DESIGN OF REAL-TIME SYSTEM BASED ON MACHINE LEARNING FOR SNORING AND OSA DETECTION

Huaiwen Luo, Lu Zhang, Lianyu Zhou, Xu Lin, Zehuai Zhang, Mingjiang Wang

Department of Electronics and Information Engineering,
Harbin Institute of Technology, Shenzhen, China, 518000
20s152151@stu.hit.edu.cn, mjwang@hit.edu.cn

ABSTRACT

Obstructive sleep apnea (OSA) is a common sleep disorder. The diagnosis of OSA based on snoring is low-cost, convenient and non-invasive. In this study, we place a microphone under the patient's bed and combined with full-night polysomnography to record audio signals. Five machine learning models and two OSA diagnostic schemes are used to classify night audio as non-snoring, snoring, or OSA-related snoring. Our experiment has achieved good results, and the highest diagnosis rate of OSA can reach 97%. Based on the trained classification model, we design a system that can diagnose OSA in real-time. Tests on the system show that it can diagnose apnea by detecting OSA-related snoring. We hope that this approach can develop into a new tool to help a large number of potential OSA patients understand their sleep health.

Index Terms— Machine learning, Deep learning, Snore signal, OSA diagnosis, Real-time system

1. INTRODUCTION

Obstructive sleep apnea (OSA) is a sleep apnea disorder that affects about one in 20 people worldwide, which seriously affects the sleep status and health of patients [1]. At present, hospitals use sleep polysomnography (PSG) to diagnose OSA diseases, collect body signals overnight through professional equipment, and report all apnea and hypopnea events. However, the process of PSG is time-consuming and labor-intensive, so it is difficult to promote. Some studies have shown that the acoustic characteristics of snoring are related to OSA [2-4]. Therefore, we diagnose OSA based on the snoring signal. Compared with PSG, snoring collection does not require professional equipment, so it is simpler, more convenient, and popular, and has great social and commercial value.

The extraction of effective acoustic features is beneficial to the detection of snoring and the diagnosis of OSA. The uniqueness of sound is determined by the resonance level of different formants and different vocal cords [5]. Different sounds have high discrimination and separability in frequency [6]. Arun Sebastian et al use the feature selection algorithm to carry out experiments, and prove that Mel Cepstrum coefficient (MFCC) provides the most discriminative information [7]. In addition, the average chromatogram (MC) [8], the linear prediction coefficient (LPC) [9] and the short-term zero-crossing rate (ZRC) [10] are also used in this kind of task. In addition to manual feature extraction, neural networks have also been proved to be able to directly obtain high-dimensional abstract information in snoring [11].

The selection of classification model can also affect the classification performance. K-nearest neighbor (KNN) [12] algorithm, support vector machine (SVM) [13], logistic regression [14] and artificial neural network (ANN) [15] have been used to extract snoring signals from night audio. Arun Sebastian et al use LDA linear classifier to extract OSA-related snoring from nocturnal audio signals [7]. In reference [16], the convolutional neural network (CNN) is used to realize the automatic detection of apnea events. The generative adversarial network (GAN) can enhance the small sample snoring data [17], and the cyclic structural neural networks RNN [18] and LSTM [19] are also used in this kind of task.

In this study, we have established a sound database containing OSA-related snoring, designed and trained five machine learning models (Bayes, Decision Tree, RNN, LSTM and TCN) for snoring detection and OSA diagnosis. After analysis and comparison, we use a multi-classification TCN model to build a system architecture for OSA diagnosis. Our proposed architecture uses short input contexts, 1.04 seconds, to make predictions, making it capable of near real-time processing.

This paper is organized as follows: Section 2 introduces our main algorithms. Experiments and results are shown in Section 3. Section 4 analyzes and discusses the results. Finally, we summarize our paper in Section 5.

