Deep Learning Model for OSA Detection using Tracheal Breathing Sounds During Wakefulness

Ali Mohammad Alqudah[1], and Zahra Moussavi[1,2]

[1] Biomedical Engineering, University of Manitoba, Winnipeg, Canada
[2]Electrical and Computer engineering, University of Manitoba, Winnipeg, Canada

Abstract — The detection of Obstructive Sleep Apnea (OSA) during sleep is a simple and well-established technique; however, its detection during wakefulness is challenging. In this paper, we propose a deep learning model for the detection of OSA using only the tracheal breathing sounds spectrum as input. We employed our team's previous dataset consisting of 109 subjects as non or mild-OSA with apnea/hypopnea index (AHI) < 15 and 90 subjects as OSA with AHI ≥ 15. All study subjects were referred to overnight polysomnography (PSG) to determine their AHI values. Tracheal breathing sounds were recorded in the supine position before proceeding to PSG while awake. The recording protocol was to have 5 deep breaths first through the mouth and then 5 deep breaths through the nose. Data were normalized and segmented into inspiratory/expiratory breathing phases; their power spectra were then calculated and fed to a deep learning model consisting of 71 layers. The results of 10 K-fold show that the proposed deep learning model achieved an accuracy of 74.9%, sensitivity of 76.1%, and specificity of 73.3%. Although these results are not as high as the previously reported analyses, they can be improved significantly by combining with anthropometric parameters and subgrouping subjects based on their age, weight, etc. This work aimed to show the potential of deep learning on this dataset despite the limited sample size. The results are encouraging to continue and improve the algorithm.

Keywords — Obstructive Sleep Apnea (OSA), Deep Learning, Tracheal Breathing Sounds, Power Spectrum.

I. Introduction

A frequent syndrome known as obstructive sleep apnea (OSA) disorder is defined by recurrent episodes of the whole (apnea) or partial (hypopnea) pharyngeal collapse while sleeping. The affected people's quality of life may suffer significantly as a result of OSA [1]. The risks of developing heart disease, high blood pressure, stroke, depression, diabetes, headaches, and traffic accidents are among many associated effects of OSA [2].

OSA is also known to increase the risk of preoperative morbidity and/or postoperative mortality [ref]; thus, identifying individuals with OSA before a surgery that requires general anesthesia would lower the post-operative mortality risks [3].

The apnea/hypopnea events index (AHI) per hour is used to gauge the severity of sleep apnea. Polysomnography (PSG) is the gold standard for diagnosing OSA. However, the expensive nature of PSG due to its long procedure and need for trained technicians, lack of sleep labs in small towns, etc., there is typically a large waiting list for PSG assessments [4].

The STOP-BANG questionnaire is one of the various subjective OSA diagnosis/screening instruments that doctors employ for quick screening of OSA when there is no possibility of performing PSG overnight [5]. However, its specificity is very low (10%). As a result, such subjective screening is not reliable for quick decision-making.

Thus, there is a great need for an improved method of diagnosing and screening OSA individuals during wakefulness as a fast and robust screening tool. Such an approach will aid in reducing the need for PSG and will significantly help with decision-making prior to full anesthesia in regard to potential complications due to OSA in a patient [5].

To screen for OSA during wakefulness, several research teams around the world have investigated objective alternatives to current procedures by using tracheal breathing sound analysis recorded while individuals were awake [6]–[10] and also using voice analysis [11], [12] with promising results.

A major shortcoming of the previous research on this topic is the dependency of the algorithms on pre-processing and feature selection; the automated feature extraction from the sound data using deep learning models was not explored [13].

Deep-learning techniques typically need large samples for learning. In this work, despite the limited sample size, our main objective was to investigate the potential of deep learning in extracting characteristic features sensitive to AHI without the need to know the type of extracted respiratory phase (either inspiration or expiration) and lengthy pre-processing.

II. Methodology

A. Dataset

Data were adopted from our team's previous work [6]. Each individual's tracheal breathing sounds were recorded by a microphone positioned at the suprasternal notch of the trachea for 5 complete cycles of deep breathing through the nose with the mouth closed and then another 5 breaths through the mouth while wearing a nasal clip using.

