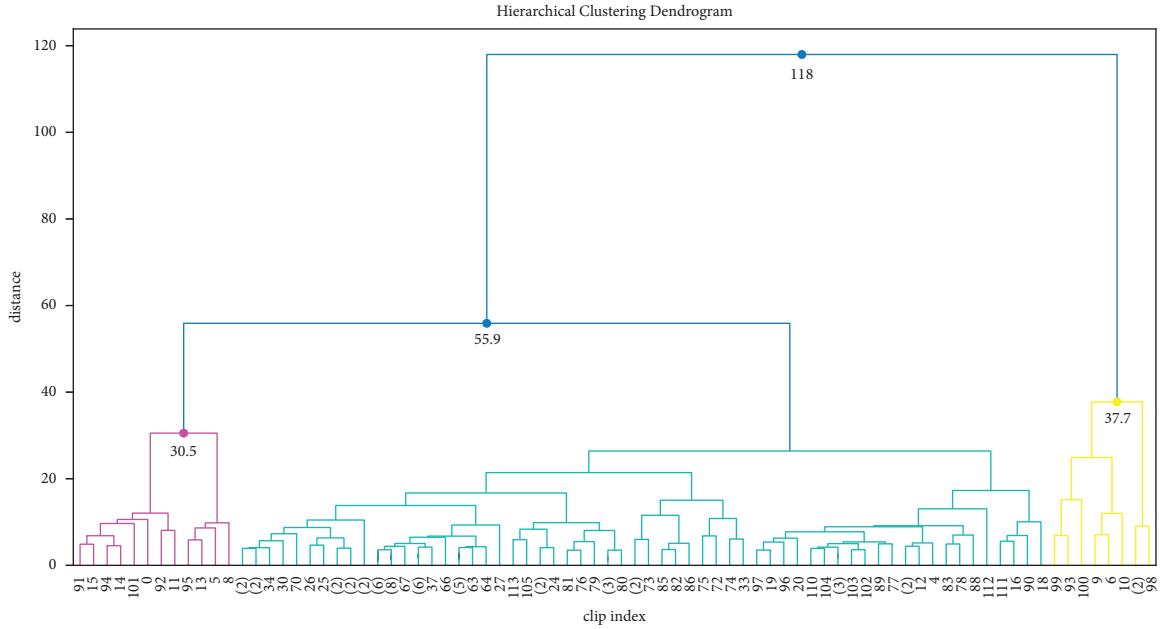

Figure 4: The dendrogram of cluster result.

Table 1: Characteristics of each cluster.

Cluster no	Property	Clip number
1	Snoring	94
2	Normal respiratory	12
3	Uncertain	14

2.2. Snoring Classification Based on K-Means Algorithm. Snoring occurs when the upper airways collapse, air moves around the floppy tissue near the back of the throat, and causes the tissue to vibrate. Simple snoring (also called benign snoring) occurs when there is a partial collapse of the soft tissues. As such, simple snoring is generally not considered a health threat. Apneic Snoring (also called obstructive sleep apnea-related snoring) is caused by partial or complete obstruction of the airway, and apneic snoring causes a partial or complete airflow stop, resulting in little or no oxygen going to the blood [14, 15]. For the apneic snoring, at the end of the obstruction, the closed upper airway is suddenly opened, and the pressures of the upper and lower airflows are suddenly balanced, causing the upper airway to repeat multiple openings and closings in a short period, producing a popping sound. The collapse degree and resistance of the upper airway may vary greatly from the beginning to the end of inspiration, thus, affecting the vibration of the upper airway tissue [16]. The snoring sounds in patients with obstructive sleep apnea and with simple snoring have different characteristics and effects on breath quality. It is essential to discriminate between these two different types of snoring for evaluating the influence on tidal volume.

Formant frequencies represent the resonance frequencies of the airways and change with the upper airway anatomy. A formant is the broad spectral maximum produced by an acoustic resonance of the human vocal tract [17]. Formants represent the direct source of pronunciation information, and the extraction and trajectory tracking of formants play an important role in speech recognition and speech synthesis. The formants $F1$–$F3$ are the three lowest resonant frequencies of the vocal tract. $F1$ is associated with the degree of pharyngeal constriction and the height of the tongue. $F2$ reflects the degree of the tongue's relative advancement position to its neutral position. $F3$ is related to the degree of lip rounding. Among $F1$–$F3$, $F1$ carries more information than others as it is associated with severity of apnea. Like speech pronunciation, snoring sounds are also produced depending on the shape and physical conditions of the upper airway, the formant of snoring can be extracted as a snoring feature [18]. Ng et al. proposed that apneic snoring has a high formant frequency than simple snoring in $F1$, and a threshold value of $F1 = 470$ Hz can be used to distinguish apneic snoring from simple snoring [19]. Sola Soler et al. suggested that the formant standard deviation of OSA snoring is higher than simple snoring [20].

These studies used the formant parameters to distinguish simple snoring from apneic snoring, and all emphasized the decisive role of $F1$. However, some cases may be misjudged by these methods. The reason is that the difference between the speech formant and the snoring formant is not considered. The most important formant analysis in speech processing is the formant tracks. The spacing between the word formant is not taken into consideration in speech processing. On the contrary, in applications such as speech recognition, the effect of spacing needs to be eliminated. The frequently used methods are dynamic time warping (DWT). By locally scaling the speech sequence, DWT eliminates the influence of speech rate and word spacing, so that the morphology of the two speech sequences is as consistent as possible, and the maximum possible similarity is obtained. But in snoring recognition, the interval between breathing is an important parameter as it is associated with airflow reduction time, and the interval of apneic snoring is usually

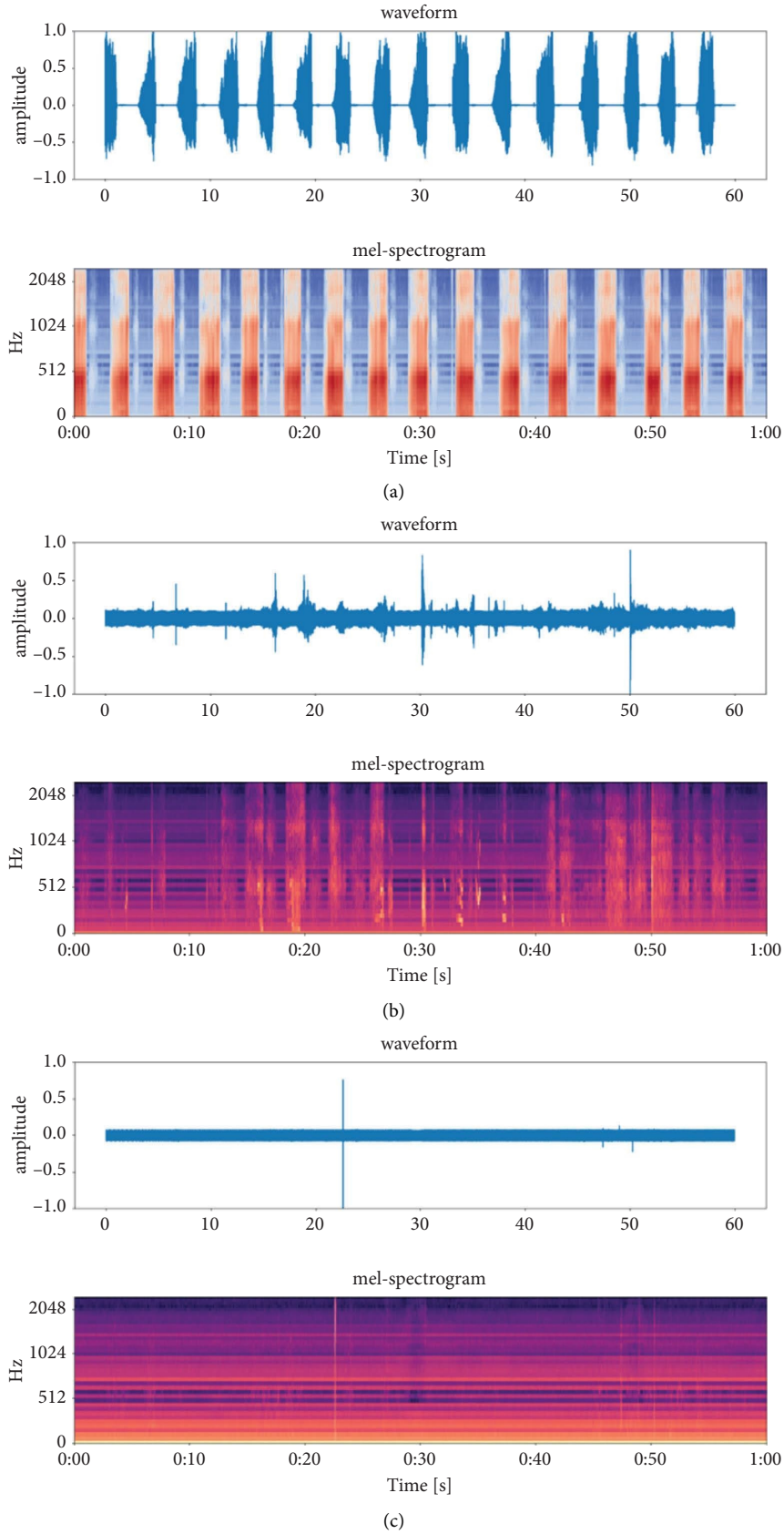

Figure 5: The example waveform and Mel-spectrum of each cluster: (a) Example of cluster 1; (b) Example of cluster 2; (c) Example of cluster 3.

larger and more irregular than simple snoring. To solve this problem, this paper extracted the standard deviation of the formant interval, together with the standard deviation of the formant frequencies as parameters, and used the *K*-means algorithm to discriminate between simple snoring and obstructive snoring by unsupervised clustering. *K*-means clustering is an unsupervised learning algorithm, it groups the unlabeled dataset into different clusters [21]. *K* defines the number of predefined clusters that need to be created in the process, here the *K* is set as 2.

The linear predictive analysis (LPC) method is one of the fast and more effective formant frequency estimation methods. The system function of the human vocal tract can be uniquely determined by a set of linear prediction coefficients, so the effect of vocal tract modulation can be estimated through LPC analysis. The formant of snoring can be obtained. The sound signals were windowed with a Hamming window of 20 ms with 50% overlap. In each window, a 14th-order LPC analysis is performed, and the LPC parameters were calculated via the Yule–Walker autoregressive method with the Levinson–Drubin recursive procedure. The standard deviation of *F*1 frequencies and the standard deviation of F1 interval are extracted to form a 2-dimensional feature vector. The snoring cluster result is shown in Figure 6. The apneic snoring and simple snoring are marked with red and cyan dots, respectively. After *K*-means clustering, the snoring cluster result of AHC is divided into 2 subclusters: cluster 0 and cluster 1. The property of all 4 clusters is shown in Table 2. The spectrum of the example clip chosen from cluster 0 and cluster 1 is shown in Figure 7, the formant is displayed with black dots on the spectrum.

2.3. Tidal Volume Level Estimation. For each cluster, different parameters are extracted and the corresponding tidal volume levels are determined based on these parameters. The tidal volume levels are divided into three grades: high, medium, and low. The tidal volume level is calculated for each cluster.

Cluster 2 contains the normal breathing clips, although there are fluctuations during normal respiration, the tidal volume levels of normal breathing can roughly be set as high.

Cluster 1 contains simple snoring. According to Hoffstein's research, simple snoring does not cause a sustained deterioration of MnO_2 (mean nocturnal oxygen saturation) but cause significantly the variability of LoO_2 (lowest nocturnal oxygen saturation) [22]. Based on this research, the tidal volume level during simple snoring beginning is similar to normal respiration, but after a certain duration, the fluctuation of nocturnal oxygen saturation increases and deteriorates ventilation quality at a moderate level. Although the accurate SpO_2 drop time is not clear, according to the research by Gruber, the interval to equilibration of oxygen saturation is within 4.5 minutes [23]. Therefore, the SpO_2 drop threshold is set at 4 minutes, meaning that when the normal breathing ends and simple snoring starts, after approximately 4 minutes, the SpO_2 drops to a medium level with high probability.

Cluster 0 contains apneic snoring. The breathing pause lasts longer than normal breathing during apnea. Based on the research by Ma et al., nocturnal hypoxemia severity is proportional to the pause time [24]. To evaluate the severity of hypoxemia, the maximum breathing pause interval (MBPI) is calculated as a parameter. According to the apnea definition, the threshold to distinguish the low/medium grade of apneic snoring is set to 10 seconds. The criterion for tidal volume level estimation is listed in Table 3.

3. Results and Discussion

The SpO_2 is a reading that shows the amount of oxygen available in human blood to deliver to the heart, brain, lungs, and other muscles and organs. The LoO_2 (lowest nocturnal oxygen saturation) is the lowest SpO_2 value during a certain time and has a high correlation with tidal volume. The LoO_2 is divided into 3 levels: large than 95% is considered a high level, less than 90% is considered low (hypoxemia), and between 95% and 90% is considered medium (mild) hypoxemia. The summarized results are shown in Figure 8. The first row is the clustering result, the *x*-axis represents the clip index, and each clip is 60 seconds in length. Each clip is classified into apneic snoring/simple snoring/breathing/uncertain types. The second row is the tidal volume level calculated by the proposed algorithm. The third row is the LoO_2, which is divided into high/medium/low levels, and the uncertain level corresponds to the uncertain clustering type. The fourth row is the SpO_2 level that is used to calculate the third row.

Six clips were selected as representatives, which are shown in Figure 9. Figure 9(a) is a normal respiration state at the 13th minute, the corresponding SpO_2 is stable and LoO_2 is above 95%. Figure 9(b) is apneic snoring with MBPI ≤ 10 at the 19th minute, the SpO_2 fluctuates, and LoO_2 is between 95% and 90%. Figure 9(c) is apneic snoring with MBPI > 10 at the 20th minute, the SpO_2 fluctuates dramatically, and LoO_2 is below 90%. Figure 9(d) is simple snoring at the 16th minute, the SpO_2 is at a high level as in (a). Figure 9(e) is simple snoring at the 43th minute, the SpO_2 drops slightly, and the LoO_2 drops to between 95% and 90%. Figure 9(f) is an uncertain case by which the signal is insufficient to calculate the SpO_2 level.

The accuracy is calculated by equation (4). Six patients with different apean-hypopnea index (AHI) were selected to test the effectiveness and robustness of the proposed method. AHI is defined as the number of apnea or hypopnea per hour during sleep. It is used as a parameter for the evaluation of the OSA severity. AHI less than 15 is considered mild apnea. AHI between 15 and 30 denotes moderate apnea, while a greater than 30 is considered severe. The characteristic of selected data and algorithm performance are shown in Table 4. The algorithm accuracy is 88.3% in the group with mild apnea. As for the moderate apnea group, the algorithm accuracy slightly drops to 85.8%. In the severe apnea group where the sound signal contains ambient noise, the algorithm accuracy is still above 83%.

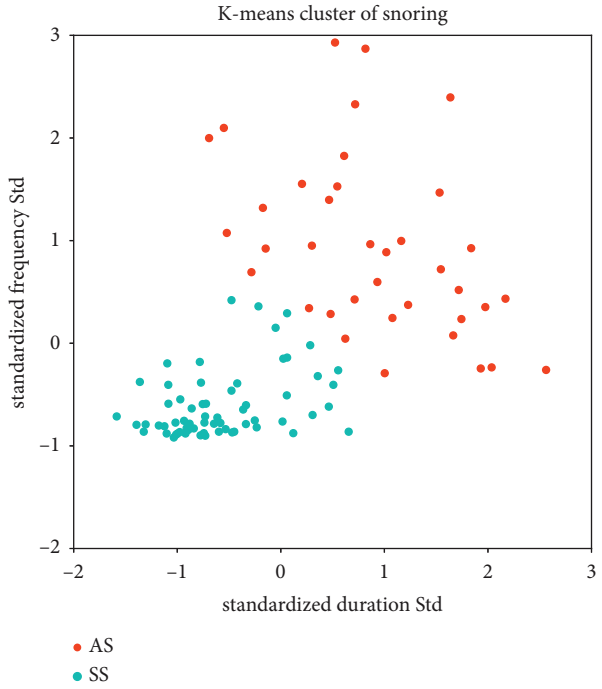

Figure 6: The apneic snoring/simple snoring cluster result.

Table 2: Characteristics of after SS/AS classification.

