Obstructive Sleep Apnea Classification Using Snore Sounds Based on Deep Learning

Apichada Sillaparaya[†], Apichai Bhatranand[†], Chudanat Sudthongkong[††], Kosin Chamnongthai[†], and Yuttapong Jiraraksopakun[†]
[*]Electronic and Telecommunication Engineering Department, Faculty of Engineering, King Mongkut's University of Technology Thonburi, Bangkok, Thailand
E-mail:Apichada.sill@gmail.com, yuttapong.jir@mail.kmutt.ac.th
[††]Medical and Science Media Department, School of Architecture and Design, King Mongkut's University of Technology Thonburi, Bangkok, Thailand

Abstract— **Early screening for the Obstructive Sleep Apnea (OSA), especially the first grade of Apnea-Hypopnea Index (AHI), can reduce risk and improve the effectiveness of timely treatment. The current gold standard technique for OSA diagnosis is Polysomnography (PSG), but the technique must be performed in a specialized laboratory with an expert and requires many sensors attached to a patient. Hence, it is costly and may not be convenient for a self-test by the patient. The characteristic of snore sounds has recently been used to screen the OSA and more likely to identify the abnormality of breathing conditions. Therefore, this study proposes a deep learning model to classify the OSA based on snore sounds. The snore sound data of 5 OSA patients were selected from the opened-source PSG-Audio data by the Sleep Study Unit of the Sismanoglio-Amalia Fleming General Hospital of Athens [1]. 2,439 snoring and breathing-related sound segments were extracted and divided into 3 groups of 1,020 normal snore sounds, 1,185 apnea or hypopnea snore sounds, and 234 non-snore sounds. All sound segments were separated into 60% training, 20% validation, and 20% test sets, respectively. The mean of Mel-Frequency Cepstral Coefficients (MFCC) of a sound segment were computed as the feature inputs of the deep learning model. Three fully connected layers were used in this deep learning model to classify into three groups as (1) normal snore sounds, (2) abnormal (apnea or hypopnea) snore sounds, and (3) non-snore sounds. The result showed that the model was able to correctly classify 85.2459%. Therefore, the model is promising to use snore sounds for screening OSA.**

I. INTRODUCTION

Obstructive Sleep Apnea (OSA) is the most common sleep-related breathing disorder and comprises up to one-seventh of the world's adult population [2]. The consequences of OSA both physically and mentally affect the health condition because insufficient sleep may cause hypersomnia, leading to microsleep and narcolepsy, diabetes mellitus, coronary artery disease, heart attack, ischemic stroke, and depression, etc [3-5]. At present, Polysomnography (PSG) is the gold standard technique to screen sleep apnea [4-6] that identify sleep disorders from the physiological changes of the body signals, i.e. electroencephalogram (EEG), electrocardiogram (ECG), heart rate, eye movement, depth and breathing patterns, snore sounds, blood oxygen levels, and skeletal muscle activity. However, the technique requires a specialized laboratory with an expert and many sensors attached to a patient, hence costly and not convenient for a self-test. Screening OSA using the characteristics of snore sound is then a potential alternative because it is convenient, simple, and useful for subjects to proceed by themselves. Snore sounds can be one of the information to diagnosis OSA because it is directly related to abnormality of breathing conditions caused by obstruction of the upper respiratory tract [4, 5].

Detecting sleep apnea has currently gained significant interest. According to the literature review [7], ECG sensor-based signals are most commonly and widely used for sleep apnea classification, but which sensors or signals are the best is still an active question. On the other hand, obstructive sleep apnea is directly related to snore sounds. Therefore, obstructive sleep apnea classification using snore sounds is also a potential method and interesting for study.

In snore sound classification, the sound characteristics between normal and abnormal snore sounds are different due to different sources of sounds [8-10]. Therefore, the characteristic of snore sounds can be exploited for OSA screening application of the normal and abnormal snore sounds. Feature extraction is an important step before the learning process of sound classification. The feature can be divided into three main groups: the time domain features [11], frequency domain features [8], and time and frequency domain feature or wavelet transform [10]. Among many feature extraction methods used in snore sound classification, the Mel-Frequency Cepstral Coefficients (MFCC) is widely chosen as it provides promising accuracy [12]. In addition, the various classification methods such as Bayes classifier [13], logistic regression [9], AdaBoost classifier [14], Support vectors machines (SVM) [15] can also be used along the sound features to classify the severity of OSA. However, no studies clearly identify which feature extraction and classification method provide the optimum result.

Deep learning-based classification is of great interest to many classification research due to its promising accuracy and ability to adapt to new data. In this study, a deep learning technique to classify OSA using only snore sounds is proposed.

