



Regularized logistic regression for obstructive sleep apnea screening during wakefulness using daytime tracheal breathing sounds and anthropometric information

Farahnaz Hajipour¹ · Mohammad Jafari Jozani² · Ahmed Elwali¹ · Zahra Moussavi^{1,3}

Received: 21 February 2019 / Accepted: 27 September 2019 / Published online: 6 November 2019
© International Federation for Medical and Biological Engineering 2019

Abstract

Obstructive sleep apnea (OSA) is a prevalent health problem. Developing a technology for quick OSA screening is momentous. In this study, we used regularized logistic regression to predict the OSA severity level of 199 individuals (116 males) with apnea/hypopnea index (AHI) ≥ 15 (moderate/severe OSA) and AHI < 5 (non-OSA) using their tracheal breathing sounds (TBS) recorded during daytime, while they were awake. The participants were guided to breathe through their nose, and then through their mouth at their deep breathing rate. The least absolute shrinkage and selection operator (LASSO) feature selection approach was used to select the discriminative features from the power spectra of the TBS and the anthropometric information. Using a five-fold cross-validation procedure, five different training sets and their corresponding blind-testing sets were formed. The average blind-testing classification accuracy over the five different folds was found to be $79.3\% \pm 6.1$ with the sensitivity (specificity) of $82.2\% \pm 7.2\%$ ($75.8\% \pm 9.9\%$). The accuracy for the entire dataset was found to be 81.1% with sensitivity (specificity) of 84.4% (77.0%). The feature selection and classification procedures were intelligible and fast. The selected features were physiologically meaningful. Overall, the results show that TBS analysis can be used as a quick and reliable prediction of the presence and severity of OSA during wakefulness without a sleep study.

Keywords Obstructive sleep apnea · Tracheal breathing sounds · Wakefulness screening · Regularized logistic regression · LASSO feature selection

1 Introduction

Obstructive sleep apnea-hypopnea disorder (OSA) is characterized by repetitive narrowing or complete closure of the upper airway (UA) during sleep, leading to complete cessation (apnea) and/or partially ($\geq 50\%$) reduction (hypopnea) of airflow that lasts at least 10 s and is associated with a minimum of 4% drop in oxygen saturation level of blood (S_aO_2) [1]. The apneic episodes often contribute to frequent arousals or awakenings in order to restore airway functioning. These frequent

arousals reduce sleep continuity and quality [2]. Untreated OSA is associated with many deficits, including excessive daytime sleepiness, increased risk of motor vehicle accidents [1], and increased risk of cardiovascular disease [3].

OSA is a common health problem which affects all age groups. Between 9 and 38% of the general adult, the population is suffering from OSA, while it is higher in men compared to women and much higher in the elderly groups [4]. Since OSA is still underdiagnosed, these statistics are believed to underestimate the actual numbers [5]. The severity of OSA is currently measured by the number of apneic episodes per hour of sleep using the apnea/hypopnea index (AHI). AHI < 5 considered as non-OSA, and $5 \leq \text{AHI} < 15$, $15 \leq \text{AHI} < 30$, AHI ≥ 30 considered as mild, moderate, and severe OSA, respectively [6]. AHI is measured by polysomnography (PSG) assessment over the night sleep. PSG records various signals including electroencephalogram, electrocardiogram, electrooculogram, electromyography of chin and leg, body position, nasal airflow, S_aO_2 , as well as abdominal and thoracic movements in order to provide a full assessment of sleep

✉ Zahra Moussavi
Zahra.Moussavi@umanitoba.ca

¹ Biomedical Engineering Program, University of Manitoba, Winnipeg, Canada

² Dept of Statistics, University of Manitoba, Winnipeg, Canada

³ Electrical and Computer Engineering Department, University of Manitoba, Winnipeg, Canada

quality [7]. While full-nocturnal PSG is considered as the gold standard for sleep apnea assessment, it is time-consuming, laborious, expensive, and not easily accessible, particularly in small towns and remote areas. Thus, in emergency situations that the OSA status of a patient is needed, it might not be feasible to perform a quick assessment of OSA status using PSG approach, hence underdiagnose OSA. For example, around 80% of patients presenting for operation are undiagnosed at the time of surgery [5]. Inadequate preoperative assessment of OSA in these patients may increase their postoperative complications risks. Therefore, a quick and reliable screening OSA assessment during daytime without a full-nocturnal sleep study is highly desirable.

The current quick preoperative OSA screening method in hospitals is based on using subjective questionnaires, such as Berlin Questionnaire, Epworth sleepiness scale, STOP, and STOP-BANG [8], that collect anthropometric information (gender, neck circumference, age) [9]. These methods have shown to have a high sensitivity (90%) but at the cost of a very low specificity (< 40%) [8]. Consequently, they could potentially result in identifying a much higher number of participants as high-risk patients and reducing the cost-effectiveness of such assessment. On the other hand, our team and a few other researchers around the world have used tracheal breathing sounds (TBS) for OSA monitoring and screening during daytime to predict their OSA condition with better accuracy than the commonly used method based on questionnaires [10–12]. The premise of our technique to screen OSA during wakefulness is based on the fact that the TBS change as the UA structure varies in OSA individuals [13, 14]. It has been shown that these structural changes are present not only during sleep [15, 16] but also during daytime when patients are awake (wakefulness) [17–20].

In a study reported in [12], the formant frequencies of TBS and their variation from inspiration to expiration were investigated for 10 mild/moderate (AHI < 30) and 13 severe (AHI \geq 30) OSA patients during wakefulness before getting asleep. They used an LDA classification approach using formant features and some anthropometric information. They reported sensitivity (specificity) of 88.9% (84.6%) and accuracy of 86.4%. Since the features were extracted from the entire dataset (due to the small sample size), those results are considered as biased. Also, LDA is a generative approach that makes strong and sometimes unrealistic assumptions such as normality as well as equal variance structure of all extracted features for each class. In our team's previous study [10], many features were extracted from the TBS' power spectra of 105 randomly selected participants (56 non-OSA, AHI \leq 5 and 49 OSA, AHI \geq 10) out of a total of 130 participants. Next, using data of all participants and within an exhaustive leave-two-out routine (by leaving one subject from each of the two groups as testing and the remaining as training), six features with the best discriminative power were selected.