2. METHODS

Our research mainly includes four parts: the establishment of database, effective feature extraction, classification model design and real-time system construction. The database establishment method is shown in Figure 1:

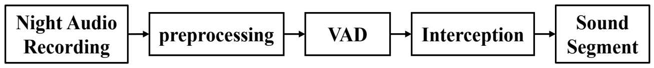

Fig.1: Establishment of sound database.

The processing of the original data mainly includes three stages: the first step is the preprocessing, including denoising and filtering. The second step is to complete the endpoint detection. The last step is to build the database through manual tagging. A supervised snoring detection algorithm can be carried out on the database, and the algorithm framework is shown in Figure 2:

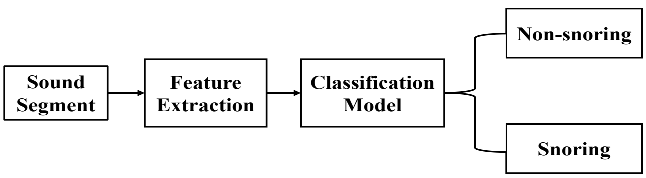

Fig.2: Block diagram of snoring detection.

The duration of the audio clip in the voice database is about 1-3 seconds. After feature extraction, the feature data with the same dimension and high separability can be obtained, which can be used to train the machine learning model and complete the snoring detection task. On this basis, the OSA diagnosis algorithm can be determined by improvement. To achieve this goal, we can add a two-step classifier to the original model to achieve two-step diagnosis, as shown in Figure 3, or directly train a multi-classifier to achieve single-step diagnosis, as shown in Figure 4.

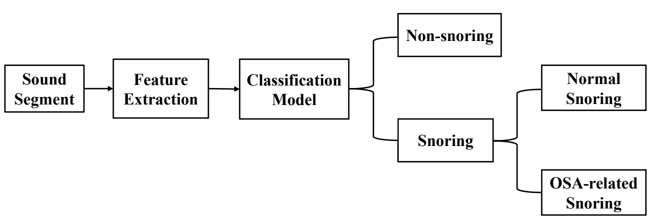

Fig.3: Block diagram of two-step OSA diagnosis.

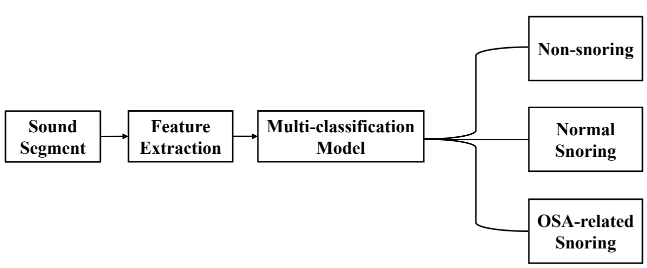

Fig.4: Block diagram of single-step OSA diagnosis.

Finally, after the analysis and comparison of the two diagnosis algorithms and a variety of machine learning models, the final classification model can be determined, which can be used to build a real-time system.

2.1. Database Establishment

The data of the snoring incident comes from Luohu people's Hospital. A total of 100 potential OSA patients are tested with PSG overnight. We place the microphone under the patient's bed to get the audio signal, and the sampling rate is 44.1 kHz. Among them, the normal snoring comes from the snoring event that is not labeled as OSA and the subjects who do not have apnea, the OSA-related snoring comes from the apnea events given in the PSG report. The collection of non-snoring events is from the internet, including the sound of vehicle sirens at night, the operation of electrical appliances (such as air conditioners, fans), natural sounds (wind, rain) and human voices and so on.

After getting the night audio signal, based on the frequency characteristics of snoring, we first use a 150-2000 Hz band-pass filter for denoising, and then design a multi-threshold endpoint detection algorithm to extract the sound segment. Finally, we label the extracted sound clips manually. Through this process, we set up a database with a total of 20465 sound segments, including 6545 non-snoring segments, 7502 normal snoring segments and 6418 snoring segments related to OSA. The duration of the sound segment is about 1-3 seconds.