Cluster no	Property	Clip number
0	Apneic snoring (AS)	29
1	Simple snoring (SS)	65
2	Normal respiration	12
3	Uncertain	14

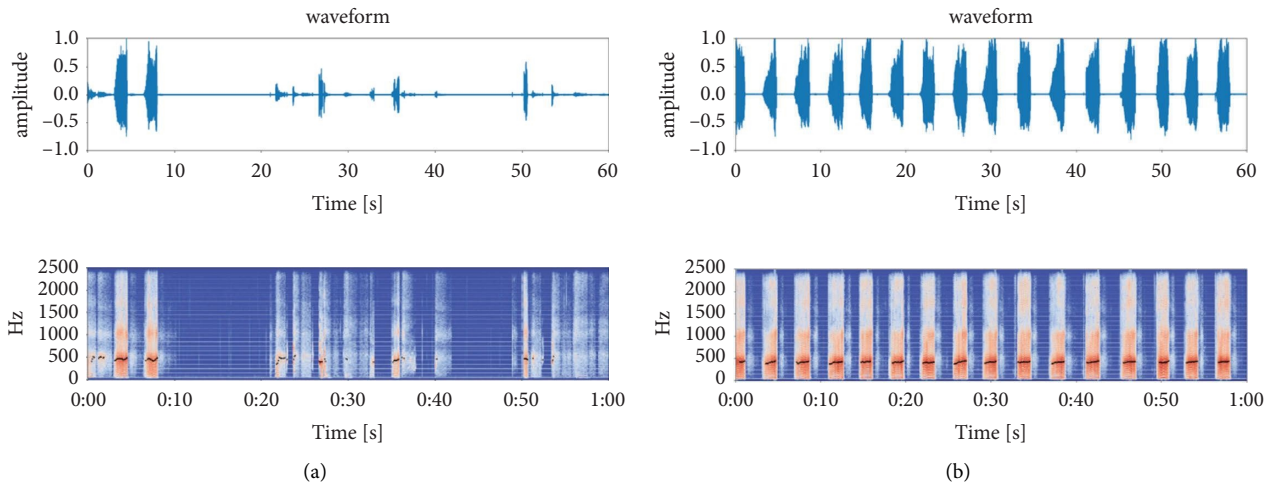

Figure 7: Example spectrum and formant of; (a) cluster 0 and (b) cluster 1.

Table 3: The criterion for tidal volume calculation.

Cluster no	Property	Criterion	Breathing quality
0	Apneic snoring	MBPI ≤ 10 second	Medium
		MBPI > 10 second	Low
1	Simple snoring	Last time < 4 minutes	High
		Last time ≥ 4 minutes	Medium
2	Normal respiration	All	High

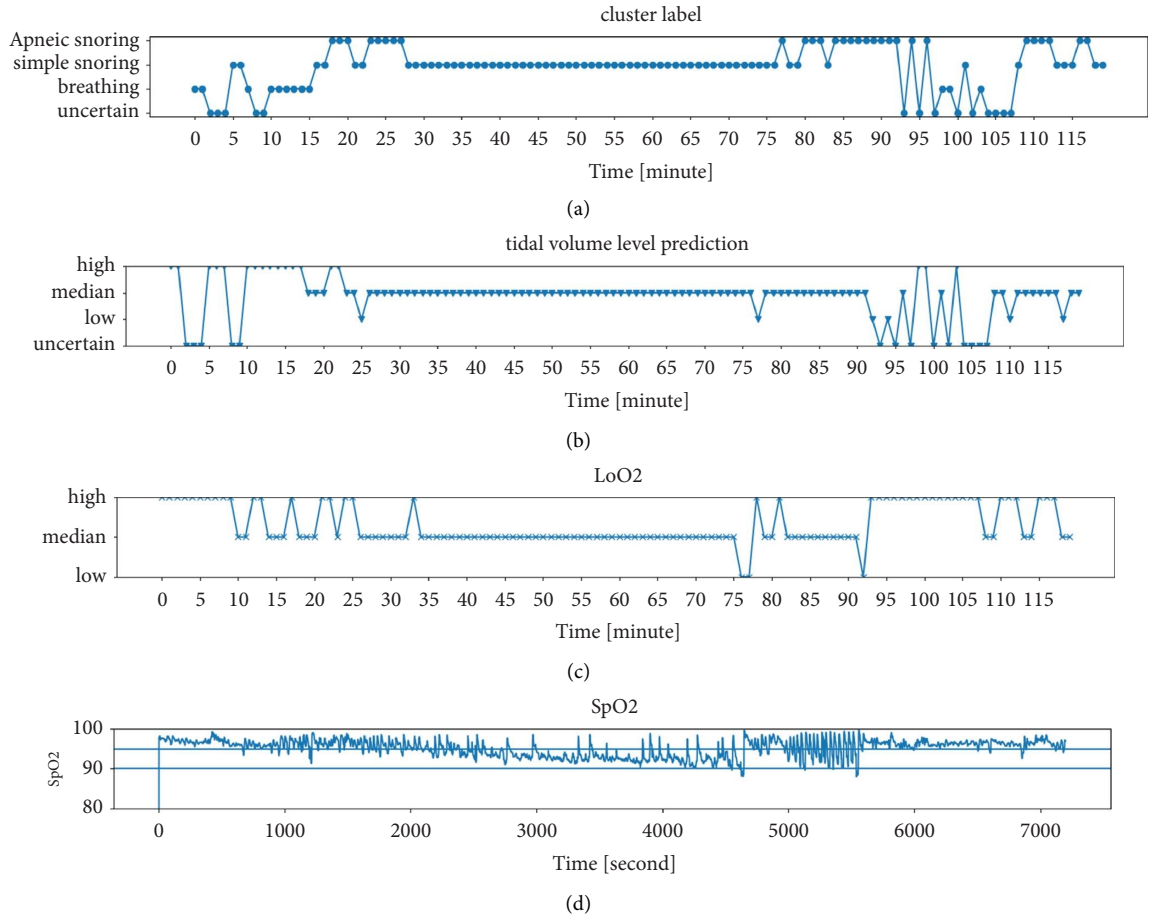

Figure 8: The prediction result: (a) cluster label, (b) tidal volume level prediction, (c) LoO_2, and (d) SpO_2.

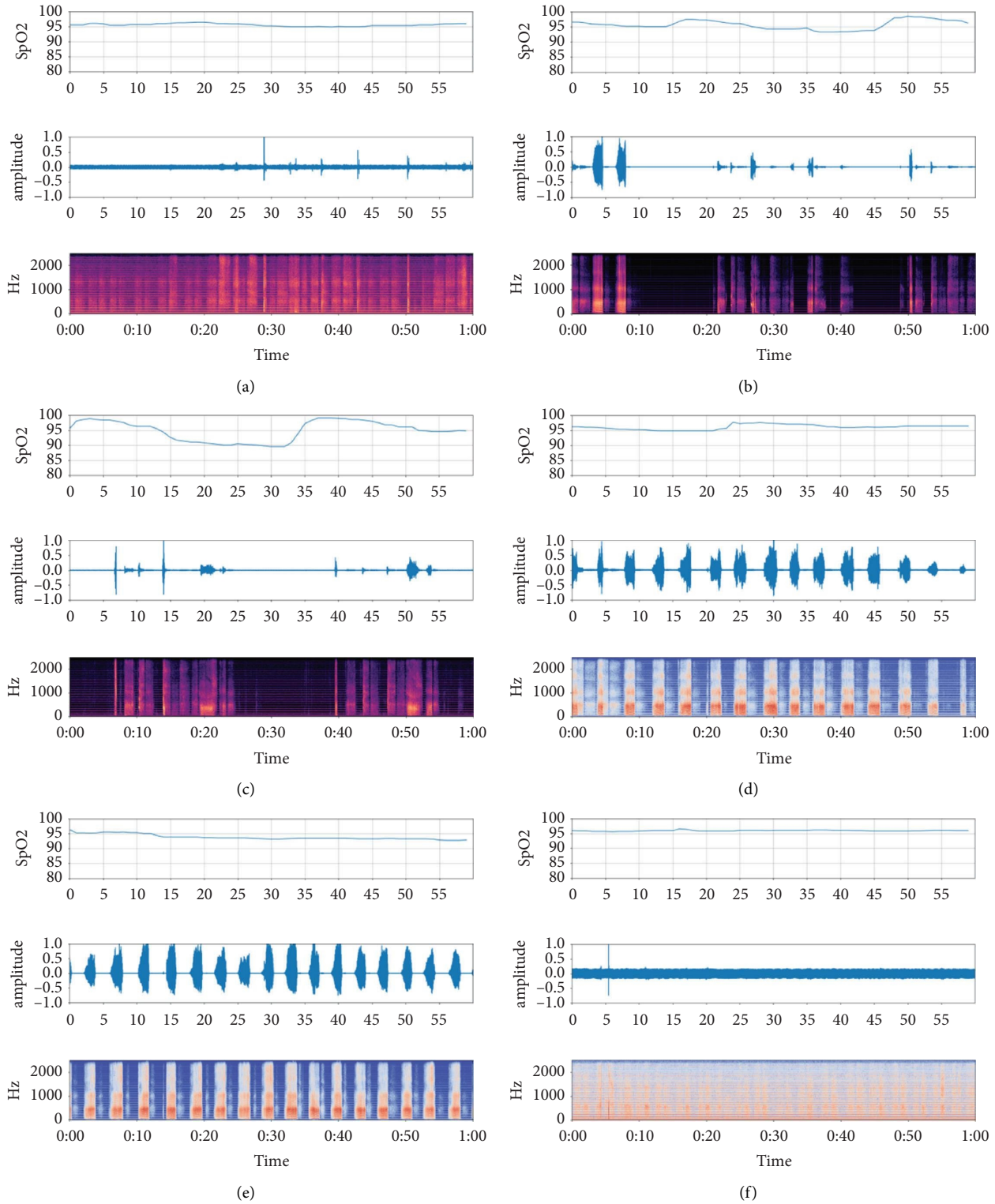

Figure 9: The representative state in prediction result. (a) Normal respiration; (b) apneic snoring with MBPI ≤ 10; (c) apneic snoring with MBPI > 10; (d) simple snoring with normal SpO_2; (e) simple snoring with SpO_2 drops slightly; (f) uncertain signal.

Table 4: Algorithm accuracy on data with different characteristics.

Apnea severity	Patient number	Data length (hour)	Data characteristic	Algorithm accuracy (%)
Mild	1	2	Mild apnea, no simple snoring	88.3
Moderate	3	6	Moderate apnea, little simple snoring	85.8
Severe	2	4	Severe apnea, little simple snoring, containing ambient noise	83.3

$$\text{Accuracy} = \frac{\text{correct prediction number}}{\text{total number} - \text{uncertain number}}. \quad (4)$$

4. Conclusion

In this study, a tidal volume level prediction method is proposed based on unsupervised clustering and snoring parameters. This method can provide a coarse-grained tidal volume level estimation that does not need any calibration. In addition, this method can be used for sleep breathing monitoring in a home environment. However, the accuracy of the method in this study is not very well because noise such as ambient noise will cause misjudgement, also breathing during sleep is affected by many other factors such as sleep position, pulmonary disease, and body movement, these factors cannot be captured by breathing sound. We are going to improve the performance by incorporating other factors in the future.

Data Availability

The data supporting the current study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This research was partially supported by the Ministry of Education of Japan, Science, Sport and Culture, Grant-in-Aid for Scientific Research, 2019–2022, 19K04257.

References

[1] E. M. Wickwire and N. A. Collop, "Insomnia and sleep-related breathing disorders," *Chest*, vol. 137, no. 6, pp. 1449–1463, 2010.

[2] P. P. Terragni, G. Rosboch, A. Tealdi et al., "Tidal hyperinflation during low tidal volume ventilation in acute respiratory distress syndrome," *American Journal of Respiratory and Critical Care Medicine*, vol. 175, no. 2, pp. 160–166, 2007.

[3] N. Gavriely and D. W. Cugell, "Airflow effects on amplitude and spectral content of normal breath sounds," *Journal of Applied Physiology*, vol. 80, no. 1, pp. 5–13, 1996.

[4] Y. L. Yap and Z. Moussavi, "Acoustic airflow estimation from tracheal sound power," *IEEE CCECE2002. Canadian Conference on Electrical and Computer Engineering*, vol. 2, pp. 1073–1076, 2002.

[5] N. Reljin, B. A. Reyes, and K. H. Chon, "Tidal volume estimation using the blanket fractal dimension of the tracheal sounds acquired by smartphone," *Sensors*, vol. 15, no. 5, pp. 9773–9790, 2015.

[6] A. Yadollahi and Z. Moussavi, "Comparison of flow-sound relationship for different features of tracheal sound," in *Proceedings of the 2008 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pp. 805–808, Vancouver, BC, Canada, Augest 2008.

[7] J. Fu, W. N. Teng, and W. Li, "Estimation of respiratory nasal pressure and flow rate signals using different respiratory sound features," *IRBM*, vol. 43, no. 6, pp. 694–704, 2021.

[8] G. Chen, I. de la Cruz, and E. Rodriguez-Villegas, "Automatic lung tidal volumes estimation from tracheal sounds," in *Proceedings of the 2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pp. 1497–1500, Chicago, IL, USA, November 2014.

[9] A. Yadollahi and Z. M. K. Moussavi, "The effect of anthropometric variations on acoustical flow estimation: proposing a novel approach for flow estimation without the need for individual calibration," *IEEE Transactions on Biomedical Engineering*, vol. 58, no. 6, pp. 1663–1670, 2011.

[10] G. Korompili, A. Amfilochiou, L. Kokkalas et al., "PSG-Audio, a scored polysomnography dataset with simultaneous audio recordings for sleep apnea studies," *Scientific Data*, vol. 8, no. 1, pp. 197–213, 2021.

[11] V. Tiwari, "MFCC and its applications in speaker recognition," *International Journal on Emerging Technologies*, vol. 1, no. 1, pp. 19–22, 2010.

[12] T. Pellegrini, J. Portêlo, and I. Trancoso, "Hierarchical clustering experiments for application to audio event detection," in *Proceedings of the 13th International Conference on Speech and Computer*, New Paltz, NY, USA, October 2009.

[13] I. Gronau and S. Moran, "Optimal implementations of UPGMA and other common clustering algorithms," *Information Processing Letters*, vol. 104, no. 6, pp. 205–210, 2007.

[14] F. Shen, S. Cheng, Z. Li, K. Yue, W. Li, and L. Dai, "Detection of snore from OSAHS patients based on deep learning," *Journal of Healthcare Engineering*, vol. 2020, Article ID 8864863, 10 pages, 2020.

[15] E. S. Arnardottir and T. Gislason, "Quantifying airflow limitation and snoring during sleep," *Sleep Medicine Clinics*, vol. 11, no. 4, pp. 421–434, 2016.

[16] K. Qian, Z. Xu, H. Xu, Y. Wu, and Z. Zhao, "Automatic detection, segmentation and classification of snore related signals from overnight audio recording," *IET Signal Processing*, vol. 9, no. 1, pp. 21–29, 2015.

[17] I. C. Yoo, H. Lim, and D. Yook, "Formant-based robust voice activity detection," *IEEE/ACM Transactions on audio, speech, and language Processing*, vol. 23, no. 12, pp. 2238–2245, 2015.

[18] C. Janott, M. Schmitt, Y. Zhang et al., "Snoring classified: the Munich-Passau snore sound corpus," *Computers in Biology and Medicine*, vol. 94, pp. 106–118, 2018.

[19] A. K. Ng, T. S. Koh, E. Baey, T. H. Lee, U. R. Abeyratne, and K. Puvanendran, "Could formant frequencies of snore signals be an alternative means for the diagnosis of obstructive sleep apnea," *Sleep Medicine*, vol. 9, no. 8, pp. 894–898, 2008.

[20] J. Sola-Soler, R. Jane, and J. A. Fiz, "Spectral envelope analysis in snoring signals from simple snorers and patients with obstructive sleep apnea," in *Proceedings of the 25th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, vol. 3, pp. 2527–2530, Cancun, Mexico, April 2003.

[21] A. Likas, N. Vlassis, and J. J Verbeek, "The global *K*-means clustering algorithm," *Pattern Recognition*, vol. 36, no. 2, pp. 451–461, 2003.

[22] V. Hoffstein, "Snoring and nocturnal oxygenation: is there a relationship?" *Chest*, vol. 108, no. 2, pp. 370–374, 1995.

[23] P. Gruber, T. Kwiatkowski, R. Silverrnan, E. Flaster, and C. Auerbach, "Time to equilibration of oxygen saturation using pulse oximetry," *Academic Emergency Medicine*, vol. 2, no. 9, pp. 810–815, 1995.

[24] C. Ma, Y. Zhang, J. Liu, and G. Sun, "A novel parameter is better than the AHI to assess nocturnal hypoxaemia and excessive daytime sleepiness in obstructive sleep apnoea," *Scientific Reports*, vol. 11, no. 1, pp. 4702–4708, 2021.

Hindawi
Journal of Healthcare Engineering
Volume 2023, Article ID 5287043, 11 pages
https://doi.org/10.1155/2023/5287043

Research Article

MCFN: A Multichannel Fusion Network for Sleep Apnea Syndrome Detection

Xingfeng Lv,[1,2] **Jinbao Li**,[3] **and Qianqian Ren**[2]

[1]*College of Electronic Engineering, Heilongjiang University, Harbin 150080, China*
[2]*Department of Computer Science and Technology, Heilongjiang University, Harbin 150080, China*
[3]*Shandong Artificial Intelligence Institute, Qilu University of Technology (Shandong Academy of Science), Jinan 250353, China*

Correspondence should be addressed to Jinbao Li; lijinb@sdas.org

Received 7 May 2022; Revised 24 June 2022; Accepted 24 November 2022; Published 23 January 2023

Academic Editor: Yatao Zhang

Sleep apnea syndrome (SAS) is the most common sleep disorder which affects human life and health. Many researchers use deep learning methods to automatically learn the features of physiological signals. However, these methods ignore the different effects of multichannel features from various physiological signals. To solve this problem, we propose a multichannel fusion network (MCFN), which learns the multilevel features through a convolution neural network on different respiratory signals and then reconstructs the relationship between feature channels with an attention mechanism. MCFN effectively fuses the multichannel features to improve the SAS detection performance. We conducted experiments on the Multi-Ethnic Study of Atherosclerosis (MESA) dataset, consisting of 2056 subjects. The experiment results show that our proposed network achieves an overall accuracy of 87.3%, which is better than other SAS detection methods and can better assist sleep experts in diagnosing sleep disorders.