The recording was made while the subject was in a supine position and awake. In this study and like our previous works the Data (n=199) were grouped as mild/non-OSA (n=109, AHI< 15) and OSA (n=90, AHI ≥ 15) [6]. Table 1 displays the total number of subjects and the total number of breathing phases (both inspiration and expiration) for each class for the dataset used in this study.

Table 1 Dataset Distribution

Class	AHI	Number of Subjects	Total Breathing Phases
non-OSA	AHI< 15	109	1840
OSA	AHI ≥ 15	90	1496

B. Preprocessing and Power Spectrum

A Butterworth band-pass filter of order 4 with a cutoff frequency of 75- 3000 Hz was first used to filter out each breathing sound phase independently in order to cancel out the impacts of different interferences such as heartbeats, background noise, etc. [7]. Second, in order to eliminate the impact of plausible airflow fluctuation between breathing cycles, each filtered signal was normalized by its variance envelope using a smoothed copy by applying a Moving Average window with a size of 64 samples sequence, then by its standard deviation (energy). After that, a 50% duration around each breathing phase's maximum was chosen for further analysis using the logarithm of the variance of the phase signals, representing the respiratory flow [7]. This duration roughly equates to the upper 40% of each breathing phase's respiratory airflow, where the signal is considered stationary within the breathing sound signal. The power spectrum density (PSD) for all middle regions of the signal was computed using the welch method with 25 ms length (256 samples window size), 50% overlapping between adjacent windows [7]. These preprocessing techniques are the same techniques used in our team's previously published works using machine learning methods.

C. Deep Learning Model

Deep learning is a recently developed technique in the artificial intelligence field, which is emerged as a response to the increased number of available and recorded large datasets [14]. Deep learning models can be distinguished and defined based on their unique design and development of the set of layers, usually called architecture. Such an architecture consists of a set of consecutive layers to process the model input and get the output [15]. This paper's proposed deep learning model is based on a combination of Convolutional Neural Network (CNN) with Long Short-Term Memory (LSTM).

The model consists of 71 layers where the input of the model is the 1D power spectrum of the breathing phase. Then the CNN part of the model will extract the deep features from the input data, the layers are the convolutional and the max pooling layers. The model takes the maximum and the average power spectrum by passing it through max pooling and average pooling layers, respectively. Then, both variables over the frequency band between 1 and 3700 Hz are passed through three stacked residual blocks to efficiently extract features. The output features are then passed through a series of fully connected and scaling layers to make the model "structure agnostic" about the input. Moreover, the LSTM part will use the input data characteristics as frequency-domain-based features to learn to extract the most significant part of the frequency data. Using the combination of two models allows shallower models with fewer parameters to be developed, resulting in better performance. Figure 1 shows the developed models and Table 2 shows the layer details.

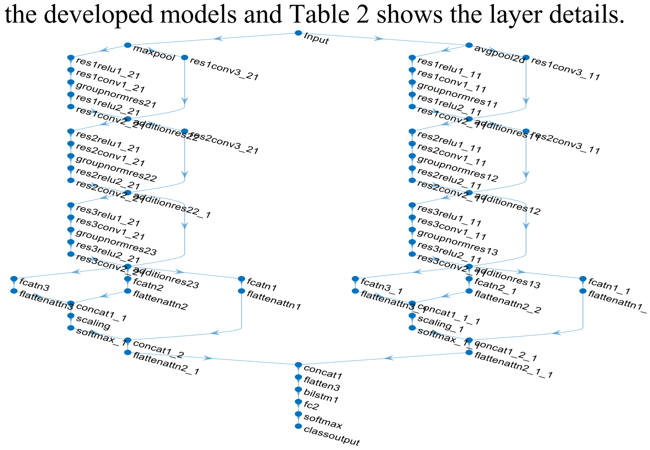


Fig. 1 The proposed deep learning model architecture.

III. Results

The deep learning model in this work was created by combining the CNN and LSTM models. All of the trials were carried out using a desktop computer with an Intel Core i7-12700H/2.3 GHz CPU, 32 GB of RAM, a 1 TB hard drive (HDD), and a 16 GB Nvidia GeForce RTX 3080 Ti GPU.