Afterward, using a support vector machine (SVM) classifier with linear kernel, the classification accuracy of all combination of three-feature sets were calculated. This procedure was repeated until each participant's data was used at least once as a test. Finally, from the most repeated 3 features, 2 of them with the lowest correlation were selected. Then, through another exhaustive leave-two-out cross-validation, the 2-class SVM classifier resulted in 83.83% and 83.92% accuracies for training and test datasets, respectively. The sensitivity (specificity) was 83.92% (85.20%) for training and 82.61% (85.22%) for test datasets [10]. Despite the high accuracy of the classification results, the proposed procedure in [10] is complicated and computationally expensive. Moreover, the random selection of training participants for feature extraction was done only once; thus, the reported results might have some bias.

In this paper, we propose to apply the logistic regression with least absolute shrinkage and selection operator (LASSO) or l_1 -penalty for classification of apneic individuals using the combination of the anthropometric information and TBS during daytime while the subjects are awake. LASSO is a powerful regularization method that performs the feature selection for statistical models [21] by shrinking and sometimes setting some of the coefficients of the regression variables to zero [21, 22]. Removing some of the coefficients can reduce the variance without a substantial increase of the bias, hence, increase the prediction accuracy. Moreover, by eliminating the irrelevant variables that are not associated with the response variable, the LASSO helps to increase the model interpretability and reduce the overfitting [21, 22]. This method has been widely used in variable selection and classification in many clinical fields [23–25]. We also use LASSO logistic regression for our classification problem, as it is a discriminant approach that does not require normality assumption or homogeneity of features in OSA classes. To the best of our knowledge, this is the first research where these techniques are used to assist in OSA diagnosis from TBS signals recorded during the daytime. We have validated our approach on a larger number of participants compared to that of previous studies. The physiological interpretation of the selected features in relation to the structural changes of UA due to OSA are also discussed.

2 Method

2.1 Data

Data of this study were collected from 199 participants suspected of OSA, prior to overnight PSG assessment at Sleep Disorder Lab at Misericordia Health Centre (Winnipeg, MB, Canada). The Biomedical Research Ethics Board of the University of Manitoba approved the study, and all participants signed an informed consent form prior to

participating in the study. After our recording, participants proceeded to overnight PSG preparation and sleep assessment. We collected their anthropometric information and calculated their AHI (from PSG assessment) afterward to compare with our proposed method's AHI prediction and calculate its accuracy. Our collected data included 74 (29 males) non-OSA with $AHI < 5$, and 90 (66 males) moderate/severe OSA with $AHI \geq 15$. The remaining 35 (21 males) mild-OSA participants with $5 \leq AHI < 15$ were dealt separately. Table 1 presents the anthropometric information of the participants.

2.2 Recording procedure

TBS data were recorded during the daytime, while the participants were awake, and prior to their PSG assessment. Participants' TBS were recorded by a miniature microphone (Sony ECM-77B) embedded in a chamber with a 2-mm cone-shaped space with skin placed over the suprasternal notch of trachea using a double-sided adhesive ring tape. The chamber was embedded in a soft neckband wrapped around the participants' neck to prevent plausible movement and comfort of the participant. The recorded breathing sounds were amplified using a Biopac (DA100C) amplifier, band-pass filtered in the frequency range of (0.05–5000 Hz), and digitized at 10240 Hz sampling rate. In supine position with head resting on a pillow, the participants were instructed to breathe deeply in two maneuvers: first through their mouth with a nose clip in place, second through their nose with their mouth closed. and then we recorded 5 full breath cycles in each maneuver.

2.3 Pre-processing and signal analysis

All of our recordings were done in a hospital setting in a relatively noisy background. Thus, the recorded sounds contained various types of noises (e.g., vocal noises, air conditioner sounds) in addition to the desired breathing sounds. We applied different processes to remove unwanted noises

from the collected TBS. As the first step, we examined the recorded sounds in the time-and-frequency domain by visual and auditory means using the spectrogram program of Matlab™ software; this step was done manually to ensure 100% accuracy in selecting noise-free breathing sounds segments. Consequently, all of the breathing cycles that included artifacts, vocal noises, swallowing, and low signal to noise ratio compared to the background noise were excluded from the analysis. Figure 1 shows a sample spectrogram that contains breathing sounds in addition to various type of noises including swallowing, coughing, vocal noises, and the microphone movement sounds due to either fat or loose tissues of the neck touching the microphone.

As the second step of noise reduction, we used a 5th-order Butterworth band-pass filter in the range of (75–2500 Hz) to keep the main frequency components of the breathing phases of each selected sound and to suppress the effect of low and high-frequency noises (including heart sounds that are less than 75 Hz [26], neck muscle artefact, fundamental frequency of the power line (60 Hz in Canada), and background noises). The reason for choosing the 2500 Hz as the high-frequency bound is that most of the TBS energy is claimed to be below 2500 Hz [27]. Next, similar to our previous studies [10, 28], we normalized the filtered sounds first by their variance envelope (64-sample sequence moving average filter of the signal), and secondly, by their energy to compensate for probable different flow rates in different breathing cycles.

In this study, we did not record the respiratory flow of participants, however, to ensure the respiratory phases all recording procedures started at the inspiration phase and marked by the voice of the experimenter. Using that auditory marker, the inspiratory and expiratory phases were separated manually and analyzed independently. Then, we used the method introduced in [14] to determine the approximately stationary portion of each normalized sound. In brief, the logarithm of the sound's variance (LogVar) was calculated, and the sounds segments corresponding to the middle part (50% duration

Table 1 Anthropometric information's mean and their corresponding standard deviations (std) for the non-OSA, moderate/severe OSA, and mild-OSA groups

	Non-OSA (AHI < 5, n = 74)	Moderate/severe OSA (AHI ≥ 15, n = 90)	Mild-OSA (5 ≤ AHI < 15, n = 35)
AHI ± std	1.2 ± 1.3	42.8 ± 32.7	8.7 ± 2.6
NC (cm) ± std	38.7 ± 4.1	44.1 ± 3.7	42.2 ± 6.3
Age (year) ± std	46.8 ± 12.9	52.2 ± 11.6	52.3 ± 11.7
BMI (kg/m ²) ± std	30.6 ± 6.3	36.4 ± 8.0	34.3 ± 8.4
Sex (male:female)	29:45	66:24	21:14
Height (cm) ± std	167.5 ± 9.4	170 ± 10.3	170.6 ± 9.9
Weight (kg) ± std	85.1 ± 19.3	106 ± 23.2	188.8 ± 20.8
MP (I-II-III-IV)	41-19-6-8	22-30-22-16	19-6-9-1