1. Introduction

Sleep apnea syndrome (SAS) is a common sleep-breathing disorder characterized by repetitive events of complete or partial cessation of breathing during sleep [1]. SAS often occurs in men and women aged 30 to 60 years or older [2]. The main symptoms of SAS are daytime sleepiness, tiredness, inattention, and so on. Most SAS patients are undiagnosed and untreated which may lead to health problems such as heart and brain diseases [3–6].

SAS includes two important sleep events: obstructive sleep apnea (OSA) and hypopnea. According to an American Academy of Sleep Medicine (AASM) manual [7], OSA is scored when there is a 90% or more reduction in the prevent baseline of the airflow amplitude. However, there is a continued respiratory effort in the thoracic and abdominal belts. Hypopnea is scored when there is a 30% or more reduction in the preevent baseline of the airflow and 3% or more significant oxygen desaturation from the preevent baseline. Every OSA and hypopnea event lasts longer than 10 s. Normal sleep is scored when there is no OSA and hypopnea event or their duration time is less than 10 s.

Diagnosing SAS traditionally uses polysomnography (PSG), which is the gold standard. PSG can measure several signals, such as respiratory, electrocardiography (ECG), blood oxygen saturation, electroencephalography (EEG), and body movement signals. However, it is expensive and inconvenient because the patients need to attach a variety of sensors to their bodies. Moreover, it is time-consuming due to the manual analysis of signals. Therefore, it is necessary to propose alternative methods to automatic SAS detection using fewer physiological signals.

Various physiological signals have been used to detect sleep events [8–10]. Among these signals, respiratory signals can directly reflect the breathing situation during sleep [11]. The respiratory signal can be measured directly from the airflow sensor and thoracic and abdominal belts. Some methods have been used for SAS detection, such as threshold, support vector machine (SVM), logical regression (LR), and k nearest-neighbor (k-NN) [12–16]. These

methods extracted the time domain, frequency domain, and other nonlinear features from physiological signals. However, manual feature extraction is difficult to perform in noisy signals and requires domain knowledge.

Deep learning networks are alternatives as they can learn informative features without prior domain knowledge. Many researchers use long- and short-term memory (LSTM) and convolutional neural networks (CNNs) to classify physiological signals [17–32]. In particular, CNN is a popular class of deep learning networks that can automatically learn and find features from physiological signals. Haidar et al. [22] have demonstrated the efficacy of CNN models in classifying apnea or hypopnea events using airflow respiratory signals, with an accuracy of 77.6%. When a wavelet spectrogram of airflow respiratory signals input the network, the accuracy was 79.8%. If we use abdominal and thoracic respiratory signals simultaneously, the performance can reach 83.5% [23]. Urtnasan et al. [24] proposed a method for automated OSA detection from a single-lead ECG using CNN. Choi et al. [25] used CNN and a single-channel nasal pressure signal to detect the real-time apnea-hypopnea event. Nasal pressure signals were adaptively normalized and segmented by sliding a 10 s window at 1 s intervals. Many researchers use the LSTM model for SAS detection to learn the temporal features of sleep events. Van Steenkiste et al. [26] used LSTM to detect sleep apnea from raw respiratory signals, obtaining 77.2% accuracy. Elmoaqet et al. [27] used LSTM and bidirectional long-short-term memory (Bi-LSTM) to detect three sleep events and got an average accuracy of 83.6%. Yu et al. [32] proposed a method of sleep staging based on EEG signals combined with sleep apnea-hypopnea syndrome classification, which significantly reduced the rate of false positives that appear in the waking period. The data preprocessed by the sliding window were manipulated by LSTM and CNN to identify distinct various sleep events. Although these networks can automatically extract and learn deep-level features from physiological signals, there are still some shortcomings. First, they only focus on extracting deep features, ignoring the effect of shallow features, which can provide rich information for sleep events. Our initial conference paper solved this problem using a multilevel feature fusion network in [33]. Second, these networks did not consider the impact of channel features obtained by different respiratory signals. Some channel features can clearly distinguish sleep events, while others have little effect on SAS detection. We propose a multichannel fusion network (MCFN) to address this problem. MCFN effectively utilizes the shallow features of respiratory signals and fuses the multichannel features by an attention mechanism. We design a multichannel fusion block to calibrate the feature channel of various respiratory signals adaptively. Since the significance of each respiratory signal feature channel is different, this block can automatically obtain the importance of each feature channel, selectively enhance the useful channel feature, and restrain the useless ones. We evaluate our proposed network on a publicly available dataset with 2056 subjects. The MCFN can achieve an overall accuracy of 87.3%.

2. Material and Methods

MCFN can effectively fuse the features of different levels and channels. This network mainly includes signal preprocessing, multilevel feature concatenation, and multichannel attention fusion. We show the framework in Figure 1. First, we segment the various respiratory signals into a series of the 30 s length of epochs. The preprocessing block standardizes the respiratory signals, and each epoch is labeled as an event of OSA, hypopnea, and normal sleep according to the AASM guidelines. Second, the multilevel feature concatenation block obtains abundant features from shallow and deep layers through skip connections. Shallow features also contain some valuable identification information. Third, the multichannel fusion block uses an attention mechanism to learn different weights. The channel features that significantly affect SAS detection can obtain larger weights; otherwise, they get smaller weights. Finally, the feature vectors are input into two convolution layers and the max-pooling layer. The sleep classification is performed in the fully connected layer by sigmoid activation functions. In the following subsections, we detail the main block of this network.

2.1. Dataset. We conducted our experiments on a large dataset called the Multi-Ethnic Study of Atherosclerosis (MESA) [28, 29]. This dataset is retrieved from the National Sleep Research Resource (NSRR). NSRR is a new National Heart, Lung, and Blood Institute resource designed to provide extensive data resources to the sleep research community. MESA contains PSG recordings of 2056 subjects. The subjects, aged 45 to 84, come from different ethnic groups, including black, white, Hispanic, and Chinese men and women. Each PSG recording included various physiological signals such as EEG, respiration signals, and ECG. Our network only used three types of respiratory signals extracted from nasal thermal sensors and conductive belts around the thorax and abdomen. The sampling frequency of these signals is 32 Hz. Sleep experts labeled the start time and duration time of OSA and hypopnea events.

2.2. Data Preprocessing. In our network, three types of respiratory signals need to be preprocessed. First, we delete some subjects from the dataset which only contain normal sleep events. Second, due to different detection environments and equipment, the amplitude of each respiratory signal is very different. Therefore, the respiratory signal is individually standardized by subtracting the mean and dividing it by the standard deviation. Finally, according to the time of each sleep event, each 30 s epoch was labeled as OSA, hypopnea, or normal sleep event. If the epoch contains only obstructive sleep apnea or hypopnea lasting more than 10 seconds, it is labeled OSA or hypopnea. We excluded the epoch with obstructive sleep apnea and hypopnea events lasting more than 10 seconds. If an epoch contains obstructive sleep apnea or hypopnea events lasting less than 10 seconds, it is labeled as normal sleep.

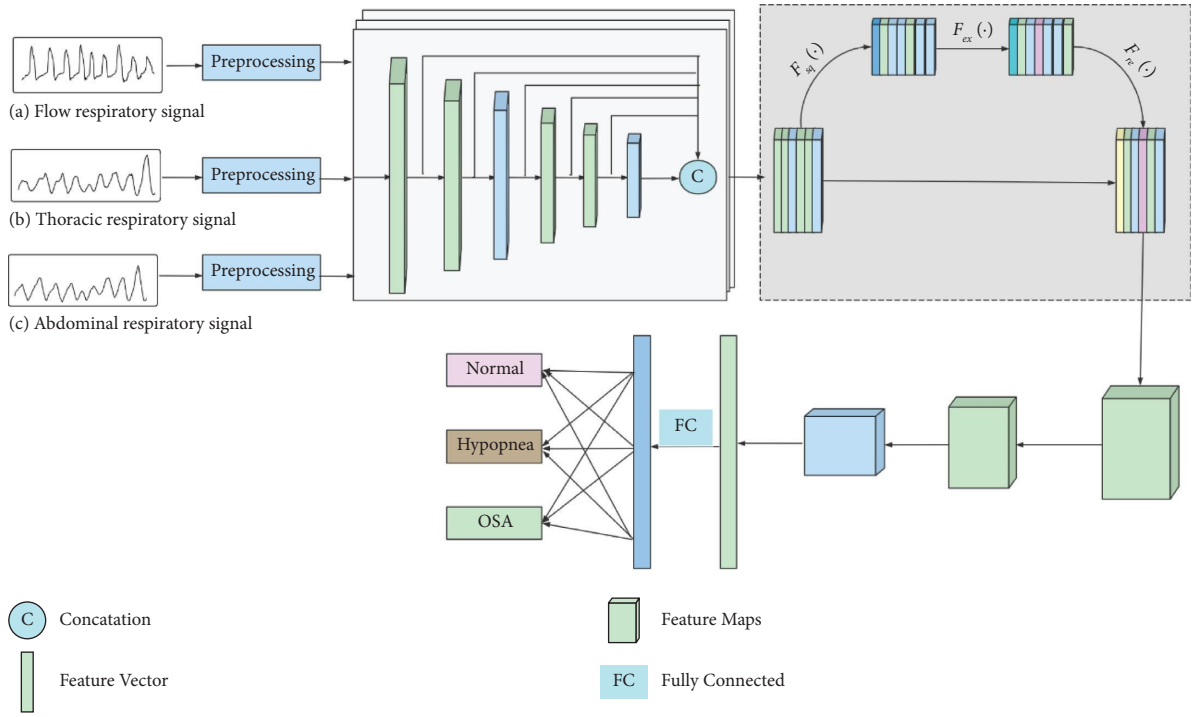

Figure 1: Overall framework of MCFN for SAS detection.

We also need to consider the balance classification of sleep events in preprocessing blocks. Typically, sleep events such as normal sleep are more than OSA or hypopnea. When learning a detection network with imbalanced classes, the result detects the most frequent sleep events. One way to address this issue is to employ balanced sampling. We randomly select the same number from the majority sleep event as the minority sleep event and then feed the network with batches of data that contain as many epochs from each sleep event.

2.3. Multilevel Feature Concatenation Block. A simple CNN architecture has been used for SAS detection [23, 33]. It was composed of convolution, pooling, and classification layers. The convolution layer extracts a feature map by applying a filter to the input respiratory signal. The pooling layer makes the feature more distinct and reduces the amount of data. The convolution layer can filter out some high-frequency information and make the signal smoother. In Figure 2, the partial feature map of the airflow respiratory signal after four convolution layers is shown. We find that with the increase of convolution layers, the receptive field of features becomes larger, and more high-frequency information is filtered. Although some networks use deep-level features to detect SAS, some high-frequency features are lost. Multilevel feature concatenation is realized through five skip connections to keep more high-frequency features in the network.

The multilevel feature concatenation block includes four convolution layers, two pooling layers, five skip connections, and one concatenation. We detail the parameters of different layers, which are summarized in Table 1. Each convolution layer has 32 filters with a rectified linear unit activation function, and each max-pooling layer has a pool size of (1, 2) with two strides. The convolutional kernel size is (1, 3) with three strides or (1, 2) with two strides. Following each convolution and pooling layer, the features of this level are obtained by average pooling to down-sampling. Then, these features are concatenated to generate multilevel feature maps. These features include shallow and deep features and provide more basic information. They can improve detection performance.

2.4. Multichannel Attention Fusion Block. Different respiratory signals such as airflow, thoracic, and abdominal have additional predictive power for SAS detection [27]. We fuse the multichannel features with an attention mechanism to fully use multilevel features from three types of respiratory signals. This block adaptively recalibrates channel-wise feature responses by explicitly modeling interdependencies between channels. It can learn to emphasize informative features and restrain less useful ones selectively.

As shown in Figure 1, we obtain the $C \times W \times H$ features through the multilevel feature concatenation block, where C is the number of channels, and each channel contains $W \times H$ features. Each respiratory signal has 192 channels, and each channel includes 1×7 features. The features of each respiratory signal are concatenated to obtain 576 channel features, which are the input of the multichannel attention fusion block. We recalibrate the multichannel features as follows.

First, the $F_{sq}(\ )$ operation compresses the features along the spatial dimension, turning each two-dimensional feature channel into an actual number. The global average pooling

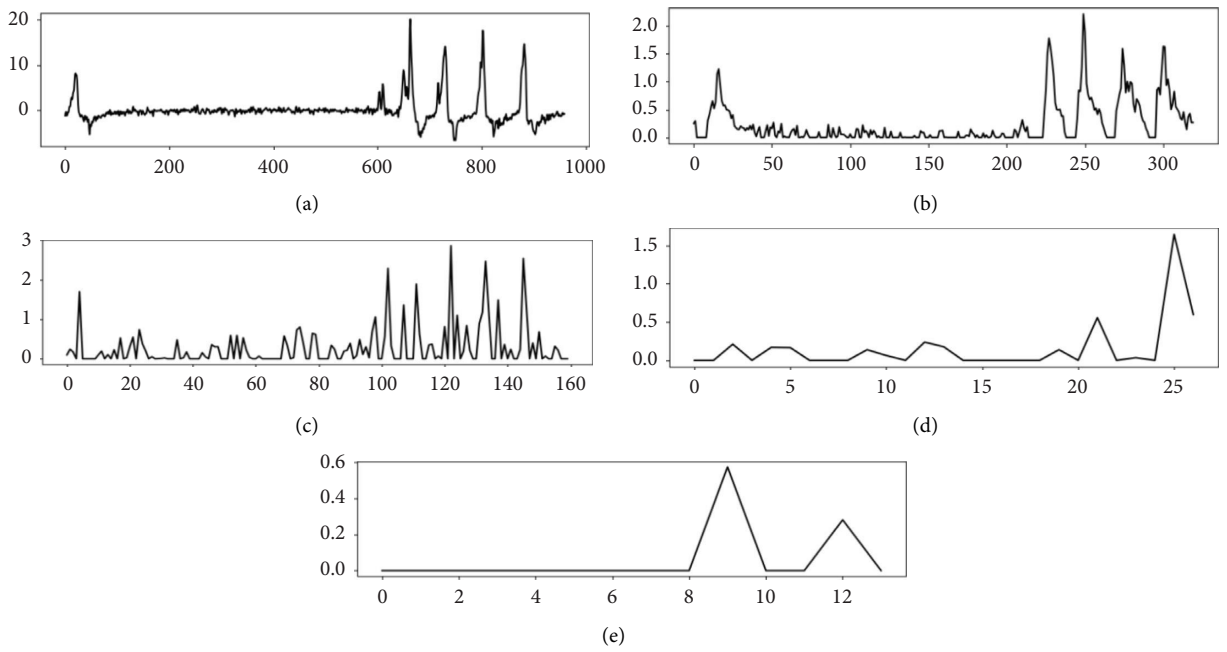

Figure 2: Feature maps of different convolution layers. (a) Airflow respiratory signal of OSA. (b) Feature map of the first convolution layer. (c) Feature map of the second convolution layer. (d) Feature map of the third convolution layer. (e) Feature map of the fourth convolution layer.

Table 1: Parameters in multilevel feature fusion block.

Layer	Size	Stride	#Filter	Activation	Dropout
Conv1	(1, 3)	3	32	ReLU ( )	0.2
Conv2	(1, 2)	2	32	ReLU ( )	0.2
Pooling	(1, 2)	2	32	—	—
Conv3	(1, 3)	3	32	ReLU ( )	0.2
Conv4	(1,2)	2	32	ReLU ( )	0.2
Pooling	(1, 2)	2	32	—	—
Concat	—	—	192	—	—

completes this operation to make the actual number have a global receptive field. The output dimension is the same as the number of input channels. F_{sq} (μ_c) is calculated as follows:

$$z_c = F_{\text{sq}}(\mu_c) = \frac{1}{H \times W} \sum_{i=1}^{H} \sum_{j=1}^{W} \mu_c(i, j), \tag{1}$$

where μ_c represents the feature map of the c-th channel feature map and i and j represent the row and column of the feature map, respectively.