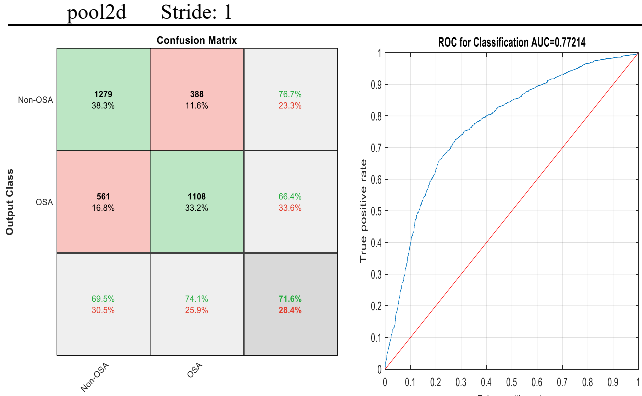
Our proposed model was tested using the 10-Kfold methodology, and one of the nine training folds was employed as validation during the cross-validation procedure; for each fold, the training was done using the SGDM optimizer with a cross-entropy loss function, initial learning rate of 0.001, and mini-batch size of 25. These hyperparameters are selected using trial and error. Since there is a high correlation between breathing sounds of ins/exp phases of the same subject, the cross-validation classification process in this research was done based on each subject; where all power spectrum data related to the same subject's breathing phases are either in the training set, validation set, or testing set.

A two-step classification using the deep learning model was used to detect OSA. In the first step, called "phase-based", the deep learning model was used to classify each breathing phase (either inspiration or expiration) into non-OSA or OSA groups. The outcomes of this step are representative of the ability of deep learning to detect OSA using phases regardless of knowing if they are inspiratory or expiratory phases. Figure 2 shows the confusion matrix and the ROC for this step.

In the second step, called "subject-based" detection, a final decision for each subject is generated by finding the most frequent class using a majority voting classifications technique among all breathing phase related to the same subject found in the first step. The outcomes of this step are representative of the ability of the proposed deep learning model to screen for OSA subjects. Figure 3 shows the confusion matrix and the ROC for the second step.

Table 2 The proposed architecture layer details.

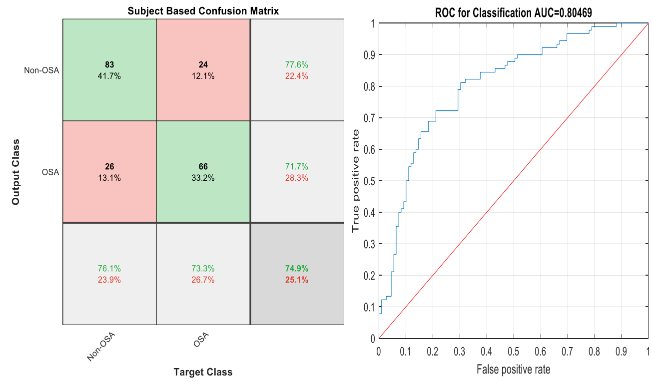
#	Name	Parameters	#	Name	Parameters
1	Input	Length: 3700	36	res1conv1_11	Filter: 8 *3,3 Stride: 1
2	maxpool	Pool Size: 5,5 Stride: 1	37	group-normres11	Channels: 8
4	res1conv1_21	Filter: 8*8, 8 Stride: 1	39	res1conv2_11	Filter: 8* 3,3 Stride: 1
5	group-normres21		40	res1conv3_11	Filter: 8*3,3 Stride: 1
7	res1conv2_21	Filter: 8*8, 8 Stride: 1	43	res2conv1_11	Filter: 8*3,3 Stride: 1
8	res1conv3_21	Filter: 8* 8, 8 Stride: 1	44	group-normres12	Channels: 8
11	res2conv1_21	Filter: 8* 8, 8 Stride: 1	46	res2conv2_11	Filter: 8*3,3 Stride: 1
12	group-normres22	Channels: 8	47	res2conv3_11	Filter: 8*3,3 8 Stride: 1
14	res2conv2_21	Filter: 8*8, 8 Stride: 1	50	res3conv1_11	Filter: 8*3,3 Stride: 1
15	res2conv3_21	Filter: 8*8, 8 Stride: 1	51	group-normres13	Channels: 8
18	res3conv1_21	Filter: 8*8, 8 Stride: 1	53	res3conv2_11	Filter: 8*3,3 Stride: 1
19	group-normres23	Channels: 8	55	fcatn3_1	Output: 512
21	res3conv2_21	Filter: 8*8, 8 Stride: 1	57	fcatn2_1	Output: 512
23	fcatn3	Output: 512	59	fcatn1_1	Output: 512
25	fcatn2	Output: 512	68	bilstm1	Num Hidden Units: 512
27	fcatn1	Output: 512	69	fc2	Output: 2
34	avg-pool2d	PoolSize: 5 5 Stride: 1			



(A) (B)

Fig. 2 The output of Phases-based training; A) Confusion matrix, B) ROC.