2.2.Feature Extraction

Our previous research has shown that MFCC contains the key information in the snoring signal. Therefore, this paper extracts the MFCC coefficient from the sound segment, and calculates its statistical features to further compress the dimension. The process of feature extraction is as follows:

- Convert the sampling rate of the sound segment to 8 kHz.
- The spectrogram is obtained by short-time Fourier transform.
- The Mel spectrum is obtained by filtering with triangular filter banks.
- The signal of each frame is compressed by DCT, and the 12-dimensional MFCC coefficient of each frame is obtained.
- The MFCC coefficients are divided into four groups according to frames, and each group calculates four statistical features according to dimensions to form a 4*48-dimensional feature matrix.

For later network classification, we unify the sampling rate of all audio signals to 8 kHz. After feature extraction, the characteristic dimensions of audio segments with different duration can be unified and a certain correlation between frames can be retained, which is helpful for the sequential network model to exert its performance.

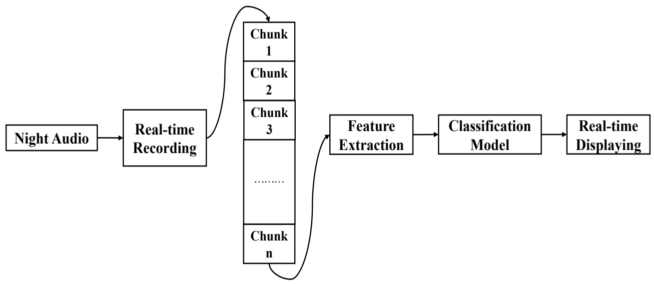

Fig.5: Block diagram of real-time system architecture.

2.3. Machine Learning

Temporal convolutional network (TCN) is a sequential neural network based on convolution connection, which has been proved to have excellent performance in processing sequential data. TCN has two characteristics, which are causal convolution and dilated convolution. Causal convolution abandons the connection with the anterior layer neurons, so that the data value at a certain time is only associated with the data at the previous moment, and there will be no information disclosure. Causal convolution can help the network to remember the past information, while the introduction of dilated convolution can make the network remember information for a longer time. Dilated convolution has a larger receptive field than ordinary convolution, but it does not increase the convolution filter parameters. It skips some input values to obtain information farther away from the current moment. We design an 18-layer TCN model and introduce the residual structure to avoid the problems of gradient disappearance, gradient explosion and network degradation. In addition, we also apply Bayes, Decision Tree (DT) classifier, RNN and LSTM network for analysis and comparison.

2.4. Architecture of Real-time System

The architecture of the system capable of real-time diagnosis of OSA is shown in Figure 5. It consists of two parts, one is the recording module, the other is the detection module. Among them, the recording module is responsible for collecting external sound, and the audio sampling rate is 8 kHz. The detection module is responsible for short-term processing of the collected audio to achieve real-time discrimination. The two modules can complete their own work independently and complement each other. The recording module continuously collects external sounds and stores them in the queue to form an audio chunk. At the same time, the detection module continues to extract the audio chunk from the storage queue, extract features from it, make a discrimination, and display the corresponding results in real-time. The chunk in the storage queue are audio blocks, which are similar in size to the audio clips in the database. Considering that if the size of the audio chunk is too large, multiple sound fragments may be mixed into it. We unify the length of the audio block to 1.04 seconds.

3. RESULTS

We first experiment with the snoring detection algorithm on the data set, then realize the diagnosis of OSA with two schemes, and finally build a real-time system. The data ratio of training set, verification set and test set is 10:1:5.

3.1. Snoring detection

The sound clips in the database are classified as snoring and non-snoring. The experimental results show that the snoring detection rates of DT, Bayes, RNN, LSTM and TCN are 94%, 84%, 95%, 95% and 98%, respectively. The detailed results are shown in Table 1, where Acc represents the average classification accuracy, Recall represents the average recall rate, and Snoring represents the single-class accuracy of snoring, showing the detection rate of snoring.

Table 1: Snoring detection results using five classification models.