AHI is apnea/hypopnea index, NC is neck circumference, BMI is body mass index, and MP is Mallampati score and the numbers in that row show the number of subjects with MP level I, II, III, and IV

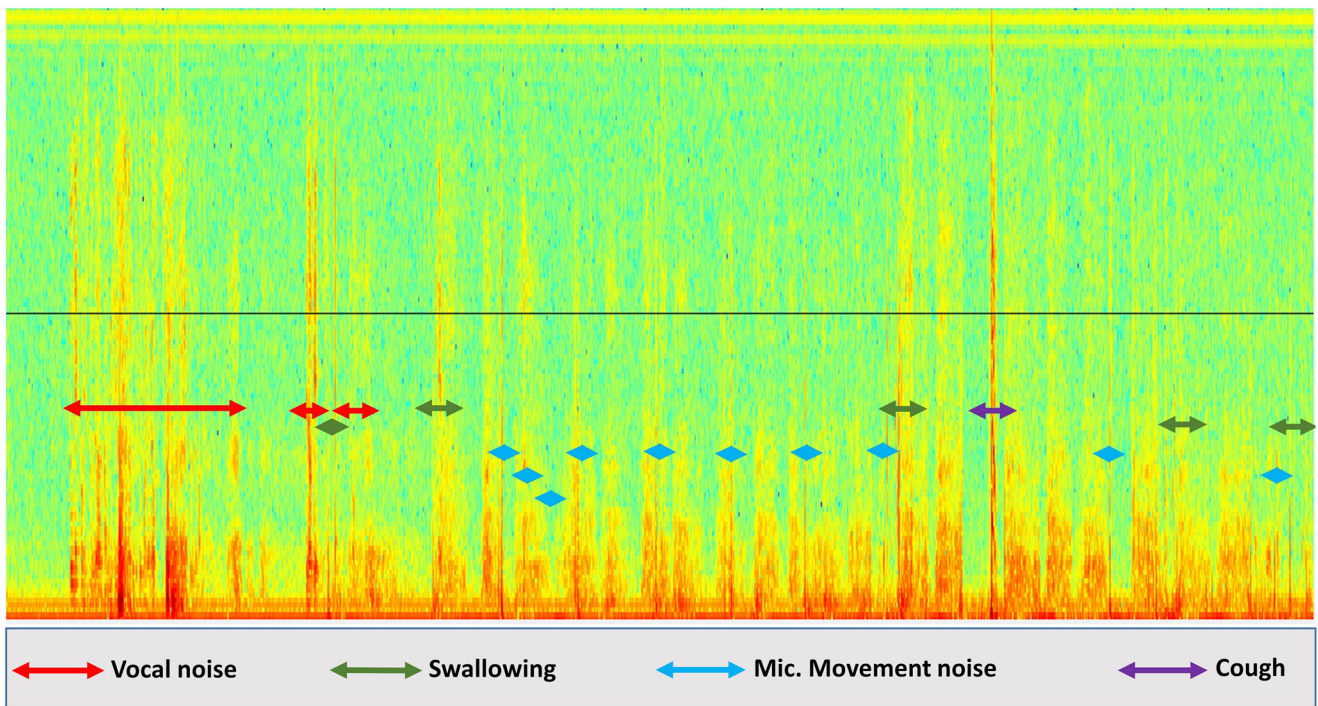


Fig. 1 Spectrogram of a sample recording with vocal noises, swallowing, coughing, and microphone movement sounds due to attachment of the microphone to the participant's skin

around the maximum) was extracted as the stationary portion of each breathing phase. Finally, we estimated the power spectrum density (PSD) of the stationary portion of each TBS signal using Welch's method [29] in windows of 205-point (~20 ms) with 50% overlap between the successive windows. The values of optimum window size and overlap for segmenting the TBS signal were selected based on the results of our previous study [28] and our various pilot studies that showed these values estimates the mostly differentiable power spectrum density (PSD) signals between the non-OSA and OSA subjects.

Following the above routine, for each participant, we estimated the PSD in both linear and logarithmic scale for TBS recorded in each mouth and nasal maneuvers. The PSD estimation was done for each inspiratory/expiratory phase separately and also for their combination (summation/subtraction). This routine resulted in 16 different PSD signals for each participant; they were then averaged over the breathing cycles of each participant's data. Figure 2 outlines the abovementioned pre-processing and signal analysis.

2.4 Feature extraction

We used the 5-fold cross-validation routine to randomly split the set of 164 participants associated with the non-OSA and

moderate/severe OSA groups into five non-overlapping groups (folds), each consisting of approximately 33 participants (20% non-OSA and 20% moderate/severe OSA). Each time, participants' TBS data of one fold were left out and considered as blind-testing data, and data of the remaining four folds (131 participants) were considered as training data and used for feature extraction. The test error was estimated by averaging the five resulting error estimates. Data of the mild-OSA group ($5 < \text{AHI} < 15$, $n = 35$) were not used during feature selection, and they were dealt separately as described later in Section 2.5.