Second, the F_{ex} ( ) operation is similar to the gate mechanism in the recurrent neural network (RNN). This operation can learn a nonlinear interaction between channels, and it can learn a nonmutually exclusive relationship. The operation is completed by two fully connected layers (FC). F_{ex} $(z,\ W)$ is calculated as follows:

$$s = F_{ex}(z, W) = \sigma(W_2 \delta(W_1, z)), \tag{2}$$

where δ refers to the ReLU function and the parameter W_1 multiplied by Z is the first FC layer. To limit model complexity and aid generalization, dimensions are reduced according to $c/16 \times c$. A dimensionality-increasing layer returns to the channel dimension of the transformation output.

$$X_c = F_{re}(u_c, s_c) = u_c \cdot s_c. \tag{3}$$

Finally, the F_{re} ( ) operation regards the output weight of the excitation as the importance of each feature channel. Then, the original feature is recalibrated on the channel dimension by weighting the previous feature by channel. F_{re} (μ_c, s_c) is calculated.where s_c indicates the importance of the feature channel, and μ_c represents the feature map of channel C.

After recalibration, there are 576 channel feature maps. The size of each feature map is 1×7. After two convolutions and one pooling operation, the convolution kernel sizes are (1, 3) and (1, 2), and the strides are 3 and 2, respectively. The max-pooling size is (1, 2), and the stride is 2. Finally, the flatten operation obtains the 576 features. Then, two fully connected layers and the sigmoid function output the probability of each sleep event. According to the probability value, this block outputs the sleep events.

2.5. Performance Evaluation. We evaluate and compare the performance of different methods using classification accuracy, sensitivity (recall), specificity, precision, and $F1$ score. They are defined as follows:

$$\begin{aligned}
\text{Accuracy} &= \frac{TP + TN}{TP + TN + FN + FP} \times 100\%, \\
\text{Specificity} &= \frac{TN}{TN + FP} \times 100\%, \\
\text{Precision} &= \frac{TP}{TP + FP} \times 100\%, \\
\text{Sensitivity (Recall)} &= \frac{TP}{TP + FN} \times 100\%, \\
F1 &= \frac{\text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}} \times 100\%,
\end{aligned} \tag{4}$$

where TP, FP, TN, and FN represent the number of true positive, false positive, true negative, and false negative epochs. The proportion of the correctly identified epochs is measured by sensitivity. Specificity reflects the detection effect of negative samples.

The confusion matrix also is used. Each row of the confusion matrix represents the epoch in actual labels, while each column represents the epoch in the predicted labels. We also standardized the confusion matrix by rows to obtain different probabilities. We use colors with different shades to represent the probability. The darker the color, the greater the probability, vise versa.

3. Experimental Results

This section presents the experimental setup and several experimental results designed to demonstrate the role of each block. First, we showed the classification results of the MCFN model, which proves that the model has better performance. Second, we confirmed the effect of different respiratory signals on different sleep event detections. They complement each other in the SAS detection. Third, we demonstrated the advantages of the multilevel feature concatenation block. Finally, we confirmed that the attention mechanism effectively fuses the multichannel features to improve performance.

3.1. Experimental Setup. The proposed network was trained and tested on the MESA dataset. After preprocessing, we selected 1801 subjects from 2506 subjects. They included the 54517 OSA events, 209910 hypopnea events, and 2019760 normal sleep events. The training and test set consisted of a balanced number for each sleep event to prevent the model from overfitting to the majority number of the class. We randomly selected 54517 sleep events from each sleep classification and mitigated the class imbalance issue. The experiment chose 80% of the sleep events as the training set and 20% as the testing set.

The training and testing are conducted based on the TensorFlow framework of Python 3.6. The experiments used the graphics card NVIDIA GTX 2080Ti GPU. The proposed network adopted the Adam optimization method and cross-entropy as the loss function. The initial learning rate is $1e-3$, and the learning rate is $1e-4$ after 40 iterations. The size of the mini-batch is 400 sleep events. The network had training of 100 epochs.

3.2. SAS Detection Performance of MCFN. The MCFN model detects sleep events using three respiratory signals of the chest, abdomen, and nasal airflow on the MESA dataset. The average accuracy is 87.3%, and the average $F1$ score is 87.3%. Table 2 presents the detection performance of the model. We found that the performance indexes of OSA sleep event detection are the highest, recall can reach 93.7%, the $F1$ score is 93.5%, and precision is 93.3%, indicating that the MCFN model can achieve good performance in detecting OSA events. There is a contradiction between the precision and recall of normal sleep and hypopnea events, which the $F1$ score can measure. The $F1$ scores of the two events are very similar, with a difference of only 0.8%, indicating that the performance of the MCFN model in detecting these two events is the same. From the confusion matrix, we found that there are some misclassifications between normal sleep and hypopnea events, mainly because sometimes the waveforms of the two events are very similar, but there are differences in amplitude. The MCFN model can achieve good performance in detecting OSA events. The main reason is that the waveform of the respiratory signal of OSA events is very different from that of other events.

3.3. The Effects of Three Respiratory Signals. We used sensitivity and specificity to measure the effect of different respiratory signals on various sleep events. The sensitivity measures the proportion of correctly identified positives, such as the percentage of OSA events correctly identified as having the event. The specificity measures the proportion of correctly identified negatives, such as the percentage of not OSA correctly identified as not having the event.

We show the sensitivity of airflow (Flow), thoracic respiratory signal (Thor.), and abdominal respiratory signal (Abdo.) in Figure 3. The sensitivity of abdominal respiration signal in detecting OSA and hypopnea sleep events is 81.39% and 73.05%, respectively. The sensitivity of the airflow respiration signals in detecting normal sleep events is 72.9%, which was higher than the other respiratory signals.

We show the specificity in Figure 4. The specificity of the airflow respiratory signal in detecting OSA was 93.72%, and the specificity of detecting hypopnea sleep events was 87.64%, which was 4.31% higher than that of the abdominal respiratory signal. The specificity of abdominal respiratory signals in detecting normal sleep events was 47.2%. These experimental results show that different respiratory signals play different roles in detecting various sleep events, so we can use three respiratory signals simultaneously for SAS detection.

To comprehensively evaluate the role of three respiratory signals in detecting SAS, we show the accuracy in Figure 5. We input single, two, and three respiratory signals into the

Table 2: Confusion matrix and the per-class result of the MCFN model.

		MCFN output			Per-class result (%)		
		Normal	Hypopnea	OSA	Precision	Recall	*F*1 score
Ground truth	Normal	4586	836	185	87.5	81.8	84.6
	Hypopnea	574	4842	191	81.4	86.4	83.8
	OSA	79	273	5255	**93.3**	**93.7**	**93.5**

Bold values indicate the highest value of each performance index.

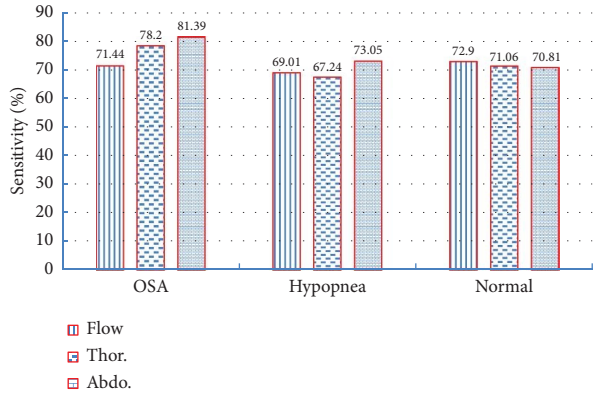

Figure 3: Sensitivity of different respiratory signals on sleep events.

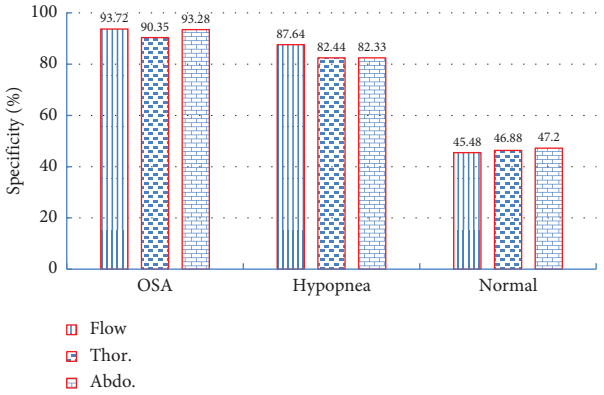

Figure 4: Specificity of different respiratory signals on sleep events.

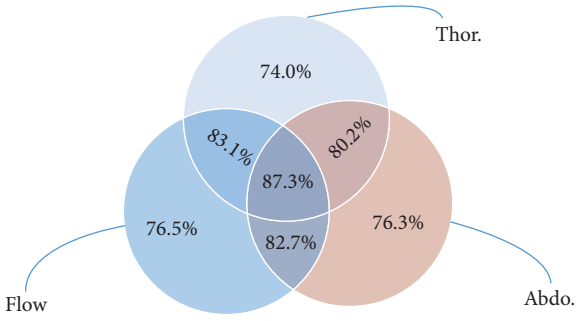

Figure 5: SAS detection accuracy of multiple respiratory signal combinations.

MCFN model, respectively. It can find that the SAS detection performance of single respiratory signals is the lowest. The accuracy of nasal airflow, abdominal, and thoracic respiratory signals was 76.5%, 76.3%, and 74.0%, respectively. When we combine the respiratory signals in pairs, the accuracy improves to varying degrees compared with that of single respiratory signals, such as the accuracy of combined flow and thoracic respiratory signals which can reach 83.1%, which is 6.6% higher than that of flow. The detection accuracy is the highest when the three respiratory signals are combined, reaching 87.3%. The detection accuracy improved by 9.1%.

We find that the detection performance of the combined respiratory signals is better than that of single respiratory signals. The three kinds of respiratory signals play different roles in detecting sleep events. The combination of multiple respiratory signals can complement each other and improve the SAS detection performance.

3.4. Multilevel Features Concatenation Block Improves Performance. In this experiment, we investigate the influence of the multilevel feature concatenation block on classification performances. First, to concatenate the features of different levels, it is necessary to down-sample the shallow features to get the same dimension. There are two methods for down-sampling: average pooling and max pooling. Through the experiment, we find that the two methods have little effect on the detection performance. We choose one way randomly, and here we choose average pooling to reduce the dimension. Then, by inputting different respiratory signals into the model with only deep-level features or multilevel features, the overall accuracy obtained is shown in Figure 6.

We find that whether it is single respiratory signals or combined respiratory signals, the detection accuracy using multilevel features is higher than that using only deep features. For airflow respiratory signals, the accuracy is only improved by 0.2%, indicating that the other level's features provide less identification information. For thoracic respiratory signals, the accuracy with only deep features was 71.1%, and the accuracy with multilevel features was 74.0%. It increased by 2.9%, indicating low-level features of thoracic respiratory signals which can provide rich identification information and improve the detection performance. For the combined respiratory signals, the accuracy can get improvement.

This result shows that the multilevel features of various respiratory signals have different effects on SAS detection. The complete learning features of thoracic and abdominal respiratory signals can improve detection accuracy. In contrast, the multilevel features of airflow respiratory signals have little impact on performance.

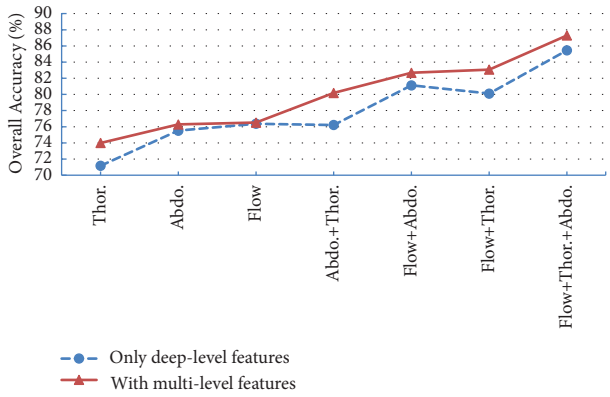

Figure 6: Influence of multilevel feature block on classification performance.

3.5. Multichannel Features Fusion Block Improves Performance. In this experiment, we investigated the influence of the relationship between different channel features on classification performances. We take two types of airflow and abdominal signals or three kinds of respiratory signals as an example. Whether or not multichannel feature fusion is used, Figure 7 shows the SAS detection confusion matrix.

Comparing the confusion matrices (a) and (b), we find that the correct classification probability of hypopnea events increased from 0.78 to 0.83, increased by 0.05. The correct classification probability of OSA events rose from 0.92 to 0.94, increasing by 0.02. The experimental results show that the respiratory signal combined with abdominal and airflow can extract rich features. After attention fusion, it can strengthen useful features and suppress useless features to improve performance. The correct classification probability of normal sleep events did not increase. Still, it decreased by 0.02, indicating that the extracted features by the two combined signals are very similar.

Comparing the confusion matrices (a) and (c), we find that the correct probability of each event classification in (c) is greater than or equal to that in (a). The experimental results confirm that the classification performance of three respiratory signals is better than that of two signals, which further verifies that various respiratory signals can provide richer information.

Comparing the confusion matrices (c) and (d), we find that the correct probability of event classification in (d) is greater than that in (c). The correct classification probability of hypopnea events increased from 0.82 to 0.86, an increase of 0.04. The correct probability of OSA event and normal event classification has increased by 0.01. The experimental results confirm that the attention mechanism improves the detection performance by fusing the multichannel features of the three respiratory signals.

The abovementioned experimental results confirm that the multichannel attention fusion block can improve the correct classification probability of hypopnea events and OSA events. The effect on normal sleep events is not very significant, mainly because the waveform of such events is relatively stable.

3.6. Learned Weight for Each Channel Feature. The attention mechanism can learn different weights for the channel features. The experiment results verify that the channel features of each respiratory signal have different effects on SAS detection. Figure 8 shows the multichannel feature weights of three respiratory signals. For the first channel of each respiratory signal, the channel weight of airflow respiratory is 0.18, the channel weight of thoracic respiratory is 0.50, and the channel weight of abdominal respiratory is 0.16. For the 64th channel of each respiratory signal, the channel weight of airflow respiratory is 0.50, the channel weight of thoracic respiratory is 0.50, and the channel weight of abdominal respiratory is 0.99. After multilevel feature concatenation of each respiratory signal, the model can obtain 192 channel features. The multichannel feature fusion block obtains 576 channel features. The attention mechanism learns the weight of each feature channel through training.

From Figure 8, we can find that the weights of each respiratory signal feature channel are different. For example, the weights of flow respiratory signals channel features are close to 1, and some are close to 0. These weights indicate that varying levels of features have different effects on sleep event detection. In addition, the weights of the feature channels 0, 32, 64, 96, 128, and 160 are marked with special graphics. The importance of channel features at the same level is also different.

Figure 9 shows the weight distribution of different respiratory signal channel features. When the weight is less than 0.25, the weight distribution of the three respiratory signals is very similar, indicating that the number of weak action feature channels is approximately equal. When the weight is in the range of 0.25~0.75, the number of feature diagrams of airflow respiratory signal is significant, indicating that the role of airflow respiratory signal is moderately important. When the weight is more powerful than 0.75, the number of the abdominal respiratory signals feature diagrams is large. This result indicates that these features contribute the most to SAS detection and contain the most identifying information. In addition, the Kolmogorov–Smirnov (KS) test further determines whether the channel weights of the two respiratory signals obey the same distribution. Since the P values are less than 0.05, they belong to different distributions. Therefore, each respiratory signal learning channel feature has different effects on SAS detection, which shows that the fusion of multiple respiratory signals is essential.

4. Discussion

Several methods have been applied to automated sleep event detection in previous studies. They can detect various sleep events, such as OSA, hypopnea, normal sleep, central sleep apnea (CSA), and mixed sleep apnea (MSA). The detection accuracy is compared with previous studies to evaluate the efficiency of MCFN.