Model	Acc(%)	Recall(%)	Snoring(%)
DT	92(91.1-92.3)	90(89.6-91.1)	94(93.5-94.1)
Bayes	81(80.9-81.8)	78(77.8-79.1)	84(83.8-85.9)
RNN	87(86.6-87.2)	82(80.2-84.3)	95(92.3-97.4)
LSTM	94(93.1-95.2)	94(93.6-93.9)	95(94.9-96.5)
TCN	96(96.3-97.4)	96(95.1-96.9)	98(98.2-98.6)

3.2. OSA diagnosis

On the basis of snoring detection, we continue to classify snoring as normal snoring and OSA-related snoring to achieve two-step diagnosis. The experimental results show that the OSA diagnosis rates of DT, Bayes, RNN, LSTM and TCN for OSA are 93%, 83%, 94%, 97% and 99%, respectively. The detailed results are shown in Table 2, where OSA is the single-class accuracy of OSA-related snoring, showing the diagnosis rate of OSA. Combined with the results of snoring detection, it can be calculated that the actual diagnosis rate of OSA in the two-step method is 87%, 70%, 89%, 93% and 97%, respectively.

Single-step diagnosis is to classify sound fragments as non-snoring, normal snoring and OSA-related snoring directly in the original database. The experimental results show that the diagnosis rate of DT, Bayes, RNN, LSTM and TCN to OSA is 88%, 81%, 88%, 95% and 97%, respectively. The detailed results are shown in Table 3.

Table 2: Results of the second stage of Two-step OSA diagnosis using five classification models.

Model	Acc(%)	Recall(%)	OSA(%)
DT	92(91.9-93.1)	92(91.9-92.9)	93(92.1-93.4)
Bayes	84(84.1-84.6)	84(83.9-84.5)	83(82.6-84.2)
RNN	87(85.4-88.6)	87(85.3-89.2)	94(93.3-95.6)
LSTM	94(93.8-95.1)	95(94.2-95.4)	97(96.8-98.8)
TCN	97(97.3-97.4)	97(97.4-97.5)	99(98.7-99.1)

Table 3:Single-step OSA diagnosis results using five classification models.

Model	Acc(%)	Recall(%)	OSA(%)
DT	87(86.4-87.1)	86(86.3-86.9)	88(87.1-89.5)
Bayes	74(73.6-75.2)	75(73.7-75.9)	81(80.1-82.7)
RNN	82(81.8-83.2)	82(81.5-83.2)	88(85.1-93.5)
LSTM	94(93.5-94.5)	94(93.4-94.5)	95(94.4-95.6)
TCN	96(95.9-96.3)	96(95.8-96.2)	97(96.4-97.6)

3.3. Real-time OSA diagnosis

TCN model has the highest detection rate of OSA-related snoring, and because of the convolution connection between networks, fewer parameters are needed. Considering the accuracy and real-time performance of the system, we choose the TCN model to build a real-time system. The result of TCN in single-step OSA diagnosis algorithm is similar to that in two-step OSA diagnosis algorithm, but only one classification model is needed to realize single-step diagnosis. To sum up, we build a real-time system based on a trained multi-classification TCN model.

The size of audio chunk in real-time system is 1.04 s, and a judgment is made about once a second. We tested the system with a night audio, as shown in Figure 6. The real-time system made a total of 28 judgments in this audio, and there was a misjudgment on the last OSA-related snoring, OSA-related snoring is classified as normal snoring, which may be due to the mixing of noise in this audio chunk. The test results show that this system can basically distinguish non-snoring, normal snoring and OSA-related snoring, and determine the time of apnea according to the location of OSA-related snoring.