978-616-590-477-3

II. Method

A. *Snore Dataset*

The snore sound data of 5 adult OSA patients (2 female/ 3 male) were selected from the opened-source PSG-Audio data for full-night PSG study by the Sleep Study Unit of the Sismanoglio-Amalia Fleming General Hospital of Athens [1]. The snore sounds were recorded together with the full-night PSG at the sleep disorders laboratory using a microphone placed over the tracheal of the patients with the sampling rate 48 kHz. 2,439 snoring and breathing-related sound segments with an average time duration of 18 seconds were extracted according to the PSG annotations of nasal and respiratory events by the medical team. All segments were divided into 3 groups of 1,020 normal snore sounds, 1,185 apnea or hypopnea snore sounds, and 234 non-snore sounds as shown in Table I. The sound segment data was then separated into 3 groups of 60% training, 20% validation, and 20% test sets for the classification model.

TABLE I
THE NUMBER OF SNORE SOUNDS AND BREATHING-RELATED SOUND SEGMENTS

Group/ Type of Abnormalities	Number of Segments
Hypopnea or Apnea Snore	1,185
Normal Snore	1,020
Non-snore	234
Total	2,439

B. *Data Preprocessing*

Feature extraction was applied to extract the characteristic of snore sounds as data input of the deep learning model. Fig. 1 shows an example of a snore sound segment from the dataset. The Mel-Frequency Cepstral Coefficients (MFCC), a representation of the short-term power spectrum of a sound based on a linear cosine transform of a log power spectrum on a nonlinear Mel-scale of frequency, was computed as the feature of the snore sounds. The number of MFCCs were typically 128 for any short window of a sound segment, hence comprising a 2D array of MFCCs for one sound segment. In this study, the mean of 2D MFCCs was computed to reduce the dimension of the input for the classification model as show in Fig. 2.

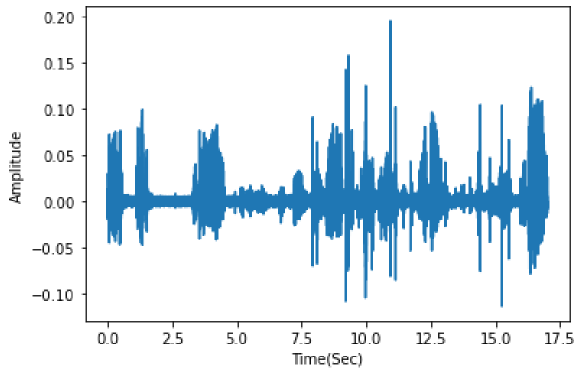

Fig. 1. The Signal Example of A Snore Sound Segment.

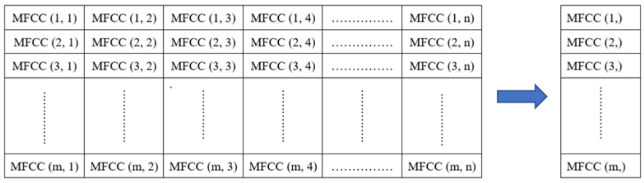

Fig. 2. The Dimension Reduction on MFCCs to Mean MFCCs for Model Inputs.

C. *Deep Learning Classification Model*

A deep learning model for the snore sound classification was constructed by three fully connected layers as shown in Fig. 3. The number of nodes in hidden layers are 100, 200, and 100 nodes, respectively. All hidden layers were added with 50% dropout to prevent the model overfitting. The rectified linear activation function (ReLu) was used as the activation function for all hidden layers. The softmax activation was used for the output layer for a multi-class classification on the three groups of snore sounds as: (1) normal snore sounds, (2) abnormal (apnea or hypopnea) snore sounds, and (3) non-snore sounds.

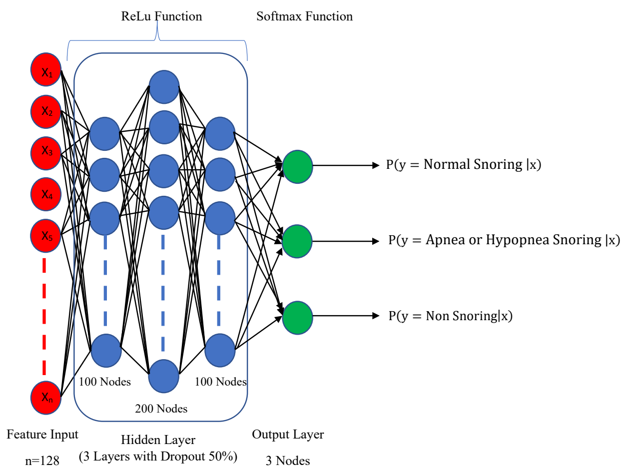

Fig. 3. The Proposed Deep Learning Classification Model.