As mentioned in Section 2.3, within each training set, we estimated the PSD for all participants separately. In order to extract the characteristic features between the mild and moderate/severe OSA groups of this study, we separately averaged the PSD signals of mild and moderate groups within each training dataset and calculated their 95% confidence intervals (CI). The 95% CI for either mild or moderate/severe OSA groups were calculated according to the Eq. 1 as follows:

$$\text{CI} = \text{PSD} \pm \text{Critical_value} \times \frac{\text{std}(\text{PSD})}{\sqrt{n}} \quad (1)$$

where PSD is the average of PSD for each OSA group, std is the standard deviation, n is the number of

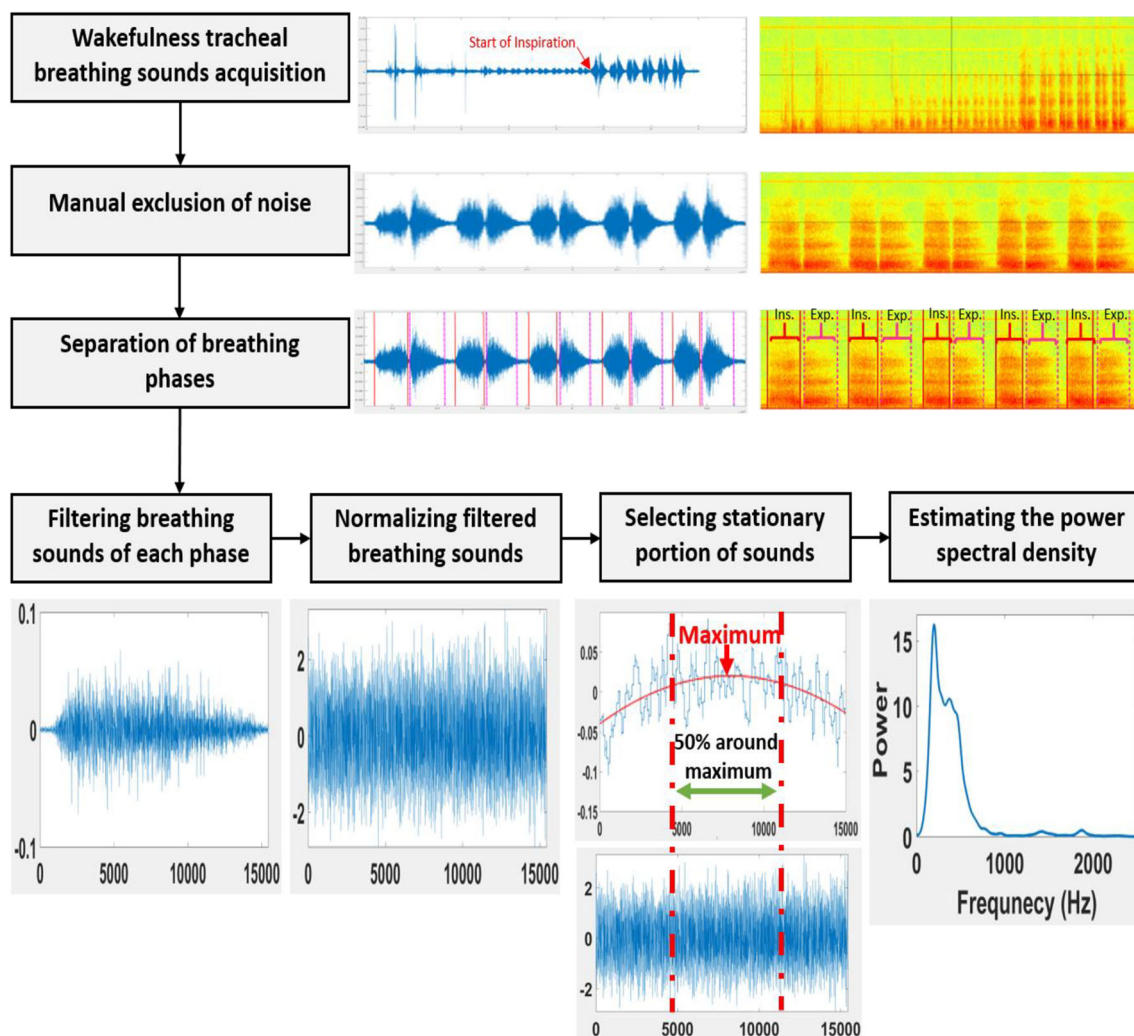


Fig. 2 Pre-processing and signal analysis framework. Tracheal breathing sounds were extracted during daytime while the participants were awake. After manual exclusion of noisy sounds, the breathing sounds were separated as inspiratory and expiratory phases. Each individual phase of

breathing sounds were filtered and normalized to purify tracheal breathing sounds. Afterward, the power spectrum density were estimated for the approximately stationary portion of each normalized sound

participants in each OSA groups, and Critical_value is the *t*-statistics (values of *t*-distributions) with *n* – 1 degree of freedom.

We considered the mean and slope of the regions with different slopes and the non-overlapped area between the average power spectra of two groups in association with their corresponding 95% CIs as characteristic features to be further selected for classification. As an example, Fig. 3 shows the average power spectra that calculated from the summation of inspiratory and expiratory mouth TBS signals in linear and logarithmic scale for the non-OSA and moderate/severe OSA groups in one of the training sets as well as their corresponding 95% CIs. TBS features were extracted from the regions between the solid, dotted, and dashed lines. Overall, we extracted 78 TBS features. Combining these TBS features with seven anthropometric features of each participant (sex, age, height, weight, body

mass index (BMI), neck circumference (NC), and Mallampati score (MP)) resulted in a total of 85 features.

2.5 Feature reduction and classification

Due to the high number of extracted features, we used a feature selection/reduction approach to find the most characteristic features for classification of the two moderate/severe OSA and non-OSA groups. The first step of feature reduction was to apply an unpaired *t* test on the features of the participants in each training set and to select the statistically significantly different features (*p* < 0.01) between non-OSA and moderate/severe OSA groups within that training set. In order to meet the statistical significance of 99% for the overall test, the significance level of each test was considered as $1 - (1 - 0.01)^{1/85} \approx 1.2 \times 10^{-4}$ [30, 31]. Within each training set, on average 11 features with *p* value > 1.2×10^{-4} were omitted,