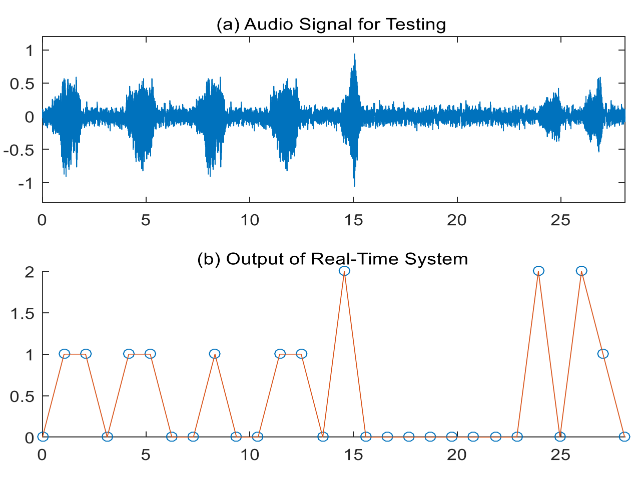

Fig.6: Part (a) shows the waveform of the test audio. There are four normal snoring clips in the first part of the audio, followed by apnea, including three OSA-related snoring clips . Part (b) gives the detection results from the real-time system, in which 0 is non-snoring, 1 is normal snoring, and 2 is OSA-related snoring.

4. DISCUSSION

Five classification models are applied in our experiment. On the whole, the classification accuracy of deep learning model is higher than that of traditional machine learning model. For the deep learning models, LSTM is better and more stable than RNN, but the training time is the longest. TCN is 2~4% higher than LSTM in the diagnostic task of OSA, and requires fewer parameters. The experimental results of the two schemes of OSA diagnosis show that single-step diagnosis can achieve better results on most machine learning models, and only one classification model is needed, which is more suitable for building real-time systems.

In the study of diagnosing OSA, our proposed method shows very competitive performance. One study uses feature selection algorithm and traditional classifier LDA to classify non-snoring, normal snoring and OSA-related snoring directly. The accuracy of extracting apnea events is 64% [7]. Another study discusses the snoring records of 49 Thai subjects, and the classification accuracy is 87.8% based on feature fusion [8]. SVM classifier is used to estimate the existence of OSA in reference [20], and the accuracy is 76.5%. In [16], CNN is used for automatic detection of apnea, with an average accuracy of 79.6%. A hybrid neural network is proposed to detect snoring, apnea and silence events during sleep in reference [19], the detection accuracy is 90.65%, 90.99% and 90.30%, respectively.We introduce TCN into sleep apnea task for the first time and achieve 97% accuracy. In addition, we have designed a system architecture that can diagnose OSA in real-time, which is a work that has not been done yet.

Our study has some limitations. First of all, when we collect audio, we do not take into account the impact of the patient's sleeping position on sound. Second, when we build the database, the sound clips are mixed with the recordings of all patients, which may lead to a low generalization ability of the model. Finally, we implement the real-time algorithm on PC, which is difficult to be used in reality. There is still a lot of work worth doing, using more professional equipment to collect audio, improving the generalization ability of classification models, and transplanting real-time algorithms to embedded devices for promotion.

5. CONCLUSION

In this study, an algorithm for real-time diagnosis of OSA is proposed, in which a multi-classification TCN model is applied. The experimental results show that the diagnosis rate of OSA can be as high as 97%, and the OSA-related snoring can be displayed in real-time, so as to determine the time when apnea occurs. The algorithm can achieve high accuracy in the diagnosis of apnea, and is expected to be transplanted to embedded devices for practical promotion and use.

6. REFERENCES

[1] K. Macarthur, C. Ryan, T. Bradley, and H. Alshaer, "Differential Effect of Snoring and Obstructive Sleep Apnea on Sleep Structure and Sleepiness," in *C77. PREDICTORS OF SLEEP DISORDERED BREATHING AND RESPONSE TO TREATMENT*: American Thoracic Society, 2018, pp. A5899-A5899.

[2] J. Sola-Soler, R. Jane, J. A. Fiz, and J. Morera, "Formant frequencies of normal breath sounds of snorers may indicate the risk of Obstructive Sleep Apnea Syndrome," in *2008 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, 2008, pp. 3500-3503: IEEE.

[3] A. Yadollahi and Z. Moussavi, "Formant analysis of breath and snore sounds," in *2009 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, 2009, pp. 2563-2566: IEEE.