Gutiérrez et al. [15] used a single airflow respiratory signal and the AdaBoost method to obtain 86.5% accuracy. They extracted features manually, detected normal sleep and

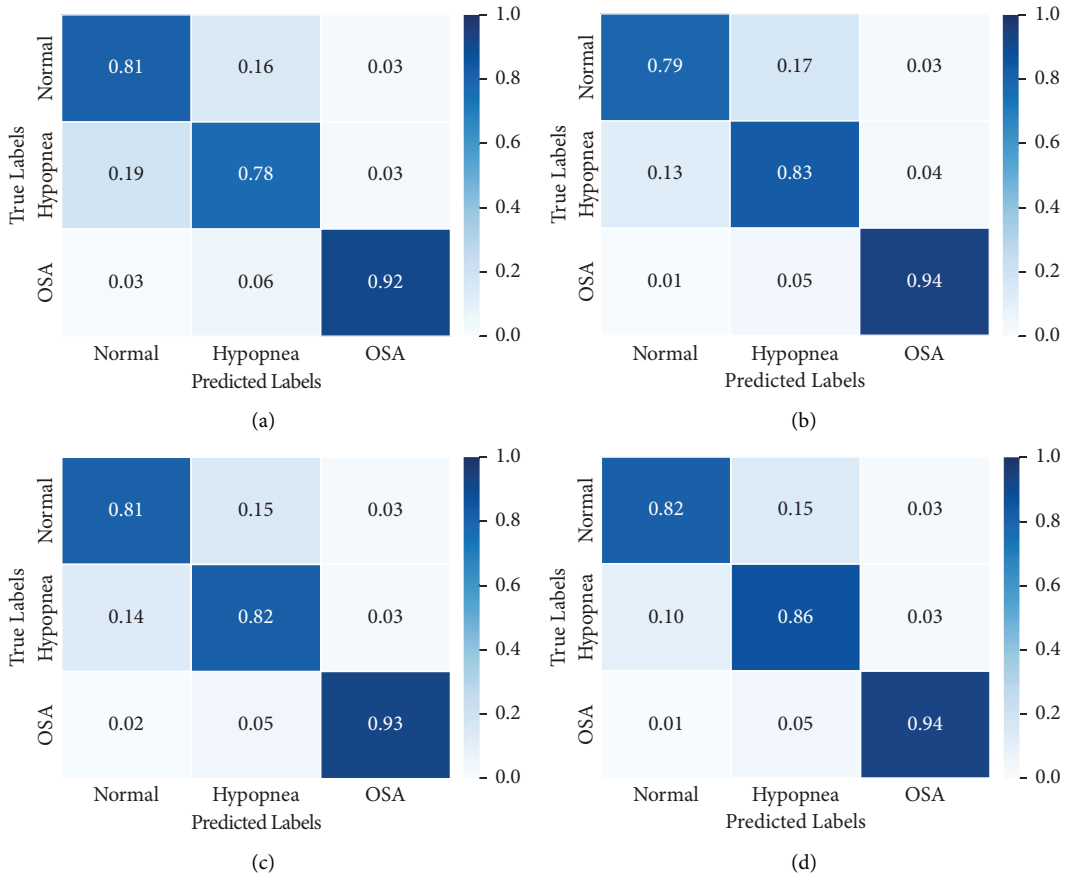

Figure 7: Influence of multichannel feature fusion (MCFF) block on multiple respiratory signals. (a) Two types of respiratory signals without MCFF. (b) Two types of respiratory signals with MCFF. (c) Three types of respiratory signals without MCFF. (d) Three types of respiratory signals with MCFF.

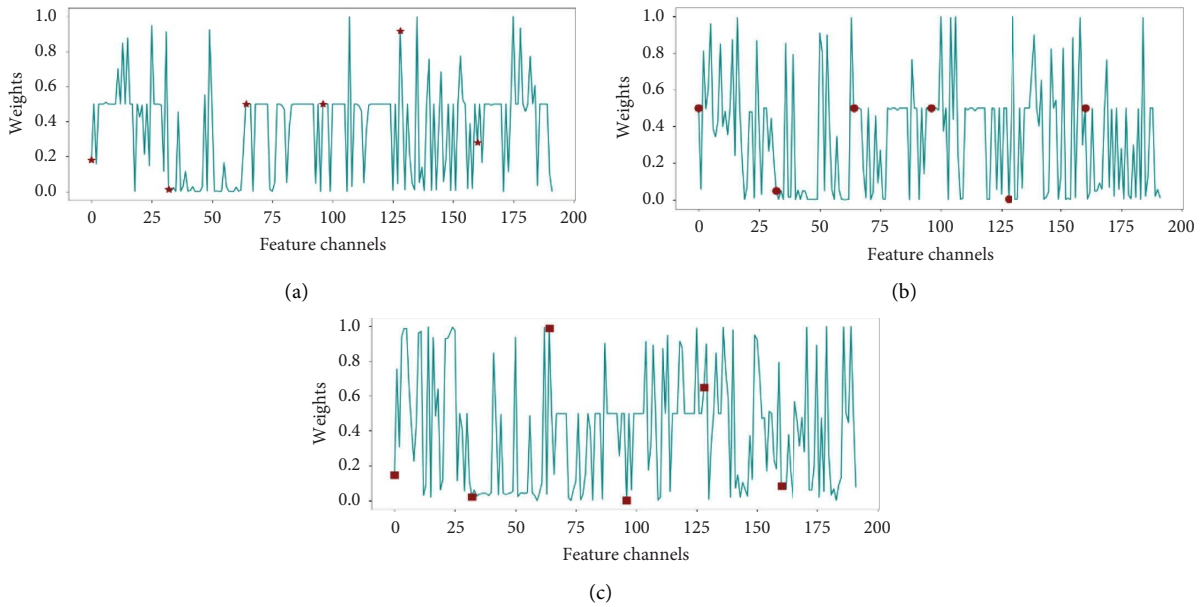

Figure 8: Weights of multichannel features of different respiratory signals. (a) Weight of multichannel features on airflow respiratory signal. (b) Weight of multichannel features on the thoracic respiratory signal. (c) Weight of multichannel features on the abdominal respiratory signal.

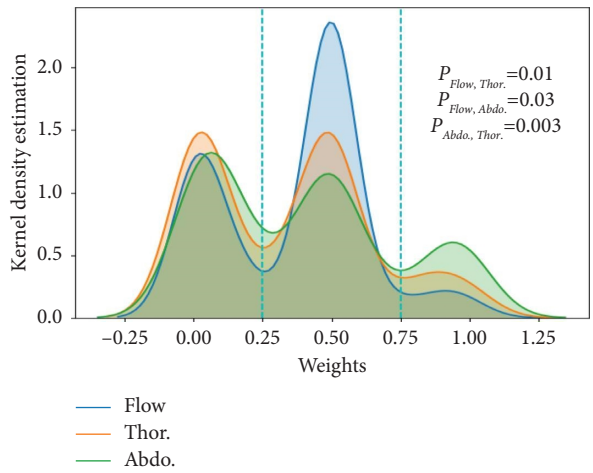

Figure 9: Distribution of weight values on different respiratory signal channel features.

Table 3: Performance comparison between MCFN and existing methods.

	Signals	Methods	Patients	Classify	Accuracy (%)
Gutiérrez et al. [15]	Flow	AdaBoost	317	Apnea/normal	86.5
Lin et al. [13]	Flow, Abdo. Thor.	SVM	34	OSA/CSA/hypopnea	81.8
Jiménez et al. [16]	Flow, SpO_2	AdaBoost	974	OSA/normal	81.3
Haidar et al. [23]	Flow, Abdo. Thor.	CNN	2056	OSA/hypopnea/normal	83.4
Van Steenkiste et al. [26]	Abdo. Thor. EDR	LSTM	2100	Apnea/normal	77.2
Elmoaqet et al. [27]	Flow, Abdo. NPRE	LSTM/Bi-LSTM	17	OSA/CSA/MSA	83.6
Barroso et al. [31]	Flow, ODI3	MLP	946	Apnea/normal	82.5
Yu et al. [32]	EEG, flow, Abdo.	LSTM_CNN	126	Normal/hypopnea/OSA/MSA	83.9
Ours	Flow, Abdo. Thor.	MCFN	2056	OSA/Hypopnea/normal	87.3

sleep apnea events, and did not classify them in detail. Lin et al. [13] explored the possibility of identifying sleep apnea events, including OSA and CSA, by solely analyzing one or both the thoracic and abdominal respiratory signals. They introduced an adaptive nonharmonic model to model the thoracic and abdominal movement signals. Then, an SVM method was applied to classify three categories of sleep events. When features from the thoracic and abdominal signals were combined, the overall classification accuracy became 81.8%. Jiménez et al. [16] evaluated the complementarity of airflow and oximetry (SpO_2) signals. They assessed the utility of a multiclass AdaBoost classifier to predict OSA severity in children.

Van Steenkiste et al. [26] used LSTM to detect normal sleep and sleep apnea on large data sets, and the accuracy was 77.2%. Although the temporal correlation of sleep events was considered, they ignored the relationship between different channel features. Elmoaqet et al. [27] developed the LSTM and Bi-LSTM framework to detect apnea events. They evaluated the framework over three respiration signals: airflow, nasal pressure (NPRE), and abdominal respiratory inductance plethysmography. They used PSG recording of 17 patients with obstructive, central, and mixed apnea events. The average accuracy was 83.6%.

Barroso et al. [31] conducted the 13 bispectral features from airflow. The oxygen desaturation index ≥3% (ODI3) was also obtained to evaluate its complementarity to the bispectral analysis. They used the fast correlation-based filter (FCBF) and a multilayer perceptron (MLP) to select the feature and recognize the pattern. The model reached 82.5% accuracy for the typical cut-offs of five events per hour. Yu et al. [32] proposed the SAS detection and classification method, which uses C4/A1 single-channel EEG signal, oronasal flow signal, and abdominal displacement signal. They utilized LSTM-CNN to identify four distinct types: normal sleep, hypopnea events, OSA, and CSA + MSA. The overall classification accuracy achieves 83.94%.

It is challenging to compare as they do not all use the same database and the number of the same sleep classification. To make a comparison on the same dataset, we have implemented the research of Haidar et al., who have carried out a lot of analysis on the MESA dataset. In the beginning, in [22], they got 77.6% accuracy with CNN by inputting airflow respiratory signal. Later, in [23], they obtained 83.5% accuracy by inputting three types of respiratory signals. All the previously mentioned research studies are summarized in Table 3. Considering the effect of shallow features on sleep classification and the relationship between different channel features in detecting sleep events, our experiment improved the accuracy by 3.9%. Our network could not only detect many types of sleep events but also improve accuracy.

5. Conclusion

We propose an MCFN model to detect OSA, hypopnea, and normal sleep. The model uses the multilevel feature concatenation block which can extract more rich information and give full play to the role of shallow features. Then, the

model utilizes an attention mechanism to effectively fuse the different level features of airflow, abdominal, and thoracic respiratory signals. The fusion block makes each channel feature of three respiratory signals have different weights, enhances the useful channel feature, and suppresses the useless channel feature. The experiments verified that multiple respiratory signals, multilevel features, multichannel fusion, and channel features affect SAS detection. MCFN model improves SAS detection performance by using the complementarity of various signals and the completeness of features. The detection accuracy is 87.3% on the MESA dataset, which is better than the other methods. In future research, we will try to study the effect of sleep apnea on sleep staging.

Data Availability

The MESA sleep dataset was supported by the National Heart, Lung, and Blood Institute (NHLBI) at the National Institutes of Health. It is available through NHLBI National Sleep Research Resource at https://www.sleepdata.org/datasets/mesa.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This paper was supported by the National Natural Science Foundation of China under grant no. 62172143), the National Key R&D Program of China under grant no. 2020YFB1710200, and Heilongjiang Province Natural Science Foundation Key Project of China under grant no. ZD2019F003.

References

[1] R. Heinzer, S. Vat, P. V. Marques et al., "Prevalence of sleep-disordered breathing in the general population:the hypnolaus study," *The Lancet Respiratory Medicine*, vol. 3, no. 4, pp. 310–318, 2015.

[2] C. V. Senaratna, J. L. Perret, C. J. Lodge et al., "Prevalence of obstructive sleep apnea in the general population: a systematic review," *Sleep Medicine Reviews*, vol. 34, no. 2, pp. 70–81, 2017.

[3] K. A. Maki, L. A. Burke, M. W. Calik et al., "Sleep fragmentation increases blood pressure and is associated with alterations in the gut microbiome and fecal metabolome in rats," *Physiological Genomics*, vol. 52, no. 7, pp. 280–292, 2020.

[4] S. Kalemci, I. Altun, F. Akin, and M. Biteker, "Obstructive sleep apnea and stroke risk in atrial fibrillation: is there a correlation or not?" *International Journal of Cardiology*, vol. 184, no. 1, p. 306, 2015.

[5] M. L. Jackson, M. Cavuoto, R. Schembri et al., "Severe obstructive sleep apnea is associated with higher brain amyloid burden: a preliminary pet imaging study," *Journal of Alzheimer's Disease*, vol. 78, no. 2, pp. 611–617, 2020.

[6] T. Harnod, Y. C. Wang, and C. H. Kao, "Association of migraine and sleep-related breathing disorder: a population-based cohort study," *Medicine*, vol. 94, no. 36, 2015.

[7] R. B. Berry, R. Budhiraja, D. J. Gottlieb et al., "Rules for scoring respiratory events in sleep: update of the 2007 aasm manual for the scoring of sleep and associated events. deliberations of the sleep apnea definitions task force of the american academy of sleep medicine," *Journal of Clinical Sleep Medicine*, vol. 8, no. 05, pp. 597–619, 2012.

[8] F. Mendonca, S. S. Mostafa, A. G. Ravelo, F. D. Morgado, and T. Penzel, "A review of obstructive sleep apnea detection approaches," *IEEE Journal of Biomedical and Health Informatics*, vol. 23, no. 2, pp. 825–837, 2019.

[9] N. Salari, A. F. Hosseinian, M. H. Mohammadi et al., "Detection of sleep apnea using machine learning algorithms based on ECG signals: a comprehensive systematic review," *Expert Systems with Applications*, vol. 187, Article ID 115950, 2022.

[10] G. Korompili, L. Kokkalas, S. A. Mitilineos, N. A. Tatlas, and S. M. Potirakis, "Detecting apnea/hypopnea events time location from sound recordings for patients with severe or moderate sleep apnea syndrome," *Applied Sciences*, vol. 11, no. 15, p. 6888, 2021.

[11] M. B. Uddin, C. M. Chow, and S. W. Su, "Classification methods to detect sleep apnea in adults based on respiratory and oximetry signals: a systematic review," *Physiological Measurement*, vol. 39, no. 3, 2018.

[12] M. Kagawa, H. Tojima, and T. Matsui, "Non-contact diagnostic system for sleep apnea-hypopnea syndrome based on amplitude and phase analysis of thoracic and abdominal Doppler radars," *Medical, & Biological Engineering & Computing*, vol. 54, no. 5, pp. 789–798, 2016.

[13] Y. Y. Lin, H. T. Wu, C. A. Hsu, P. C. Huang, Y. H. Huang, and Y. L. Lo, "Sleep apnea detection based on thoracic and abdominal movement signals of wearable piezoelectric bands," *IEEE Journal of Biomedical and Health Informatics*, vol. 21, no. 6, pp. 1533–1545, 2017.

[14] G. C. Gutiérrez-Tobal, D. Alvarez, A. Crespo, F. D. Campo, and R. Hornero, "Evaluation of machine-learning approaches to estimate sleep apnea severity from at-home oximetry recordings," *IEEE Journal of Biomedical and Health Informatics*, vol. 23, no. 2, pp. 882–892, 2019.

[15] G. C. T. Gutiérrez, D. Alvarez, F. Del Campo, and R. Hornero, "Utility of adaboost to detect sleep apnea-hypopnea syndromefrom single-channel airflow," *IEEE Transactions on Biomedical Engineering*, vol. 63, no. 3, pp. 636–646, 2016.

[16] J. G. Jiménez, G. C. Gutiérrez-Tobal, M. Garcia et al., "Assessment of airflow and oximetry signals to detect pediatric sleep apnea-hypopnea syndrome using adaboost," *Entropy*, vol. 22, no. 6, p. 670, 2020.

[17] Q. Yang, L. Zou, K. Wei, and G. Liu, "Obstructive sleep apnea detection from single-lead electrocardiogram signals using one dimensional squeeze-and-excitation residual group network," *Computers in Biology and Medicine*, vol. 140, Article ID 105124, 2022.

[18] J. Zhang, Z. Tang, J. Gao et al., "Automatic detection of obstructive sleep apnea events using a deep CNN-LSTM model," *Computational Intelligence and Neuroscience*, vol. 2021, Article ID 5594733, 10 pages, 2021.

[19] Y. Wang, Z. Xiao, S. Fang, W. Li, J. Wang, and X. Zhao, "Bi-directional long short-term memory for automatic detection of sleep apnea events based on single channel EEG signal," *Computers in Biology and Medicine*, vol. 142, Article ID 105211, 2022.

[20] K. Kido, T. Tamura, N. Ono et al., "A novel CNN-based framework for classification of signal quality and sleep

position from a capacitive ECG measurement," *Sensors*, vol. 19, no. 7, p. 1731, 2019.

[21] D. Dey, S. Chaudhuri, and S. Munshi, "Obstructive sleep apnoea detection using convolutional neural network based deep learning framework," *Biomedical Engineering Letters*, vol. 8, no. 1, pp. 95–100, 2018.

[22] R. Haidar, I. Koprinska, and B. Jeffries, "Sleep apnea event detection from nasal airflow using convolutional neural networks," in *Proceedings of the International Conference on Neural Information Processing*, Springer, Berlin, Germany, October 2017.

[23] R. Haidar, S. McCloskey, I. Koprinska, and B. Jeffries, "Convolutional neural networks on multiple respiratory channels to detect hypopnea and obstructive apnea events," in *Proceedings of the 2018 International Joint Conference on Neural Networks (IJCNN)*, pp. 1–7, Brazil, July 2018.

[24] E. Urtnasan, J. U. Park, E. Y. Joo, and K. J. Lee, "Automated detection of obstructive sleep apnea events from a single-lead electrocardiogram using a convolutional neural network," *Journal of Medical Systems*, vol. 42, no. 6, p. 104, 2018.

[25] S. H. Choi, H. Yoon, H. S. Kim et al., "Real-time apnea-hypopnea event detection during sleep by convolutional neural networks," *Computers in Biology and Medicine*, vol. 100, pp. 123–131, 2018.