D. Model Training and Validation

The mean of MFCCs of each sound segment were used as data input for the model with number of batch size=32. The data were randomly separated into 3 groups as 60% training, 20% validation, and 20% test sets, respectively. The learning rate and the number of epochs for training and validation sets was set to 0.001 and 200, respectively. The ADAptive Moment estimation (Adam) optimization algorithm was used for network weight update during the model training. The categorical cross entropy was used as a loss function for the multi-class classification. The accuracy on correct estimation was tracked during the training and also evaluated on the validation set.

E. Classification Evaluation

In this study, the performance of the deep learning model on snore sound classification for OSA screening was evaluated on the test set in terms of accuracy (ACC) and positive predictive value (PPV). The results are shown in terms of confusion matrix comparison between actual and predicted values and shown in terms of true positive (TP), false negative (FN), false positive (FP), true negative (TN), and overall accuracy and PPV. The accuracy and PPV were calculated as:

$$\text{Accuracy} = \frac{\text{TP}}{\text{TP} + \text{FN}} \times 100\% \qquad (1)$$

$$\text{PPV} = \frac{\text{TP}}{\text{TP} + \text{FP}} \times 100\% \qquad (2)$$

III. RESULTS

A. Feature Extraction

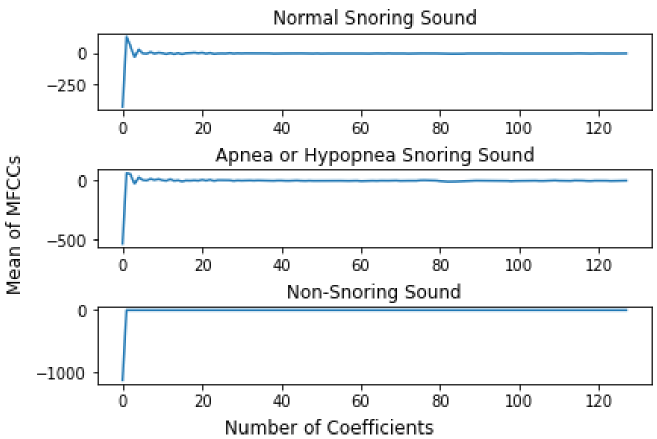

Fig. 4. The Mean MFCCs of Each Group of Snore Sounds.

Fig.4 shows an example of mean MFCC of each snore sound segment type from the dataset. The mean MFCC of normal snore sounds has evidently larger peak difference and stable than the other groups. The peak difference is smaller in the case of apnea or hypopnea snore sounds. No peak difference is clearly observed on the non-snore sounds.

B. Classification Performance

TABLE II
THE CONFUSION MATRIX COMPARISON

		Predicted			
		Normal Snore	Hypopnea or Apnea Snore	Non-snore	Total
Actual	Normal Snore	165	27	0	192
	Hypopnea or Apnea Snore	30	211	7	248
	Non-snore	6	2	40	48
	Total	201	240	47	488

$$\text{Accuracy} = \frac{165 + 211 + 40}{488} \times 100\% = 85.2459\%$$

TABLE III
THE CLASSIFICATION RESULT

Group of Data	TP	FN	FP	TN	ACC (%)	PPV (%)
Normal Snore	165	27	36	260	85.9375	82.0896
Hypopnea or Apnea Snore	211	37	29	211	85.0806	87.9167
Non-snore	40	8	7	433	83.3333	95.1034

Table II shows the confusion matrix comparison between actual and predicted values on the test set of 192 normal snore sounds, 248 apnea or hypopnea snore sounds, and 48 non-snore sounds. The overall accuracy for snore sound classification model is 85.2459% on the test set. The performance of the classification model of each group of snore sounds is summarized in Table III. The classification accuracy on normal snore sounds is 85.9375% and slightly drops to 85.0806% on hypopnea or apnea snore sounds. The accuracy on non- snore sounds apparently drops to 83.3333% possibly due to less number of data relative to the others. On the other hand, the PPV is highest on the non-snore sounds at 95.1034% and apparently reduced on the hypopnea or apnea snore sounds and normal snore sounds at 87.9167% and 82.0896, respectively.

IV. DISCUSSION AND CONCLUSION

In this study, the deep-learning classification based on snore and breathing-related sounds for screening OSA were proposed. The results showed that the snore sound characteristic obtained from the mean MFCCs of three sound groups can be used to classify the snore sounds with promising overall accuracy of 85.2459%. The classification results on each sound group are 85.9375% accuracy and 82.0896% PPV for normal snore sounds, 85.0806% accuracy and 87.9167% PPV for apnea or hypopnea snore sounds, and 83.333% accuracy and 95.1034% PPV for non-snore sounds. The accuracy of the proposed method is also in agreement to the accuracy range of 80-95% as reported in the research community [9, 12, 16]. Yet, it is worthy to note that the accuracy might not be directly compared due to different data sources and accuracy evaluation methods.