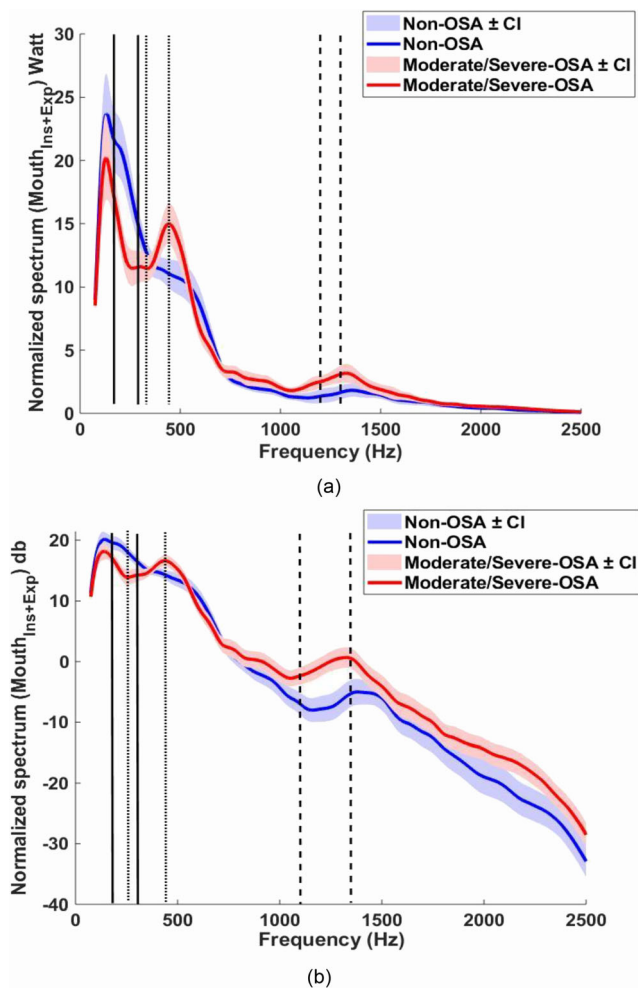


Fig. 3 Average power spectra of the summation of mouth inspiratory and expiratory breathing sounds with their 95% confidence intervals (CI, shadows) in non-OSA (blue) and moderate/severe OSA (red) groups, using one of the training sets with 131 participants in **a** linear scale, **b** logarithmic scale with base of 10. The area between solid lines, dotted lines, and dashed lines shows the regions where the features were extracted

and the features with the discrimination potential between the two groups were kept for further analysis.

To classify the two non-OSA and moderate/severe OSA groups, we considered a linear logistic regression model for the severity level of participants:

$$\log\left(\frac{\Pr(Y = 1|X = x)}{1 - \Pr(Y = 1|X = x)}\right) = \beta_0 + \beta^T x \quad (2)$$

where $Y \in \{0, 1\}$ is the true severity label of participants, X is the matrix of $(n \times p)$ for n participants and p features, β is the coefficient vector, and $\Pr(Y = 1|X = x)$ represents the probability of assigning a participant to the moderate/severe group.

To further reduce the number of features and obtain a parsimonious model, we used the LASSO approach [21] as a powerful and stable feature selection algorithm to automatically select significant features by shrinking the coefficient of

unimportant features to zero. This is done by estimating the parameters of the logistic regression via minimizing the negative log-likelihood with an l_1 -regularization defined as follows:

$$\operatorname{argmin}_{\beta} \left(-\frac{1}{n} \sum_{i=1}^n \left\{ y_i (\beta_0 + \beta^T x_i) - \log(1 + e^{\beta_0 + \beta^T x_i}) \right\} + \lambda \sum_{j=1}^p |\beta_j| \right) \quad (3)$$

where λ is the penalty parameter that is selected in a way to minimize the out of sample error (generalized error) of the model. For each value of λ , the above optimization problem is solved and the optimal value of λ is tuned using the cross-validation [32]. As the objective of this study was to balance the accuracy and computational cost, we selected λ such that it provides the smallest number of coefficients with a reasonable accuracy. Therefore, the λ that occurred within one standard error of the optimal λ was selected as the optimum value. The LASSO penalized logistic regression model was implemented using the Glmnet package of R statistical software version 3.5.1.

We applied this method to the dataset of entire non-OSA and moderate/severe OSA participants (164 participants) and also to the data of different training sets to investigate the robustness of the procedure. From the selected features of each dataset, the features with a high correlation ($> 50\%$) to other features were excluded, and the remaining were used for classification.

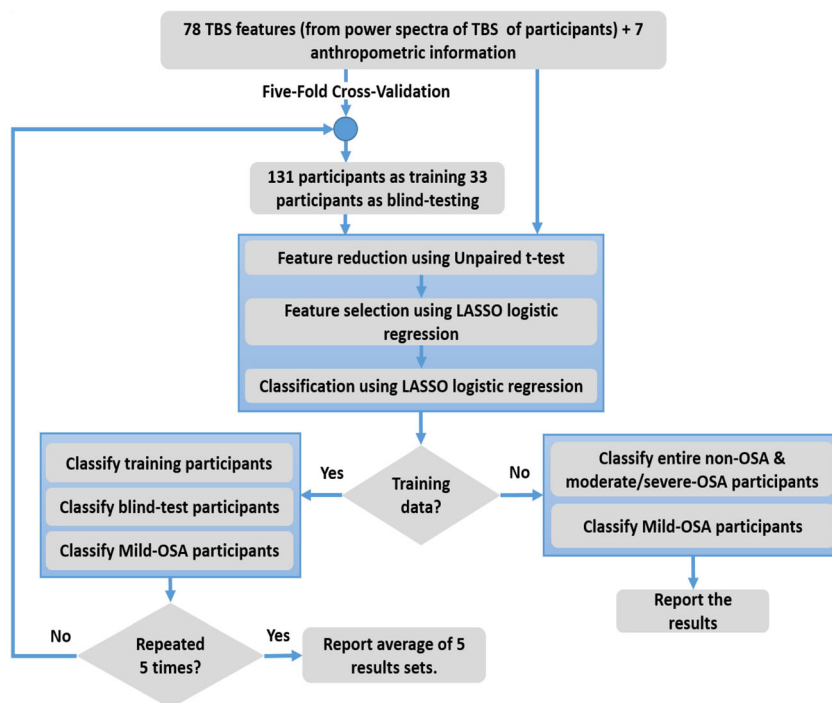
To classify each participant into one of the two groups of non-OSA or moderate/severe OSA, we used the features selected by the lasso feature selection method as input to another LASSO logistic regression method trained for the classification. Once β coefficients are determined, the probability of assigning a participant to either non-OSA or moderate/severe OSA is determined using the Eq. 2. The accuracy, specificity, and sensitivity were calculated using test and training data of each dataset as well as data of entire non-OSA and moderate/severe OSA participants. Data of the mild-OSA participants with $5 \leq \text{AHI} < 15$ were also classified at this stage with the same procedure using the models formed based on data of the entire non-OSA and moderate/severe OSA groups as well as data of each training dataset. Figure 4 shows the flowchart of this method for feature selection and classification.