[26] T. Van Steenkiste, W. Groenendaal, D. Deschrijver, and T. Dhaene, "Automated sleep apnea detection in raw respiratory signals using long short-term memory neural networks," *IEEE Journal of Biomedical and Health Informatics*, vol. 23, no. 6, pp. 2354–2364, 2019.

[27] H. Elmoaqet, M. Eid, M. Glos, M. Ryalat, and T. Penzel, "Deep recurrent neural networks for automatic detection of sleep apnea from single channel respiration signals," *Sensors*, vol. 20, no. 18, p. 5037, 2020.

[28] X. Chen, R. Wang, P. Zee et al., "Racial/ethnic differences in sleep disturbances: the multi-ethnic study of atherosclerosis (MESA)," *Sleep*, vol. 38, no. 6, pp. 877–888, 2015.

[29] G. Q. Zhang, L. Cui, R. Mueller et al., "The national sleep research resource: towards a sleep data commons," *Journal of the American Medical Informatics Association*, vol. 25, no. 10, pp. 1351–1358, 2018.

[30] T. V. Steenkiste, W. Groenendaal, J. Ruyssinck, P. Dreesen, and T. Dhaene, "Systematic comparison of respiratory signals for the automated detection of sleep apnea," in *Proceedings of the 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, Honolulu, HI, USA, July 2018.

[31] V. G. Barroso, G. C. T. Gutiérrez, L. K. Gozal et al., "Bispectral analysis of overnight airflow to improve the pediatric sleep apnea diagnosis," *Computers in Biology and Medicine*, vol. 129, Article ID 104167, 2021.

[32] H. Yu, D. Liu, J. Zhao et al., "A sleep apnea-hypopnea syndrome automatic detection and subtype classification method based on LSTM-CNN," *Biomedical Signal Processing and Control*, vol. 71, Article ID 103240, 2022.

[33] X. Lv and J. Li, "A multi-level features fusion network for detecting obstructive sleep apnea hypopnea syndrome," in *Proceedings of the 2020 International Conference on Algorithms and Architectures for Parallel Processing(ICA3PP)*, pp. 509–519, Berlin, Germany, September 2020.

Hindawi
Journal of Healthcare Engineering
Volume 2023, Article ID 3269144, 11 pages
https://doi.org/10.1155/2023/3269144

Research Article

Atrial Fibrillation Detection with Low Signal-to-Noise Ratio Data Using Artificial Features and Abstract Features

Zhe Bao,[1] **Dong Li**,[1] **Shoufen Jiang**,[2] **Liting Zhang**,[3] **and Yatao Zhang**[1]

[1]*School of Mechanical, Electrical and Information Engineering, Shandong University, Weihai 264209, China*
[2]*School of Business, Shandong University, Weihai 264209, China*
[3]*Department of Electrocardiographic, Shandong Provincial Hospital Affiliated to Shandong University, Jinan 250021, China*

Correspondence should be addressed to Yatao Zhang; zytboy@sdu.edu.cn

Received 25 April 2022; Revised 30 August 2022; Accepted 24 November 2022; Published 21 January 2023

Academic Editor: Jose Joaquin Rieta

Detecting atrial fibrillation (AF) of short single-lead electrocardiogram (ECG) with low signal-to-noise ratio (SNR) is a key of the wearable heart monitoring system. This study proposed an AF detection method based on feature fusion to identify AF rhythm (A) from other three categories of ECG recordings, that is, normal rhythm (N), other rhythm (O), and noisy (~) ECG recordings. So, the four categories, that is, N, A, O, and ~ were identified from the database provided by PhysioNet/CinC Challenge 2017. The proposed method first unified the 9 to 60 seconds unbalanced ECG recordings into 30 s segments by copying, cutting, and symmetry. Then, 24 artificial features including waveform features, interval features, frequency-domain features, and nonlinear feature were extracted relying on prior knowledge. Meanwhile, a 13-layer one-dimensional convolutional neural network (1-D CNN) was constructed to yield 38 abstract features. Finally, 24 artificial features and 38 abstract features were fused to yield the feature matrix. Random forest was employed to classify the ECG recordings. In this study, the mean accuracy (Acc) of the four categories reached 0.857. The F_1 of N, A, and O reached 0.837. The results exhibited the proposed method had relatively satisfactory performance for identifying AF from short single-lead ECG recordings with low SNR.

1. Introduction

Atrial fibrillation (AF) is a disordered and rapid atrial electrical activity characterized by supraventricular tachyarrhythmia. Its incidence increases with age, and millions of people are affected by AF every year [1]. In practice, real-time monitoring of cardiovascular disease is essential for early warning of AF. At present, wearable electrocardiogram (ECG) monitoring is the mainstream real-time monitoring system [2], which can help patients get rid of discomfort and time and place restrictions in the process of long-term health monitoring. However, the ECG recordings collected by wearable devices or mobile phones are easily contaminated by the complex external environment so that their signal-to-noise ratio (SNR) is low. Actually, many recordings with low SNR cannot be used for diagnosis because of their poor quality. Thus, the ECG recordings with low SNR also should be identified to avoid wasting clinical resources.

Traditional machine learning algorithms based on statistics were extensively used for data analysis [3–6]. Most of the current studies on AF automatic analysis do not focus on recognizing the noisy ECG recordings with low SNR. Krasteva et al. [3] used the limited feature set and combined with the optimized artificial neural network to conduct four-classification research on the CinC 2017 database. Goodfellow et al. [4] extracted three types of features, that is, template features, RRI features, and full waveform features using step-by-step machine and classified the CinC 2017 database into four categories. In general, previous studies can be divided into machine learning methods based on prior knowledge extracting artificial features and deep learning methods based on neural networks. Bin et al. [5] extracted 30 features including AF features, morphological features, and RR interval features from ECG recordings and trained a decision tree model using AdaBoost.M2 algorithm to realize AF detection. Datta et al. [6] extracted several

categories of AF features, that is, morphological features, HRV, frequency domain, and statistical features from PhysioNet/CinC Challenge 2017 database. They first transformed a four-classification problem into two binary classification problems because the performance of binary classifier is better than that of single multi-class classifier and then used a binary classifier to classify the two binary classification problems. Finally, the ECG recordings were divided into four categories, that is, normal, AF, other, and noisy ECG recordings. Pham et al. [7] first generated third-order cumulant images from four categories of ECG recordings and extracted 18 features including entropy features and other texture-based features. They used multiple classifiers to classify the recordings into four categories, that is, N_{sr}, V_{fib}, A_{fl}, and A_{fib}. The results exhibited random forest achieved the best performance than other algorithms, that is, KNN, J48 DT, PART rules, MLP, logistic regression, and Gaussian naive Bayes. Parsi et al. [8] extracted seven new features using the Poincare representation of the R-R interval series and fused the new features with classical features to predict the paroxysmal AF. Yue et al. [9] used frequency slice wavelet transform (FSWT) to analyze the ECG segments and converted the obtained two-dimensional (2-D) time-frequency matrix into a one-dimensional (1-D) feature vector. Finally, five machine learning methods were compared to classify AF, among which the Gaussian-kernel support vector machine has the best classification performance. The classical methods need a lot of artificial features that rely on the researchers' experience. However, more artificial features are not always better because some are redundant and may even descend classification accuracy.

Another method based on convolutional neural network (CNN) is widely used in physiological signal analysis [10, 11]. CNN can acquire implicit and abstract features within the ECG recordings by the convolutions of various structures without human intervention. Kachuee et al. [12] proposed a deep CNN model for heartbeat classification, which can accurately classify five different arrhythmias with the AAMI EC57 standard. Andersen et al. [13] proposed an end-to-end method combining recurrent neural network (RNN) and CNN to extract depth features from RR interval and divide the ECG recordings into AF and normal categories. Wang [14] designed an 11-layer network architecture based on CNN and Elman neural network to realize AF detection. By comparing several advanced classification methods, the combination of the two deep neural networks was confirmed to be feasible. Fan et al. [15] designed a multiscale fusion CNN structure to divide the ECG recordings into AF and normal categories. They used filters of different sizes to obtain features of different scales from 1-D ECG recordings and classified the recordings after feature fusion. Zhang et al. [16] proposed a global hybrid multiscale CNN which can fully extract features to realize the categories of AF and normal recordings. Acharya et al. [17] designed a 9-layer CNN model to automatically identify five heartbeat categories in ECG recordings, and they also tested the model in an original recording group and a noise attenuation recording group.

Actually, with the adoption of wearable devices and mobile phones, the ECG recordings collected using the devices are easy to be contaminated by noise so that the recordings cannot be used for clinical purpose because of their poor quality. So, the noisy ECG recordings should be recognized before diagnosing. Thus, it is necessary to distinguish the acceptable ECG recordings and the noisy ECG recordings from a large lot of ECG recordings with low SNR. In previous studies, entropy helped identify the inherent nonlinear property within the ECG recordings and randomness [18]. Zhang et al. [19] calculated a multiscale entropy of the ECG recordings for signal quality assessment and further studied the sensitivity of multiscale entropy on the ECG recordings with noise. Pham et al. [7] extracted a large number of entropy features to train classifiers. Fu et al. [20] extracted different entropy features, that is, approximate entropy, sample entropy, and fuzzy entropy to feed into machine learning, that is., support vector machine (SVM), least-squares SVM (LS-SVM), and long short-term memory (LSTM) for assessing the quality of the ECG recordings. Zhang et al. [21] proposed a permutation ratio entropy (PRE) based on permutation entropy to identify random components and inherent irregularities within time series. The studies exhibited a satisfying performance of entropy methods for identifying random components and inherent irregularities within the recordings. Thus, this study used the entropy feature, namely, PRE, to identify the noisy ECG recordings and other ECG recordings.

So, a novel method was proposed in this study, which used feature fusion including artificial features and abstract features to extract comprehensive information within the ECG recordings, and the entropy feature was also employed to improve classification performance of the method for noisy ECG recordings. In this study, Section 2 introduces materials and methods, including data preparation, feature extraction, and network architecture. Section 3 shows the results of this research. Section 4 discusses the effectiveness of this proposed method. Section 5 summarizes this work.

2. Materials and Methods

2.1. Database. The publicly available database provided by PhysioNet/CinC Challenge 2017 (CinC 2017) was used in this study, and it contains four categories of ECG recordings, that is, normal rhythm (N), AF rhythm (A), other rhythm (O), and noisy (~) ECG recordings. This database consisted of 8528 single-lead ECG recordings ranging in length from 9 s to over 60 s and the ECG recordings sampled at 300 Hz [22]. All recordings were identified by the clinical experts and technicians. Among them, 5076 ECG recordings were marked as N, 758 ECG recordings were marked as A, 2415 ECG recordings were marked as O, and 279 ECG recordings were marked as ~. These ECG waveforms are shown in Figure 1.

This study used a data-balanced method based on the imbalance of ECG recordings length, and the method effectively retained the critical information of the ECG

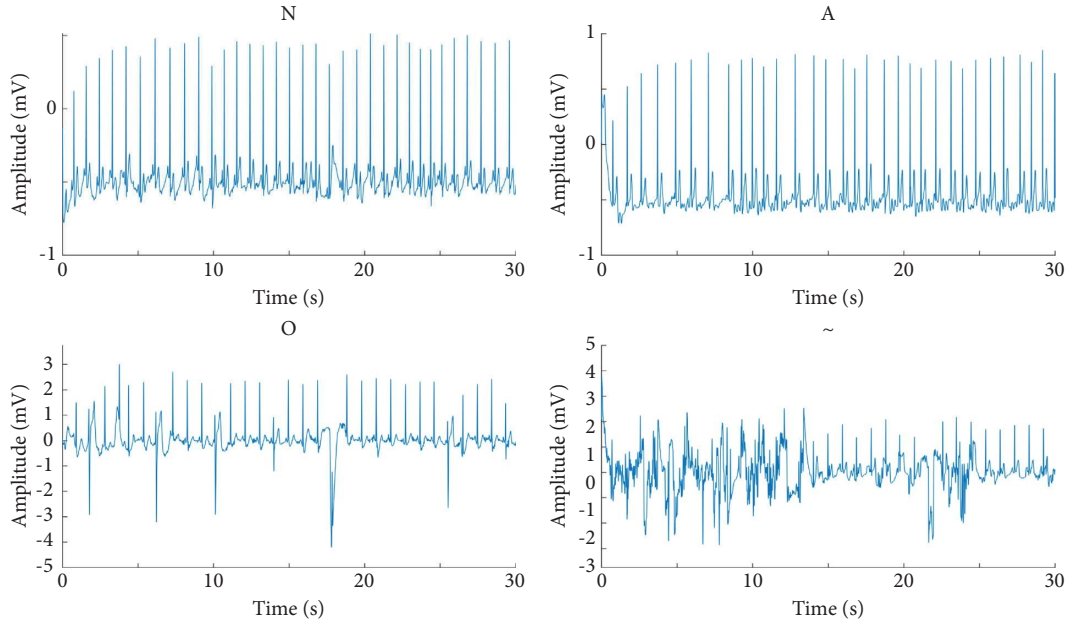

Figure 1: Examples of four categories of ECG recordings.

recordings [23]. A QRS complex location algorithm was used to locate the complex position and made the recording length consistent by copying, cutting, and symmetry. In this study, all recordings were segmented or filled to 30 s. Among them, the ECG recordings with lengths greater than 30 s were randomly segmented. The recordings with lengths less than 30 s were first located to the QRS complex using the Pan–Tompkins algorithm, then the initial downward deflection in the QRS complex was determined as the starting point of the complex, and finally the recording from the starting point of the first QRS complex to the starting point of the last QRS complex was intercepted and copied until the recording length was 30 s. After unifying the length of all segments, nearly 80% of the segments were used as training set and the remaining 20% as the test set. The performance of the proposed classification method was evaluated using the remaining segments. Table 1 shows the details of the CinC 2017 database used in this study.

2.2. Outline of the Proposed Method. In this study, the ECG recordings were first unified to the length of 30 s. Then, 62 features were calculated, including 24 artificial features, that is, 8 waveform features, 11 interval features, 4 frequency-domain features, and 1 nonlinear feature and 38 abstract features extracted by a 13-layer 1-D CNN. The abstract and artificial features constituted a feature vector for yielding the fused feature matrix. Finally, a random forest [24] containing 300 decision trees was employed to classify the AF segments. Figure 2 shows the flowchart of the proposed method.

2.3. Artificial Features. In the field of machine learning, the use of artificial features is essential. Based on a large number of previous studies, this study used four types of features, that is, waveform features, interval features, frequency-domain features, and nonlinear feature without discarding prior knowledge, and 24 specific features were calculated [4–8]. Table 2 shows the artificial features used in this study.

2.3.1. Waveform Features. In most cases, the number and amplitude of R waves within the four categories of ECG segments are significantly different, so the features based on the number and amplitude of R waves were first calculated. The Pan–Tompkins algorithm [25] was used to locate the R waves of all ECG segments. Then, the number of R waves and amplitude of all R waves were obtained by the location of R waves. Finally, the number of R waves was taken as one of the features, and the basic amplitude features, that is, maximum, minimum, mean, and median of R wave, in each segment were calculated according to the amplitude of all R waves. Suppose that there are N pieces of R waves in the time series. The r represents the amplitude of R wave. Therefore, the amplitude of all R waves is defined as $[r_1, r_2, r_3 \ldots\ldots r_N]$, so the maximum value of the amplitude is $[r_1, r_2, r_3 \ldots\ldots r_N]_{\max}$, the minimum value is $[r_1, r_2, r_3 \ldots\ldots r_N]_{\min}$, and the mean value is $[r_1, r_2, r_3 \ldots\ldots r_N]_{\text{median}}$.

In the analysis of time series, many time series exhibit irregular distribution. Still, the distribution of the mean of the series shows a certain regularity, which requires that we must have an indicator to measure the relationship between each point in the series and the mean. So, the standard deviation was used to distinguish the pseudo law of distribution in this study. Another waveform feature, namely, the feature based on standard deviation, was also calculated in this study. Suppose the time series with N points is defined as $[X_1, X_2, X_3 \ldots\ldots X_N]$, and their mean value is $\overline{X}$. The standard deviation (S) is calculated as the following:

Table 1: Details of the CinC 2017 database.