The performance metrics of both approaches (phase and subject-based) are shown in Table 3. These parameters are calculated based on the confusion matrix and ROC.



(A) (B)

Fig. 3 The output of Subject-based training; A) Confusion matrix, B) ROC.

Table 3 Performance of the Proposed Methodology.

	Accuracy	Sensitivity	Specificity	Precision	AUC
Phase-Based	71.55	69.51	74.06	76.72	0.7721
Subject-Based	74.87	76.15	73.33	77.57	0.8047

IV. Discussion

Employing tracheal breathing sounds in order to detect OSA during sleep is easy and well-established. However, using breathing sounds to detect OSA during wakefulness is very challenging as OSA individuals do not exhibit any breathing symptoms while awake [16]. While there have been research studies to use tracheal breathing sounds for screening OSA during [6]–[8], [17], this research presents the first attempt to apply deep-learning models to the tracheal breathing sounds recorded during wakefulness for screening OSA.

Out of the two classifications proposed in this study (phase-base and subject-base), the accuracy of the phase-based classification was found to be low (~71%). This is due to the shallow deep learning model, where the development of deep models requires more time for optimization and ablation studies for the effect of adding any new blocks such as attention. Moreover, the outcomes of this classification were fed to a majority vote classifier to have a subject-based classification based on the class of the majority of the phases of a subject; thus, the low accuracy of the phase classicization projects into the subsequent subject-base classification although by majority voting of the phases' class for each subject the overall accuracy is increased to about 75% (Table 3).

Comparing the performance values of the proposed deep learning model with these previous works that used the same dataset [6], [7], [17] shows that the proposed methodology has lower performance, and still needs more development and enhancement over a large dataset to increase its overall performance. At the same time, the results are encouraging given the limited sample size. Table 4 shows a comparison between the proposed work and previous similar studies.

Table 4 A Comparison Between Proposed Work and Previous Works.

Ref	AHI Threshold	Accuracy	Sensitivity	Specificity
[7]	AHI ≤ 5 AHI ≥10	83.92	82.61	85.22
[13]	AHI <15 & AHI ≥15	81.4	80.9	82.1
Phase-Based		71.55	69.51	74.06
Subject-Based		74.87	76.15	73.33

Furthermore, more experiments are required to improve the proposed model's performance and stability. This could include but is not limited to using different initialization techniques, hyperparameters fine-tuning, employing a larger dataset, and combining the extracted deep features with anthropometric features as input to other classifiers, as well as subgrouping subjects based on their age, weight, etc.

V. Conclusion

This paper presents the outcomes of a deep learning-based detection of OSA using the breathing sounds power spectra, recorded during wakefulness, as input. While the accuracy of the deep-learning model for screening OSA is lower than the previous methods using the same dataset, overall, the proposed method has the advantage of removing the need for knowing the respiratory phase as either inspiration or expiration. Given the limited sample size used in this study and yet achieving reasonable accuracy, the results encourage investigating the use of deep learning in the detection of OSA; we anticipate its accuracy to increase in a larger dataset.

References

[1] H. R. Colten and B. M. Altevogt, “Sleep Disorders and Sleep Deprivation: An Unmet Public Health Problem,” Sleep Disorders and Sleep Deprivation: An Unmet Public Health Problem, pp. 1–404, Apr. 2006, doi: 10.17226/11617.

[2] L. D. Rosenthal and D. C. Dolan, “The Epworth Sleepiness Scale in the Identification of Obstructive Sleep Apnea,” J Nerv Ment Dis, vol. 196, no. 5, pp. 429–431, 2008, doi: 10.1097/NMD.0b013e31816ff3bf.

[3] “Practice Guidelines for the Perioperative Management of Patients with Obstructive Sleep ApneaAn Updated Report by the American Society of Anesthesiologists Task Force on Perioperative Management of Patients with Obstructive Sleep Apnea,” Anesthesiology, vol. 120, no. 2, pp. 268–286, Feb. 2014, doi: 10.1097/ALN.0000000000000053.