The total computational cost of our proposed algorithm was also calculated; it is detailed in Appendix B.

3 Results

In this study, features were mainly extracted from the areas with different slopes and the non-overlapped regions of the average power spectra of non-OSA and moderate/severe OSA

Fig. 4 The flowchart of feature selection and classification methods, using the LASSO logistic regression



participants. As these regions were slightly changed among different training sets, for robustness, we chose the common area among the five-fold training datasets. Our feature selection procedure resulted in the following five features as described in Eq. 4–7; F represents a feature. Table 2 presents the features that were selected using data of each training set as follows:

$$\begin{aligned}
 &F1 : \textit{neck circumference} \\
 &F2 : \textit{slope}(10\log_{10}(\text{PSD}_{\textit{mouth-ins}} + \text{PSD}_{\textit{mouth-exp}}(260 : 440 \text{ Hz}))) \tag{4}
 \end{aligned}$$

$$F3 : \textit{mean}(10\log_{10}(\text{PSD}_{\textit{mouth-ins}} + \text{PSD}_{\textit{mouth-exp}}(1100 : 1350 \text{ Hz}))) \tag{5}$$

$$F4 : \textit{slope}((\text{PSD}_{\textit{mouth-exp}}(250 : 350 \text{ Hz}))) \tag{6}$$

$$F5 : \textit{mean}(10\log_{10}(\text{PSD}_{\textit{nose-ins}} + \text{PSD}_{\textit{nose-exp}}(1000 : 1300 \text{ Hz}))) \tag{7}$$

Table 3 presents the coefficients of each feature for different training sets and also the entire dataset of non-OSA and moderate/severe OSA participants. Figure 5 shows the correlation between the AHI and the selected features for separating non-OSA and moderate/severe OSA participants within one of the training sets as an example.

The selected features of different datasets were investigated using the unpaired *t* test to check whether they were statistically significantly different from each other. All the *p* values were found to be highly significant ($p < 7.9 \times 10^{-6}$). The discriminative power of the selected features between non-OSA

and moderate/severe OSA participants within different datasets were also investigated using other sets of unpaired *t* tests. Compared to moderate/severe OSA participants, the non-OSA participants showed to have lower NC, lower high-frequency power, and lower slope around the second formant. All the *p* values were found statistically significant ($p < 4.6 \times 10^{-6}$).

Different multivariate analysis of variance (MANOVA) tests were run over the selected features of the five different training datasets and also the entire non-OSA and moderate/

Table 2 Selected features for different training sets and the dataset containing entire non-OSA and moderate/severe OSA participants, captured using the regularized logistic regression with the LASSO penalty feature selection method

Dataset	F1	F2	F3	F4	F5
Training set 1	*	*	*		
Training set 2	*	*			*
Training set 3	*	*	*		
Training set 4	*	*			*
Training set 5	*			*	*
Entire non-OSA and moderate/severe OSA	*	*	*		

F1: neck circumference; F2: slope (10 log 10(PSD_{MI} + PSD_{ME}[260 : 440 Hz])); F3: mean (10 log 10(PSD_{MI} + PSD_{ME}[1100 : 1350 Hz])); F4: slope (PSD_{ME} [250 : 350 Hz]); F5: mean(10 log 10(PSD_{NI} + PSD_{NE}[1000 : 1300 Hz]))

PSD_{ME} is power spectrum density of mouth expiratory sounds; PSD_{MI} is power spectrum density of mouth inspiratory sounds; PSD_{NE} is power spectrum density of nose expiratory sounds; PSD_{NI} is power spectrum density of nose inspiratory sounds

Table 3 Coefficients of selected features of different training sets and the dataset containing entire non-OSA and moderate/severe OSA participants, captured using the regularized logistic regression with the LASSO penalty feature selection method

Dataset	β_0	β_1	β_2	β_3
Training set 1	- 6.5	0.2	4.8	0.05
Training set 2	- 7.2	0.2	10.5	0.05
Training set 3	- 7.5	0.2	11.6	0.03
Training set 4	- 4.5	0.1	11.6	0.08
Training set 5	- 6.8	0.2	10.2	0.08
Entire non-OSA and moderate/severe OSA	- 7.1	0.2	10.2	0.04

β_0 , intercept; β_1 , coefficient of the neck circumference; β_2 , coefficient of the feature related to the slope of the power spectrum around second formant; β_3 , coefficient of the feature related to the average power spectrum in the high-frequency range.

The reason that we decided to italicize these entries is that they are either the average over different datasets or the result of the overall data, and we just want to emphasize these findings. In addition these are the results that we reported on the text of the paper

severe OSA dataset to ensure the combination of the selected features in each feature set were (statistically) significantly different between the two groups; all were found to be highly significant ($p < 6.5 \times 10^{-7}$). Therefore, the selected features were used for classification of the two groups within each training sets as well as the entire dataset (all five-folds' data).

Table 4 provides the classification specificity, sensitivity, and accuracy of applying the regularized logistic regression with LASSO penalty over data of different training datasets and their corresponding blind-testing sets, as well as data of the entire five-folds' dataset. The classification of the entire 164 non-OSA and moderate/severe participants showed 81.1% accuracy with sensitivity and specificity of 84.4% and 77.0%, respectively. The

average accuracies of the 5 training sets were found to be $82.3\% \pm 0.7$ and $79.3\% \pm 6.1$ for training and testing, respectively. The sensitivity (specificity) were $85.0\% \pm 2.0$ ($79.1\% \pm 1.0\%$) for training and $82.2\% \pm 7.2$ ($75.8\% \pm 9.9\%$) for testing, respectively. Table 4 also reports the area under the curve (AUC) values associated with the average receiver operating characteristic (ROC) curves of the regularized logistic regression classifier for classification of non-OSA and moderate/severe OSA participants in different datasets, using their selected features. The AUC value associated with the ROC curve calculated using data of all non-OSA and moderate/severe OSA participants was shown to be 89.02. In addition, the average of the AUC values between the 5 training datasets was found to be 89.9 ± 1.7 . Figure 6 shows the ROC curve calculated using the non-OSA and moderate/severe OSA participants for one of the training sets. Table 5 shows the number of misclassified participants in each fold of training and testing as well as the entire dataset. Table 6 details the anthropometric information of the misclassified participants when applying the LASSO logistic regression model to the entire dataset.