Category	Recording	Training set	Test set	Sampling frequency (Hz)	Uniform length (s)
N	5076	4101	975	300	30
A	758	619	139		
O	2415	1916	499		
~	279	208	71		

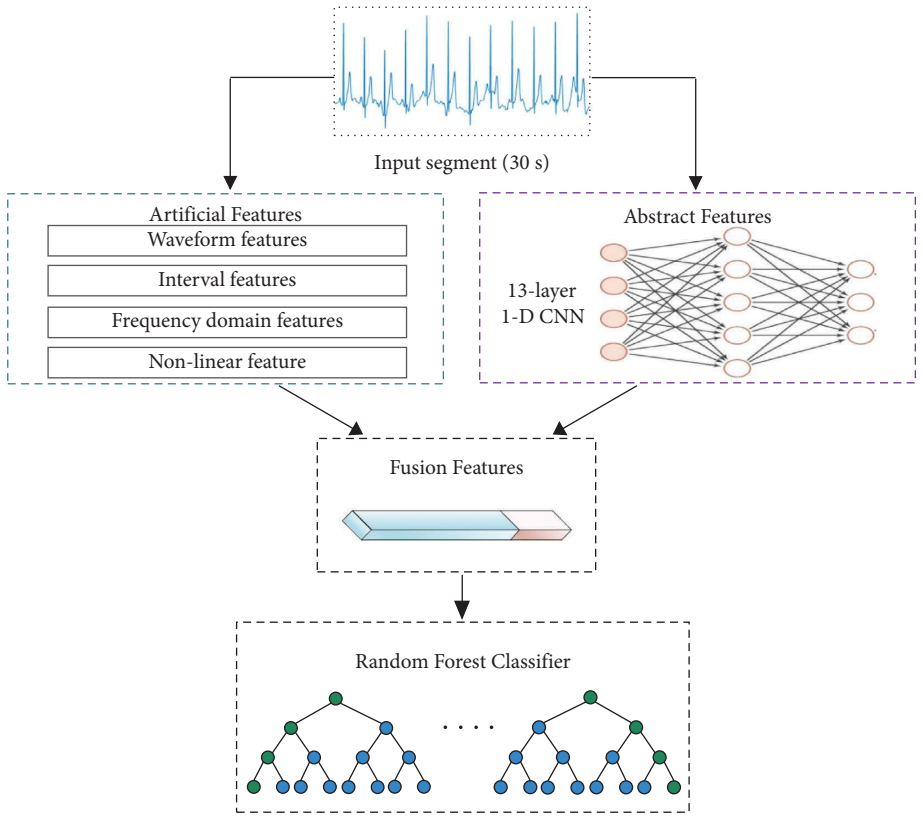

Figure 2: Flowchart of the proposed method.

$$S = \sqrt{\frac{\sum_{i=1}^{N}(X_i - \overline{X})^2}{N}}, \tag{1}$$

where i takes a non-negative integer and starts from 1 until N. According to the definition of S, the amplitude standard deviation is also calculated as one of the waveform features.

Based on the standard deviation, the skewness (SK) and kurtosis (KU) of the segments were calculated. SK represents the characteristic number of the asymmetry degree of the probability density distribution curve relative to the average value, and KU represents the characteristic number of the peak height of the probability density distribution curve at the average value. SK is calculated as the following:

$$\mathrm{SK} = \frac{\sum_{i=1}^{N}(X_i - \overline{X})^3}{(N-1)S^3}. \tag{2}$$

KU is calculated as the following:

$$\mathrm{KU} = \frac{\sum_{i=1}^{N}(X_i - \overline{X})^4}{(N-1)S^4} - 3. \tag{3}$$

To sum up, 8 waveform features were extracted from the ECG segments.

2.3.2. Interval Features. RR interval refers to the duration between two adjacent R waves in ECG, and it can reflect the duration of one heart contraction. These features of RR interval can reflect whether a person's heart rate is normal, so heart rate can be calculated by the RR interval [26]. The heart rate of patients with AF or other abnormal hearts may be irregular, and the RR interval may be too large, too small, or unstable. Therefore, the relevant features of RR interval, that is, maximum, minimum, mean, median, and standard deviation of RR interval were calculated, and the heart rate was also obtained from the RR interval as a feature.

Heart rate (HR) is calculated as the following:

Table 2: Artificial features used in this study.

Feature type	Name
Waveform features	The number of R wave
	Maximum amplitude of R wave
	Minimum amplitude of R wave
	Mean amplitude of R wave
	Median amplitude of R wave
	Amplitude standard deviation of R wave
	SK
	KU
Interval features	Maximum of RR interval
	Minimum of RR interval
	Mean of RR interval
	Median of RR interval
	Standard deviation of RR interval
	HR maximum of PR interval
	Minimum of PR interval
	Mean of PR interval
	Median of PR interval
	Standard deviation of PR interval
Frequency-domain features	FC
	MSF
	RMSF
	FV
Nonlinear feature	PRE

$$\mathrm{HR} = \frac{60}{\overline{R}}. \tag{4}$$

PR interval refers to the time interval from the starting point of the P wave to the starting point of the QRS complex on ECG. Some studies have used and proved the effectiveness of PR interval for ECG classification [3, 27, 28]. To get the PR interval, the P wave of the ECG recording should be located. P wave is easy to detect in regular ECG recordings, but it is difficult to detect in noise environment because the change is not obvious. Therefore, we used the P-wave detection method based on wavelet transform proposed by Li et al [29]. The PR interval was then calculated. Too long, too short, or variable PR interval represents different conditions of patients. Considering that there may be different situations for separating other classes in this database to locate these situations to the greatest extent, the relevant features of PR interval, that is, maximum, minimum, mean, median, and standard deviation of PR interval were extracted in this study. The calculation methods of relevant features of PR interval are the same as that of RR interval.

Finally, 6 features of RR interval and 5 features of PR interval were extracted from the ECG segments.

2.3.3. Frequency-Domain Features. In most of machine learning methods, frequency-domain features are usually used to reflect frequency and energy information within the ECG recordings. In medical diagnosis or other application scenarios, it can be used as a part of the feature vector together with time-domain features and other features to enrich the types of feature quantities and improve the diagnostic accuracy [30]. In this study, Fourier transform, a simple spectrum analysis method, was selected to obtain the spectrum of the ECG segments and the four frequency-domain features, that is, frequency center of gravity, mean-square frequency, root mean square frequency, and frequency variance were received and applied to this study.

Assuming the frequency function is $S(f)$, and S represents the spectrum and f represents the frequency of the segment. The frequency center of gravity (FC) is calculated as follows:

$$\mathrm{FC} = \frac{\int_0^\infty fS(f)\mathrm{d}f}{\int_0^\infty S(f)\mathrm{d}f}. \tag{5}$$

The mean-square frequency (MSF) is calculated as follows:

$$\mathrm{MSF} = \frac{\int_0^\infty f^2S(f)\mathrm{d}f}{\int_0^\infty S(f)\mathrm{d}f}. \tag{6}$$

The root mean square frequency (RMSF) is calculated as follows:

$$\mathrm{RMSF} = \sqrt{\mathrm{MSF}}. \tag{7}$$

The frequency variance (FV) is calculated as follows:

$$\mathrm{FV} = \frac{\int_0^\infty (f - \mathrm{FC})^2S(f)\mathrm{d}f}{\int_0^\infty S(f)\mathrm{d}f}. \tag{8}$$

Finally, 4 features of frequency domain were extracted from the ECG segments.

2.3.4. Nonlinear Feature. In some ECG classification studies, nonlinear features are widely used, especially various entropies are used to evaluate the complexity of signals. Many entropies, that is, Shannon entropy and permutation entropy, still cannot identify the nonlinear features in the signal. PRE was employed in the proposed method because it can identify nonlinear within ECG recordings, and the details of the PRE are in Reference [21]. This PRE can reflect the amplitude difference between two adjacent data points of a certain time series. Because it is sensitive to recording mutation and various changes, the classical permutation entropy is often used to measure the complexity of physiological recording sequence. However, the original time series cannot be measured by permutation entropy, so some details will be lost. Furthermore, permutation entropy is based on the ranking between data points, which also shows that permutation entropy ignores the differences between adjacent data points. Comparing with the classical permutation entropy, the PRE can reflect the relationship between adjacent data points by constructing the relationship matrix of adjacent elements and better reflecting the confusion degree of time series.

First, PRE constructs a new relationship matrix B to represent the relationship between adjacent elements and then calculates the number of new patterns c. Let $B(i)$ be the ith row vector of matrix B, and $c(i)$ be the number of the ith pattern. For $B(i)$, when another vector $B(j)$ of matrix B has the same mode as $B(i)$, $c(i)$ increases by 1, and the two have a high correlation; when each vector of matrix B represents a new mode, the maximum total number of mode c is $n-m-1$. Finally, the total number of mode c contained in matrix B can be obtained.

P_i is the probability of pattern $c(i)$, which is defined as the following:

$$P_i = \frac{c(i)}{\sum_{j=1}^{k} c(i)}, \tag{9}$$

where k is the total number of patterns c, $1 \le k \le n-m-1$.

PRE is defined as the following:

$$\mathrm{PRE} = -\sum_{j=1}^{k} P_j \ln P_j. \tag{10}$$

2.4. 1-D CNN and Abstract Features. Actually, a deeper network helps to extract deeper features within ECG segments; however, the most severe problem of deeper network was to use too many parameters, which would lead to a large amount of memory and computing resources for training and interference [31]. So, a 1-D CNN was directly used to extract abstract features in this study which was constructed from six pairs of convolutional layers and a maximum pooling layer in our proposed feature extraction network.

Larger convolution kernel size had been used on the first layer of convolution layers, and the convolution kernel size rose stepwise as the number of layers increased. Table 3 shows architecture of the 13-layer 1-D CNN and its detailed parameters. When an ECG segment was fed into the network, the segment passed through 6 pairs of convolution pooling layers. In order to obtain the abstract features, the final full connection layer changed the dimension of the output to get a 1×38 vector which meant 38 abstract features.

2.5. Fusion of Artificial and Abstract Features. Artificial features and abstract features were fused, and a feature vector of length 62 was constructed. The vector was denoted as $[R_1, R_2, R_3 \ldots R_{24}, S_1, S_2, S_3 \cdots S_{38}]^{\mathrm{T}}$. The R_i represents the ith artificial features, and $i = 1, 2, \ldots, 24$. The S_j represents the jth abstract features, and $j = 1, 2, \ldots, 38$. So, the feature matrix is defined as the following:

$$\begin{pmatrix} R_1^1 & \cdots\cdots & R_1^N \\ \vdots & & \vdots \\ R_{24}^1 & \ddots & R_{24}^N \\ S_1^1 & & S_1^N \\ \vdots & & \vdots \\ S_{38}^1 & \cdots\cdots & S_{38}^N \end{pmatrix}, \tag{11}$$

where N represents the number of input segments.

2.6. Random Forest. In CinC 2017, Zabihi et al. [32] and Kropf et al. [33] used random forest to train the extracted features to obtain classification results because random forest is interpretable explain [34]. So, random forest was employed in this study. Random forest is inherited together by several decision trees. Each decision tree is a small classifier, and random forests synthesize all classification voting results to determine the final output categories.

In this study, the classification of random forest included training and testing, and the bootstrap method was used to train the random forest. In the training process, 80% of the feature vectors were used as the training set, and a group of decision trees was trained according to the tags marked in the ECG recordings. The remaining 20% was used for testing. The training process sets the maximum number of decision trees as 300, where each node randomly selected features in the generation process. Assuming that the number of the samples was n, the number of features in the randomly selected feature subset by the decision tree node at each segmentation was set as default, that is, the square root of the total number of features, that is, $\sqrt{n}$. The minimum number of samples required for internal node division was set as 2, the maximum depth of the decision tree was set as 40, and the training ended when the maximum depth was reached. The above parameters were set to prevent overfitting. Finally, the classification category was determined by averaging the classification voting results of all decision trees.

Table 3: Architecture parameters of the 1-D CNN.

No.	Layer	Kernel size	Kernel number	Stride	Output size
0	Input	—	—	—	1×3000
1	Convolution-1	5	4	1	4×2996
2	Pooling	—	—	2	4×1498
3	Convolution-2	5	8	1	8×1494
4	Pooling	—	—	2	8×747
5	Convolution-3	7	16	1	16×741
6	Pooling	—	—	2	16×370
7	Convolution-4	7	16	1	16×364
8	Pooling	—	—	2	16×182
9	Convolution-5	9	32	1	32×174
10	Pooling	—	—	2	32×87
11	Convolution-6	11	32	1	32×77
12	Pooling	—	—	2	32×38
13	FC	—	—	—	1×38

3. Results

3.1. Evaluation Indicators. In this study, accuracy (Acc), precision, recall, and F_1 were used to evaluate performance of the proposed method.

The Acc is calculated as the following:

$$\text{Acc} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{FN} + \text{TN} + \text{FP}}, \tag{12}$$

where true positive (TP) represents the number of ECG recordings in a given category that are correctly classified as the given category, false positive (FP) represents the number of ECG recordings that other categories are misclassified as the given category, true negative (TN) represents the number of ECG recordings that other categories are not classified as the given category but are classified as the correct category, and false negative (FN) represents the number of ECG recordings that other categories are not classified as the given category and are not classified as the correct category.

The precision is calculated as the following:

$$\text{precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}. \tag{13}$$

The recall is calculated as the following:

$$\text{recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}. \tag{14}$$

Like CinC 2017, the F_{1n}, F_{1a}, F_{1o}, and F_{1p} are defined as the F_1 score of the *N*, *A*, *O,* and ~ categories, respectively, and they are calculated as the following [22]:

$$\begin{aligned} F_{1n} &= \frac{2 \times Nn}{\sum N + \sum n}, \\ F_{1a} &= \frac{2 \times Aa}{\sum A + \sum a}, \\ F_{1o} &= \frac{2 \times Oo}{\sum O + \sum o}, \\ F_{1p} &= \frac{2 \times Pp}{\sum P + \sum p}. \end{aligned} \tag{15}$$

Where *Nn, Aa, Oo, and Pp* represent the number of predicted classifications obtained by the proposed method that are consistent with the actual reference classifications of ECG recordings. $\sum N$ represents the number of recordings whose reference classification is *N* and $\sum n$ represents the number of recordings whose predicted classification is *N*, $\sum A$ represents the number of recordings whose reference classification is *A* and $\sum a$ represents the number of recordings whose predicted classification is *A*, $\sum O$ represents the number of recordings whose reference classification is *O* and $\sum o$ represents the number of recordings whose predicted classification is *O*, and $\sum P$ represents the number of recordings whose reference classification is ~ and $\sum p$ represents the number of recordings whose predicted classification is ~. Table 4 clearly showed the counting rules of the above variables. The total of F_1 is defined according to the rules of the CinC 2017 and it is obtained by taking the macro average of the three scores, and it is defined as the following:

$$F_1 = \frac{F_{1n} + F_{1a} + F_{1o}}{3}. \tag{16}$$

3.2. Results. In this study, 80% of the ECG segments were used as training set, and the rest 20% were used as test set for evaluating the proposed method. For the training set, we used 10-fold cross-validation which randomly selected 90% of the data for training and 10% for validation. The results are shown in Table 5. The corresponding recall, precision, and F_1 of the *N* category achieved the highest 0.896, 0.910, and 0.913 than that of other three categories, that is, *A*, *O,* and ~. In addition, the average of indicators of four categories, that is, recall, precision, and F_1, is higher than 0.800, at 0.816, 0.813, and 0.809, respectively.

Table 6 shows a confusion matrix of the proposed method for the test set and the corresponding recall, precision, F_{1n}, F_{1a}, F_{1o}, F_{1p}, *Acc,* and F_1. The *N* category yields the highest recall of 0.893, precision of 0.901, and F_{1n} of 0.901 than other categories, that is, *A*, *O,* and ~. The ~ category yields the lowest recall of 0.761, precision of 0.711, and F_{1p} of 0.735 among all categories. In addition, the F_1 and the Acc reached 0.837 and 0.857, respectively.

Table 4: Counting rules for some variables.

		Predicted classification				
		N	A	O	~	Total
Reference classification	N	Nn				$\sum N$
	A		Aa			$\sum A$
	O			Oo		$\sum O$
	~				Pp	$\sum P$
	Total	$\sum n$	$\sum a$	$\sum o$	$\sum p$	

Table 5: Results of using 10-fold cross-validation against the training set.

Label	Recall	Precision	F_{1n}	F_{1a}	F_{1o}	F_{1p}
N	0.896	0.910	0.913	—	—	—
A	0.814	0.827	—	0.806	—	—
O	0.808	0.788	—	—	0.795	—
~	0.745	0.726	—	—	—	0.721
Average	0.816	0.813		0.809		

Table 6: Confusion matrix of 1-DCNN for test set.

True	Predicted				Recall	Precision	F_{1n}	F_{1a}	F_{1o}	F_{1p}	Acc	F_1
	N	A	O	~								
N	871	13	83	8	0.893	0.908	0.901	—	—	—	0.857	0.837
A	11	110	15	3	0.791	0.815	—	0.803	—	—		
O	72	7	409	11	0.820	0.796	—	—	0.808	—		
~	5	2	7	54	0.761	0.711	—	—	—	0.735		—

Table 7 collected the results of some previous studies and compared them with the results of the proposed method. The proposed method achieved the highest Acc of 0.857, F_{1p} of 0.735 than all studies and the highest F_1 of 0.837 than all studies except the F_1 0.841 of Wang et al. [35]. Actually, Wang et al. ignored the ~ category of ECG recordings and used only three categories of ECG recordings of the CinC 2017, that is, N, A, and O for classification. Zihlmann et al. [39] combined LSTM and CNN to extract abstract features, and the total F_1 score reached 0.820. The classification results of ~ category in the training process were low, and the F_{1p} was only 0.645.