[4] A. K. Ng, T. Koh, E. Baey, and K. Puvanendran, "Diagnosis of obstructive sleep apnea using formant features of snore signals," in *World Congress on Medical Physics and Biomedical Engineering 2006*, 2007, pp. 967-970: Springer.

[5] C. Wang, J. Peng, L. Song, X. J. A. p. Zhang, and e. s. i. medicine, "Automatic snoring sounds detection from sleep sounds via multi-features analysis," vol. 40, no. 1, pp. 127-135, 2017.

[6] A. Levartovsky, E. Dafna, Y. Zigel, and A. J. J. o. C. S. M. Tarasiuk, "Breathing and snoring sound characteristics during sleep in adults," vol. 12, no. 3, pp. 375-384, 2016.

[7] A. Sebastian, P. A. Cistulli, G. Cohen, and P. J. a. p. a. de Chazal, "Automatic Classification of OSA related Snoring Signals from Nocturnal Audio Recordings," 2021.

[8] P. Temrat, Y. Jiraraksopakun, A. Bhatranand, and K. Wea-asae, "Suitable feature selection for OSA classification based on snoring sounds," in *2018 15th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology (ECTI-CON)*, 2018, pp. 1-4: IEEE.

[9] A. K. Ng, T. Koh, E. Baey, and K. Puvanendran, "Speech-like analysis of snore signals for the detection of obstructive sleep apnea," in *2006 International Conference on Biomedical and Pharmaceutical Engineering*, 2006, pp. 99-103: IEEE.

[10] A. Sebastian, P. A. Cistulli, G. Cohen, and P. de Chazal, "Characterisation of Upper Airway Collapse in OSA Patients Using Snore Signals: A Cluster Analysis Approach," in *2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, 2020, pp. 5124-5127: IEEE.

[11] E. Dafna, A. Tarasiuk, and Y. J. P. o. Zigel, "Automatic detection of whole night snoring events using non-contact microphone," vol. 8, no. 12, p. e84139, 2013.

[12] K. Qian, Z. Xu, H. Xu, Y. Wu, and Z. J. I. S. P. Zhao, "Automatic detection, segmentation and classification of snore related signals from overnight audio recording," vol. 9, no. 1, pp. 21-29, 2015.

[13] C. Janott, C. Rohrmeier, M. Schmitt, W. Hemmert, and B. Schuller, "Snoring-An Acoustic Definition," in *2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 2019, pp. 3653-3657: IEEE.

[14] R. Nonaka *et al.*, "Automatic snore sound extraction from sleep sound recordings via auditory image modeling," vol. 27, pp. 7-14, 2016.

[15] V. R. Swarnkar, U. R. Abeyratne, and R. V. Sharan, "Automatic picking of snore events from overnight breath sound recordings," in *2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 2017, pp. 2822-2825: IEEE.

[16] L. Cen, Z. L. Yu, T. Kluge, and W. Ser, "Automatic system for obstructive sleep apnea events detection using convolutional neural network," in 2018 40th annual international conference of the IEEE engineering in medicine and biology society (EMBC), 2018, pp. 3975-3978: IEEE.

[17] Z. Zhang *et al.*, "Snore-GANs: Improving automatic snore sound classification with synthesized data," vol. 24, no. 1, pp. 300-310, 2019.

[18] B. Arsenali *et al.*, "Recurrent neural network for classification of snoring and non-snoring sound events," in *2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 2018, pp. 328-331: IEEE.

[19] B. Kang, X. Dang, and R. Wei, "Snoring and apnea detection based on hybrid neural networks," in *2017 International Conference on Orange Technologies (ICOT)*, 2017, pp. 57-60: IEEE.

[20] R. M. Simply, E. Dafna, and Y. Zigel, "Obstructive sleep apnea (OSA) classification using analysis of breathing sounds during speech," in *2018 26th European Signal Processing Conference (EUSIPCO)*, 2018, pp. 1132-1136: IEEE.