[4] C. A. Kushida et al., “Practice Parameters for the Indications for Polysomnography and Related Procedures: An Update for 2005,” Sleep, vol. 28, no. 4, pp. 499–523, Apr. 2005, doi: 10.1093/SLEEP/28.4.499.

[5] M. Nagappa et al., “Validation of the stop-bang questionnaire as a screening tool for obstructive sleep apnea among different populations: A systematic review and meta-Analysis,” PLoS One, vol. 10, no. 12, 2015, doi: 10.1371/journal.pone.0143697.

[6] Z. Moussavi and A. Elwali, “System and Methods for Screening Obstructive Sleep Apnea During Wakefulness Using Anthropometric Information and Tracheal Breathing Sounds.” Google Patents, Dec. 2021.

[7] A. Elwali and Z. Moussavi, “Obstructive Sleep Apnea Screening and Airway Structure Characterization During Wakefulness Using Tracheal Breathing Sounds,” Ann Biomed Eng, 2016, doi: 10.1007/s10439-016-1720-5.

[8] J. Solà-Soler, J. A. Fiz, A. Torres, and R. Jané, “Identification of obstructive sleep apnea patients from tracheal breath sound analysis during wakefulness in polysomnographic studies,” in Engineering in Medicine and Biology Society (EMBC), 2014 36th Annual International Conference of the IEEE, 2014, pp. 4232–4235.

[9] J. Solà-Soler, R. Jané, J. A. Fiz, and J. Morera, “Formant frequencies of normal breath sounds of snorers may indicate the risk of obstructive sleep apnea syndrome,” Proceedings of the 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS’08 - “Personalized Healthcare through Technology,” pp. 3500–3503, 2008, doi: 10.1109/IEMBS.2008.4649960.

[10] R. M. Simply, E. Dafna, and Y. Zigel, “Obstructive Sleep Apnea (OSA) classification using analysis of breathing sounds during speech,” European Signal Processing Conference, vol. 2018-September, pp. 1132–1136, Nov. 2018, doi: 10.23919/EUSIPCO.2018.8553353.

[11] J. Solé-Casals et al., “Detection of severe obstructive sleep apnea through voice analysis,” Appl Soft Comput, vol. 23, pp. 346–354, Oct. 2014, doi: 10.1016/J.ASOC.2014.06.017.

[12] R. F. Pozo, J. L. B. Murillo, L. H. Gmez, E. L. Gonzalo, J. A. Ramírez, and D. T. Toledano, “Assessment of Severe Apnoea through Voice Analysis, Automatic Speech, and Speaker Recognition Techniques,” EURASIP Journal on Advances in Signal Processing 2009 2009:1, vol. 2009, no. 1, pp. 1–11, Jun. 2009, doi: 10.1155/2009/982531.

[13] A. Elwali and Z. Moussavi, “A novel Decision Making procedure during Wakefulness for Screening obstructive Sleep Apnea using Anthropometric information and tracheal Breathing Sounds,” Sci Rep, no. June 2018, pp. 1–12, 2017, doi: 10.1038/s41598-019-47998-5.

[14] Y. Bengio, D.-H. Lee, J. Bornschein, T. Mesnard, and Z. Lin, “Towards Biologically Plausible Deep Learning,” Feb. 2015, doi: 10.48550/arxiv.1502.04156.

[15] H. Schulz and S. Behnke, “Deep Learning: Layer-Wise Learning of Feature Hierarchies,” KI - Kunstliche Intelligenz, vol. 26, no. 4, pp. 357–363, Nov. 2012, doi: 10.1007/S13218-012-0198-Z/FIGURES/4.

[16] M. Younes, “Role of respiratory control mechanisms in the pathogenesis of obstructive sleep disorders,” J Appl Physiol, vol. 105, no. 5, pp. 1389–1405, Nov. 2008, doi: 10.1152/JAPPLPHYSIOL.90408.2008/ASSET/IMAGES/LARGE/ZDG0110882170007.JPEG.

[17] Z. Moussavi and A. Elwali, “System and Methods for Screening Obstructive Sleep Apnea During Wakefulness Using Anthropometric Information and Tracheal Breathing Sounds.” Google Patents, Dec. 2021.