4. Discussion

4.1. Evaluating Effectiveness of PRE for Noisy Recording Recognition. Two feature schemes, that is, all features and all features except the PRE were compared to evaluate the effectiveness of the PRE for recognizing noisy ECG segments. Table 8 shows the comparison results for the two feature schemes using this proposed method. The Acc of 0.857, F_1 of 0.837, and F_{1p} of 0.735 for all features are higher than for all features except the PRE. The results indicate the PRE helps to classify the noisy ECG segments because the F_{1p} of 0.735 for all features is higher than the F_{1p} of 0.679 for all features without PRE. Meanwhile, a radar chart was also designed to show more clearly the differences between results of the two schemes. Figure 3 shows a radar map of results for the two feature schemes. The F_{1p} for all features is obviously higher than that for all features except the PRE.

PRE was an improvement based on permutation entropy for identifying nonlinear chaotic character within time series instead of randomness. In PRE, a new relationship matrix B was constructed. This matrix was based on the relationship between adjacent elements and can closely reflect the gap between two points, especially in complex signals. The generation of the new mode c can avoid the repeated counting of the vector and was conducive to the complexity analysis of the whole signal. The ablation experiment showed the PRE not only played a role in noise classification but also helped the overall classification indicators.

4.2. Comparison of Effectiveness of Artificial, Abstract, and Fusion Features. In this study, the corresponding Accs of the three feature schemes, that is, artificial features, abstract features, and fusion features were also calculated to evaluate the effectiveness of the schemes. Table 9 shows the corresponding Accs of artificial features, abstract features, and fusion features. The Acc of 0.820 was obtained for the scheme using only artificial features. Similarly, the Acc for only abstract features generated by the 13-layer 1-D CNN was the lowest 0.734 than that for all feature schemes.

Actually, deep learning can extract effective abstract features with the support of a large amount of data. However, the existed ECG databases are small so that deep learning algorithms cannot make full use of its power for

Table 7: Comparison of classification results.

Author	Year	Database	Feature extraction	Task	Method	Acc	F_{1p}	F_1
Datta et al. [6]	2017	CinC 2017 AF DB	HRV, frequency domain, and statistical features	4-Class	Multilayer cascaded binary classifiers	—	—	0.830
Cao et al. [23]	2020	CinC 2017 AF DB	Abstract features	3-Class	2-Layer LSTM	0.844	—	0.827
Zabihi et al. [32]	2017	CinC 2017 AF DB	Time domain, frequency domain, time-frequency domain, and nonlinear features	4-Class	Random forest	—	0.504	0.830
Kropf et al. [33]	2017	CinC 2017 AF DB	Time-domain and frequency-domain features	4-Class	Random forest	—	0.648	0.830
Wang et al. [35]	2020	CinC 2017 AF DB	Abstract features	3-Class	DMSFNet	—	—	0.841
Gao et al. [36]	2021	CinC 2017 AF DB	Abstract features	3-Class	RTA-CNN	0.851	—	—
Mahajan et al. [37]	2017	CinC 2017 AF DB	Time domain, frequency domain, linear, and nonlinear features	4-Class	Random forest	—	—	0.780
Xiong et al. [38]	2017	CinC 2017 AF DB	Abstract features	4-Class	CNN	—	—	0.820
Zihlmann et al. [39]	2017	CinC 2017 AF DB	Abstract features	4-Class	CNN + LSTM	0.823	0.645	0.820
Gliner and Yanav [28]	2018	CinC 2017 AF DB	Time-frequency domain, statistical features, and morphological features	4-Class	SVM	—	—	0.800
Athif et al. [40]	2018	CinC 2017 AF DB	Statistical features and morphological features	4-Class	SVM	—	—	0.780
Chen et al. [41]	2018	CinC 2017 AF DB	Morphological features and heart rate variability features	4-Class	XGBoost	—	—	0.810
This work	2022	CinC 2017 AF DB	Time domain, interval, frequency domain, and nonlinear features and abstract features	4-Class	Fusion features + random forest	0.857	0.735	0.837

Table 8: Comparison results for the two feature schemes.

	Feature scheme	
	All features	All features except the PRE
Acc	0.857	0.836
F_1	0.837	0.822
F_{1p}	0.735	0.679

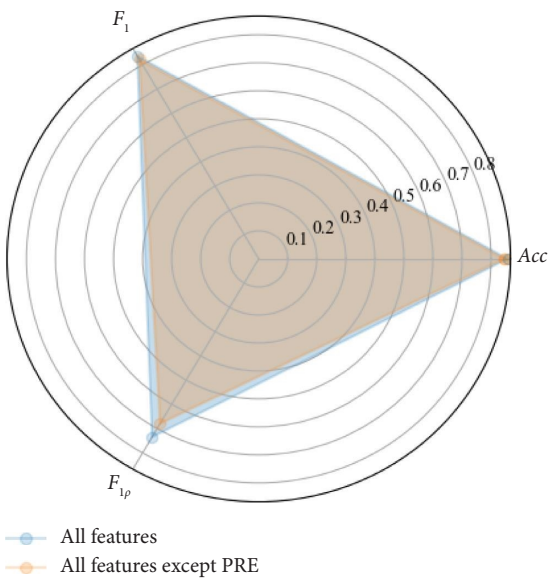

Figure 3: Radar map of results for the two feature schemes.

Table 9: Accs of three feature schemes, that is, artificial features, abstract features, and fusion features.

	Feature scheme		
	Artificial feature	Abstract feature	Fusion feature
Acc	0.820	0.734	0.857

acquiring abstract features. The artificial features were summarized on the basis of expert experience and a large number of experiments, and the features can reflect information within the ECG recordings. Therefore, abstract and artificial features were combined to make the model have the advantages of both, thus improving the classification performance of the model. After fusing artificial features and abstract features, the Acc was improved to the highest 0.857 among the Accs for all schemes. The fusion features gave full play to the advantages of the two types of features and can more comprehensively reflect the information in ECG recordings, so fusion features can improve the classification performance of such models.

5. Conclusions

In this study, an AF detection method that combined artificial features with abstract features was proposed, and it yielded the higher results, that is, Acc of 0.857, F_1 of 0.837, and F_{1p} of 0.735 for the database provided by the CinC 2017 than the previous studies. In addition, the nonlinear feature, that is, PRE, helps to identify the noisy ECG recordings from

other recordings because the PRE can identify, to some extent, nonlinear irregularities within the ECG recordings instead of randomness caused by noise. Finally, the proposed method exhibits relatively satisfied performance for the ECG recordings with low SNR.

Data Availability

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

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding this work.

Acknowledgments

This work was supported in part by the National Natural Science Foundation of China under Grant 82072014, 62076149 and 61702138, in part by the China Postdoctoral Science Foundation under Grant 2019M662360 and 2020T130368, in part by the Intergovernmental Project of National Key Research and Development Program/Hong Kong, Macao and Taiwan Key Projects under Grant SQ2019YFE010670, in part by Key R&D project of Shandong Province 2018GSF118133, in part by Young Scholars Program of Shandong University, Weihai under Grant 1050501318006, and in part by Science and Technology Development Plan of Weihai City of Shandong Province under Grant 1050413421912.

References

[1] K. M. Ryder and E. J. Benjamin, "Epidemiology and significance of atrial fibrillation," *The American Journal of Cardiology*, vol. 84, no. 9, pp. 131–138, 1999.

[2] P. S. Pandian, K. Mohanavelu, K. P. Safeer et al., "Smart Vest: wearable multi-parameter remote physiological monitoring system," *Medical Engineering & Physics*, vol. 30, no. 4, pp. 466–477, 2008.

[3] V. Krasteva, I. Christov, S. Naydenov, T. Stoyanov, and I. Jekova, "Application of dense neural networks for detection of atrial fibrillation and ranking of augmented ECG feature set," *Sensors*, vol. 21, no. 20, p. 6848, 2021.

[4] S. D. Goodfellow, A. Goodwin, R. Greer, P. C. Laussen, M. Mazwi, and D. Eytan, "Atrial fibrillation classification using step-by-step machine learning," *Biomedical Physics & Engineering Express*, vol. 4, no. 4, Article ID 45005, 2018.

[5] G. Bin, M. Shao, G. Bin, J. Huang, D. Zheng, and S. Wu, "Detection of atrial fibrillation using decision tree ensemble," in *Proceedings of the 2017 Computing in Cardiology (CinC)*, pp. 1–4, IEEE, Rennes, France, September 2017.

[6] S. Datta, C. Puri, A. Mukherjee et al., "Identifying normal, AF and other abnormal ECG rhythms using a cascaded binary classifier," in *Proceedings of the 2017 Computing in cardiology (CinC)*, pp. 1–4, IEEE, Rennes, France, September 2017.

[7] T. H. Pham, V. Sree, J. Mapes et al., "A novel machine learning framework for automated detection of arrhythmias in ECG segments," *Journal of Ambient Intelligence and Humanized Computing*, vol. 12, no. 11, Article ID 10145, 2021.

[8] A. Parsi, M. Glavin, E. Jones, and D. Byrne, "Prediction of paroxysmal atrial fibrillation using new heart rate variability features," *Computers in Biology and Medicine*, vol. 133, Article ID 104367, 2021.

[9] Y. Yue, C. Chen, P. Liu, Y. Xing, and X. Zhou, "Automatic detection of short-term atrial fibrillation segments based on frequency slice wavelet transform and machine learning techniques," *Sensors*, vol. 21, no. 16, p. 5302, 2021.

[10] B. Pourbabaee, M. J. Roshtkhari, and K. Khorasani, "Deep convolutional neural networks and learning ECG features for screening paroxysmal atrial fibrillation patients," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 48, no. 12, pp. 2095–2104, 2018.

[11] O. Yildirim, U. B. Baloglu, R. S. Tan, E. J. Ciaccio, and U. R. Acharya, "A new approach for arrhythmia classification using deep coded features and LSTM networks," *Computer Methods and Programs in Biomedicine*, vol. 176, pp. 121–133, 2019.

[12] M. Kachuee, S. Fazeli, and M. Sarrafzadeh, "Ecg heartbeat classification: a deep transferable representation," in *Proceedings of the 2018 IEEE International Conference on Healthcare Informatics (ICHI)*, pp. 443-444, IEEE, New York, NY, USA, June 2018.

[13] R. S. Andersen, A. Peimankar, and S. Puthusserypady, "A deep learning approach for real-time detection of atrial fibrillation," *Expert Systems with Applications*, vol. 115, pp. 465–473, 2019.

[14] J. Wang, "A deep learning approach for atrial fibrillation signals classification based on convolutional and modified Elman neural network," *Future Generation Computer Systems*, vol. 102, pp. 670–679, 2020.

[15] X. Fan, Q. Yao, Y. Cai, F. Miao, F. Sun, and Y. Li, "Multiscaled fusion of deep convolutional neural networks for screening atrial fibrillation from single lead short ECG recordings," *IEEE journal of biomedical and health informatics*, vol. 22, no. 6, pp. 1744–1753, 2018.

[16] P. Zhang, C. Ma, Y. Sun et al., "Global hybrid multi-scale convolutional network for accurate and robust detection of atrial fibrillation using single-lead ECG recordings," *Computers in Biology and Medicine*, vol. 139, Article ID 104880, 2021.

[17] U. R. Acharya, S. L. Oh, Y. Hagiwara et al., "A deep convolutional neural network model to classify heartbeats," *Computers in Biology and Medicine*, vol. 89, pp. 389–396, 2017.

[18] M. Kumar, R. B. Pachori, and U. Rajendra Acharya, "Automated diagnosis of atrial fibrillation ECG signals using entropy features extracted from flexible analytic wavelet transform," *Biocybernetics and Biomedical Engineering*, vol. 38, no. 3, pp. 564–573, 2018.

[19] Y. Zhang, S. Wei, Y. Long, and C. Liu, "Performance analysis of multiscale entropy for the assessment of ECG signal quality," *Journal of Electrical and Computer Engineering*, vol. 2015, Article ID 563915, 2015.

[20] F. Fu, W. Xiang, Y. An et al., "Comparison of machine learning algorithms for the quality assessment of wearable ECG signals via lenovo H3 devices," *Journal of Medical and Biological Engineering*, vol. 41, no. 2, pp. 231–240, 2021.

[21] Y. Zhang, C. Liu, S. Wei, Y. Liu, and H. Liu, "Complexity analysis of physiological time series using a novel permutation-ratio entropy," *IEEE Access*, vol. 6, Article ID 67653, 2018.

[22] G. D. Clifford, C. Liu, B. Moody et al., "AF classification from a short single lead ECG recording: the PhysioNet/computing

in cardiology challenge 2017," in *Proceedings of the 2017 Computing in Cardiology (CinC)*, pp. 1–4, IEEE, Rennes, France, September 2017.

[23] P. Cao, X. Li, K. Mao et al., "A novel data augmentation method to enhance deep neural networks for detection of atrial fibrillation," *Biomedical Signal Processing and Control*, vol. 56, Article ID 101675, 2020.

[24] C. Vimal and B. Sathish, "Random forest classifier based ECG arrhythmia classification," *International Journal of Healthcare Information Systems and Informatics*, vol. 5, no. 2, pp. 1–10, 2010.

[25] J. Pan and W. J. Tompkins, "A real-time QRS detection algorithm," *IEEE Transactions on Biomedical Engineering*, vol. BME-32, no. 3, pp. 230–236, 1985.

[26] S. Islam, N. Ammour, and N. Alajlan, "Atrial fibrillation detection with multiparametric RR interval feature and machine learning technique," in *Proceedings of the 2017 International Conference on Informatics, Health & Technology (ICIHT)*, pp. 1–5, IEEE, Riyadh, Saudi Arabia, February 2017.

[27] R. Banerjee, A. Ghose, and S. Khandelwal, "A novel recurrent neural network architecture for classification of atrial fibrillation using single-lead ECG," in *Proceedings of the 2019 27th European Signal Processing Conference (EUSIPCO)*, pp. 1–5, IEEE, A Coruna, Spain, September 2019.

[28] V. Gliner and Y. Yaniv, "An SVM approach for identifying atrial fibrillation," *Physiological Measurement*, vol. 39, no. 9, Article ID 94007, 2018.

[29] C. Li, C. Zheng, and C. Tai, "Detection of ECG characteristic points using wavelet transforms," *IEEE Transactions on Biomedical Engineering*, vol. 42, no. 1, pp. 21–28, 1995.

[30] A. S. Udawat and P. Singh, "An automated detection of atrial fibrillation from single-lead ECG using HRV features and machine learning," *Journal of Electrocardiology*, vol. 75, pp. 70–81, 2022.

[31] H. I. Lin and C. S. Cheng, "A study on accelerating convolutional neural networks," in *AIP Conference Proceedings*, vol. 2186, no. 1, AIP Publishing LLC, Article ID 150002, December 2019.

[32] M. Zabihi, A. B. Rad, A. K. Katsaggelos, S. Kiranyaz, S. Narkilahti, and M. Gabbouj, "Detection of atrial fibrillation in ECG hand-held devices using a random forest classifier," in *Proceedings of the 2017 Computing in Cardiology (CinC)*, pp. 1–4, IEEE, Rennes, France, September 2017.

[33] M. Kropf, D. Hayn, and G. Schreier, "ECG classification based on time and frequency domain features using random forests," in *Proceedings of the 2017 Computing in Cardiology (CinC)*, pp. 1–4, IEEE, Rennes, France, September 2017.

[34] L. Breiman, J. H. Friedman, R. A. Olshen, and C. J. Stone, *Classification and Regression Trees*, Routledge, London, Uk, 2017.

[35] R. Wang, J. Fan, and Y. Li, "Deep multi-scale fusion neural network for multi-class arrhythmia detection," *IEEE journal of biomedical and health informatics*, vol. 24, no. 9, pp. 2461–2472, 2020.

[36] Y. Gao, H. Wang, and Z. Liu, "An end-to-end atrial fibrillation detection by a novel residual-based temporal attention convolutional neural network with exponential nonlinearity loss," *Knowledge-Based Systems*, vol. 212, Article ID 106589, 2021.

[37] R. Mahajan, R. Kamaleswaran, J. A. Howe, and O. Akbilgic, "Cardiac rhythm classification from a short single lead ECG recording via random forest," in *Proceedings of the 2017 Computing in Cardiology (CinC)*, pp. 1–4, IEEE, Rennes, France, September 2017.

[38] Z. Xiong, M. K. Stiles, and J. Zhao, "Robust ECG signal classification for detection of atrial fibrillation using a novel neural network," in *Proceedings of the 2017 Computing in Cardiology (CinC)*, pp. 1–4, IEEE, Rennes, France, September 2017.

[39] M. Zihlmann, D. Perekrestenko, and M. Tschannen, "Convolutional recurrent neural networks for electrocardiogram classification," in *Proceedings of the 2017 Computing in Cardiology (CinC)*, pp. 1–4, IEEE, Rennes, France, September 2017.

[40] M. Athif, P. C. Yasawardene, and C. Daluwatte, "Detecting atrial fibrillation from short single lead ECGs using statistical and morphological features," *Physiological Measurement*, vol. 39, no. 6, Article ID 64002, 2018.

[41] Y. Chen, X. Wang, Y. Jung et al., "Classification of short single-lead electrocardiograms (ECGs) for atrial fibrillation detection using piecewise linear spline and XGBoost," *Physiological Measurement*, vol. 39, no. 10, Article ID 104006, 2018.