The prediction error on each sound group is possibly due to the imbalance of the number of the data in each group. The confusion of classification between the group of normal snore and apnea or hypopnea snore might be from the similarity of the mean MFCCs of both groups. The recent update on the event annotation of the PSG signal from the data source [1] can also affect to the performance of the model. The hyperparameters such as number of learning rate, number of epochs, as well as the number of layers and nodes to be used in the model can be investigated more to improve the classification results.

On the future work, more input data of each sound group for the model with optimized hyperparameters of training is one approach to improve the performance of the model. The balanced number of each snore sound group and updated annotation of the PSG signal on the event of snoring and breathing-related sound should be also considered.

REFERENCES

[1] K. Georgia et al. "PSG-Audio", Science Data Bank, March 14, 2022. Available: http://doi.org/10.11922/sciencedb.00345

[2] M. M. Lyons, N. Y. Bhatt, A. I. Pack, and U. J. Magalang, "Global burden of sleep-disordered breathing and its implications," Respirology, vol. 25, no. 7, pp. 690-702, Jul 2020, doi: 10.1111/resp.13838.

[3] Chulalongkornhospital. "Obstructive sleep apnea" Retrieved from http://chulalongkornhospital.go.th/ sleepcenter

[4] Mayo Clinic Staff. "Obstructive sleep apnea." Retrieved from https://www.mayoclinic.org/diseases-conditions/obstructive-sleep-apnea/symptoms-causes/syc-20352090

[5] N. M. Punjabi, "The epidemiology of adult obstructive sleep apnea," Proceedings of the American Thoracic Society, vol. 5, no. 2, pp. 136-143, 2008.

[6] Easmed. "Sleep Apnea Diagnosis." Retrieved from https://www.easmed.com/sleep-apnea-diagnosis/

[7] S. S. Mostafa, F. Mendonça, A. G. Ravelo-García, and F. Morgado-Dias, "A Systematic Review of Detecting Sleep Apnea Using Deep Learning," Sensors, vol. 19, no. 22, p. 4934, 2019.

[8] B. Calabrese et al., "A System for the Analysis of Snore Signals," Procedia Computer Science, vol. 4, pp. 1101-1108, 2011/01/01/ 2011, doi: 10.1016/j.procs.2011.04.117.

[9] J.-W. Kim et al., "Prediction of Obstructive Sleep Apnea Based on Respiratory Sounds Recorded Between Sleep Onset and Sleep Offset," Clin Exp Otorhinolaryngol, vol. 12, no. 1, pp. 72-78, 2019, doi: 10.21053/ceo.2018.00388.

[10] K. Qian et al., "A Bag of Wavelet Features for Snore Sound Classification," Annals of Biomedical Engineering, vol. 47, no. 4, pp. 1000-1011, 2019/04/01 2019, doi: 10.1007/s10439-019-02217-0.

[11] S. S. Mostafa, F. Mendonça, F. Morgado-Dias, and A. Ravelo-García, "SpO2 based sleep apnea detection using deep learning," in 2017 IEEE 21st International Conference on Intelligent Engineering Systems (INES), 20-23 Oct. 2017 2017, pp. 000091-000096, doi: 10.1109/INES.2017.8118534.

[12] 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, p. 8864863, 2020/12/12 2020, doi: 10.1155/2020/8864863.

[13] J. Solà-Soler, J. A. Fiz, J. Morera, and R. Jané, "Bayes classification of snoring subjects with and without Sleep Apnea Hypopnea Syndrome, using a Kernel method," in 2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society, 30 Aug.-3 Sept. 2011 2011, pp. 6071-6074, doi: 10.1109/IEMBS.2011.6091500.

[14] E. Dafna, A. Tarasiuk, and Y. Zigel, "Automatic detection of whole night snoring events using non-contact microphone," PLoS One, vol. 8, no. 12, p. e84139, 2013, doi: 10.1371/journal.pone.0084139.

[15] J. Sun, X. Hu, C. Chen, S. Peng, and Y. Ma, "Amplitude spectrum trend-based feature for excitation location classification from snore sounds," Physiol Meas, vol. 41, no. 8, p. 085006, Sep 10 2020, doi: 10.1088/1361-6579/abaa34.

[16] T. Kim, J. W. Kim, and K. Lee, "Detection of sleep disordered breathing severity using acoustic biomarker and machine learning techniques," Biomed Eng Online, vol. 17, no. 1, p. 16, Feb 1 2018, doi: 10.1186/s12938-018-0448-x.