The computational cost for each step of our proposed method is detailed in Appendix B. The overall computational cost of our proposed approach is shown in Eq. 8.

$$16n \times O(N \log(L)) + p O(n) + O(pn) + O(3n) \approx 16n \times O(N \log(L)) + O(pn) \quad (8)$$

where n is the number of participants, N is the length of the TBS after normalization, L is the window length, and p is the number of features. The computational cost of our proposed method is linear in terms of the number of features and the number of participants; hence, the method is very fast.

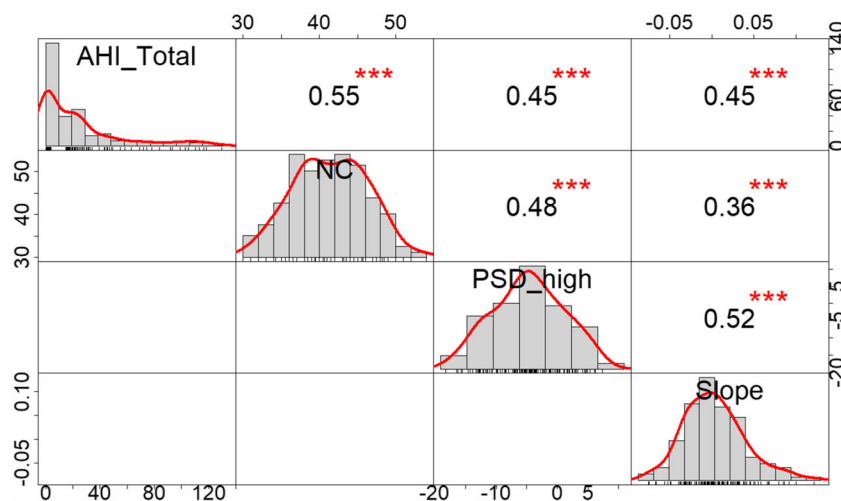


Fig. 5 Visualization of the correlation between AHI and the selected features of one of the training datasets, using LASSO feature selection method. On top, the values of the correlation between features plus their significance level as star (***) $p < 0.001$). The overall distribution of features showed as histograms with a fitted curve. The selected features

are neck circumference, the average of the power spectrum calculated from summation of mouth inspiratory and expiratory breathing sounds in the high-frequency range of (1100–1350 Hz), and the slope of the power spectrum calculated from summation of mouth inspiratory and expiratory breathing sounds in the frequency range of (260–440 Hz)

Table 4 Regularized logistic regression classification results for different training sets and their corresponding blind-testing sets, the average of all five datasets and their standard deviations (std), and also for

the dataset containing entire non-OSA and moderate/severe OSA participants, using their corresponding features selected by the LASSO penalized logistic regression feature selection method.

Dataset		Specificity (%)	Sensitivity (%)	Accuracy (%)	AUC (%)
Training set 1	Training	78	87.5	83.2	92.1
	Blind testing	60	83.3	72.7	
Training set 2	Training	78	86.1	82.4	90
	Blind testing	73.3	72.2	72.7	
Training set 3	Training	80	82	81.1	87.3
	Blind testing	85.7	77.8	81.3	
Training set 4	Training	79.7	84.7	82.4	89.4
	Blind testing	80	88.8	84.8	
Training set 5	Training	79.7	84.7	82.4	90.5
	Blind testing	80	88.9	84.8	
Average	<i>Training ± std</i>	<i>79.1 ± 1.0</i>	<i>85 ± 2.0</i>	<i>82.3 ± 0.7</i>	<i>89.9 ± 1.7</i>
	<i>Blind testing ± std</i>	<i>75.8 ± 9.9</i>	<i>82.2 ± 7.2</i>	<i>79.3 ± 6.1</i>	
Entire non-OSA and moderate/severe OSA	–	77.0	84.4	81.1	89.02

The reason that we decided to italicize these entries is that they are either the average over different datasets or the result of the overall data, and we just want to emphasize these findings. In addition these are the results that we reported on the text of the paper

4 Discussion

In this study, we proposed to use a regularized logistic modeling approach as a simple and fast method for analyzing TBS recorded during daytime when the subjects were awake (wakefulness) with the purpose of predicting their OSA severity by a few minutes breathing sound recording and data analysis. Our proposed approach is discriminative in nature and does not rely on strong model-based assumptions. We developed a robust feature selecting approach, and used it in classifying subjects with $AHI < 5$ and $AHI \geq 15$. The first set of features considered for this study ($p = 85$) were a combination of the TBS features and the anthropometric information of participants. As there were two breathing maneuvers and we

investigated respiratory phases both separately and in combination, the number of features was large. Thus, we used a feature selection approach based on regularized logistic regression methodology with a LASSO penalty to select the best subset of features.

In LASSO feature selection method, shrinking and removing some of the model’s coefficients lead to reducing the variance without a considerable increase in bias; hence, it provides desirable prediction accuracy [32]. This is useful when the number of participants is not much larger than the number of features, as it was the case in our study. The LASSO method is very fast and provides an interpretable model by weighting the selected features. Therefore, it is easy to explain which features are more important for predicting the true class labels; the coefficients of the selected features show their importance ranking (Table 3).

As the LASSO method uses a cross-validation procedure to automatically estimate the optimal value of λ (the penalty parameter), there is no need to perform another cross-validation procedure for obtaining the training and testing results. However, we generated five sets of randomly selected training and blind-testing data to ensure the unbiased test classification. Our findings (Tables 2 and 3) show that in different training sets the selected features and their corresponding coefficient values were different; however, in all of the training sets and also in the entire five-folds’ dataset three selected feature types were common. Those features were (1) the neck circumference, (2) the slope of the average PSD around the second formant, and (3) the average of the high-frequency portion of the average PSD. These features are congruent with the findings of our previous study on a separate dataset [28]. In all training and testing datasets of this study, the unpaired t test showed that the selected features were statistically significantly different from each other and also

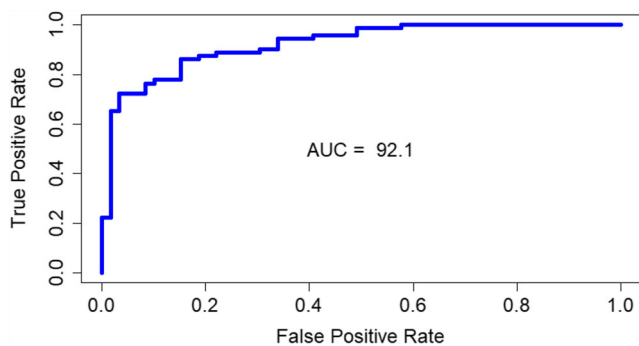


Fig. 6 ROC plot of the regularized logistic regression classification over selected features of one of the training datasets, found using the LASSO feature selection method. The selected features are neck circumference, the average of the power spectrum calculated from summation of mouth inspiratory and expiratory breathing sounds in the high-frequency range of (1100–1350 Hz), and the slope of the power spectrum calculated from summation of mouth inspiratory and expiratory breathing sounds in the frequency range of (260–440 Hz)

Table 5 Number of misclassified participants found using the regularized logistic regression classification for different training sets and their corresponding blind-testing sets, the average of all five datasets, and also for the dataset containing entire non-OSA and moderate/severe OSA participants

Dataset		Total misclassified	Non-OSA misclassified	Moderate/severe OSA misclassified
Training set 1	Training	22	13	9
	Blind testing	9	6	3
Training set 2	Training	23	13	10
	Blind testing	9	4	5
Training set 3	Training	25	12	13
	Blind testing	6	2	4
Training set 4	Training	23	12	11
	Blind testing	5	3	2
Training set 5	Training	24	13	11
	Blind testing	4	3	1
Average	<i>Training</i>	<i>24</i>	<i>13</i>	<i>11</i>
	<i>Blind testing</i>	<i>11</i>	<i>4</i>	<i>7</i>
Entire non-OSA and moderate/severe OSA		<i>31</i>	<i>17</i>	<i>14</i>

The reason that we decided to italicize these entries is that they are either the average over different datasets or the result of the overall data, and we just want to emphasize these findings. In addition these are the results that we reported on the text of the paper

different between the two groups of non-OSA and moderate/severe OSA. These features also showed high correlations with AHI. In addition, the ROC curves of the different LASSO logistic regression classifiers over different training sets and also on the entire dataset showed high performance with high values of AUC (Table 4). These high AUCs indicate the robustness of the LASSO logistic regression classifiers with respect to our selected features. Appendix A reports the results of using a simple logistic regression for classification as opposed to the LASSO logistic regression method. The simple logistic regression classification of the entire 164 non-OSA and moderate/severe participants showed 80.5% accuracy with sensitivity (specificity) of 82.2% (78.4%), which are lower than those of LASSO logistic regression classification (Table 4). The reason that LASSO logistic regression classification resulted in better than the simple logistic

regression is that regularization term of LASSO logistic regression classification method (λ) improves the generalization performance (performance on new data) by reducing the variance of parameter estimates (shrinks the coefficients that contribute most to the error).

It is also important to investigate the characteristics of misclassified participants. For the dataset containing all non-OSA and moderate/severe OSA participants, the overall accuracy of the LASSO logistic regression was 81.1%; out of 164 subjects, 31 were misclassified, out of which, 17 were from non-OSA and 14 were from moderate/severe OSA group (Table 5). Further investigation on the anthropometric information of misclassified subjects (Table 6) and comparing them with the non-OSA and moderate/severe participants of this study, revealed that overall, the non-OSA misclassified individuals are

Table 6 Anthropometric information's mean and their corresponding standard deviations (std) for the misclassified non-OSA and moderate/severe OSA participants of the dataset containing entire non-OSA and

moderate/severe OSA participants, found using the regularized logistic regression with LASSO penalty classifier

	Non-OSA (AHI < 5, $n = 74$)	Moderate/severe OSA (AHI ≥ 15 , $n = 90$)
AHI \pm std	1.1 \pm 1.1	29.5 \pm 20.2
NC (cm) \pm std	43.6 \pm 3.1	39.04 \pm 2.1
Age (year) \pm std	46.6 \pm 13.4	57.1 \pm 10.2
BMI (kg/m ²) \pm std	35.3 \pm 7.3	33.6 \pm 4.7
Sex (male:female)	13:4	5:9
Height (cm) \pm std	170.9 \pm 9.4	164.3 \pm 12.1
Weight (kg) \pm std	105.7 \pm 21.7	88.3 \pm 15.5
MP (I-II-III-IV)	8-6-1-2	5-7-1-1

AHI is apnea/hypopnea index; NC is neck circumference; BMI is body mass index; MP is Mallampati score and the numbers in that row show the number of subjects with MP level I, II, III, and IV

characterized with higher NC, BMI, height, weight, and being male, while the moderate/severe OSA misclassified individuals are characterized with lower NC, BMI, height, weight, and being female. This finding is not surprising, as it is known that these anthropometric parameters affect the breathing sounds while being correlated with AHI [33]. Thus, ideally, we should have the two non-OSA and moderate/severe OSA groups of the training and testing matched in these characteristics. However, that requires a much larger dataset.

In this study, similar to all other similar studies, we used the AHI values as the true label of each individual to be either in non-OSA or moderate/severe OSA group. Although AHI is one of the most important and commonly used indicators for diagnosis of OSA, there are many individuals with similar AHI but different daytime symptoms and different levels of OSA severity. The clinical diagnosis of OSA is not only based on AHI; it is based on a combination of frequency of arousals, daytime symptoms, lack of deep sleep, etc. [34]. We speculate that the TBS features are indeed good representatives of the OSA severity as they are affected by the pathophysiology of OSA. However, the UA pathology of two individuals with similar AHI (e.g., AHI of 14 and 16) may not be that much different to affect the TBS differently enough to be detected by sound analysis. For this reason, to avoid the misclassification in the borderline ranges with artificially crisp nature, we allowed a small gap ($5 \leq \text{AHI} < 15$) in continuous AHI values of the non-OSA and moderate/severe OSA groups to form the two groups. Once the classifiers were trained, the participants with AHI in this gap range were classified using our proposed method. Using the model obtained based on data of 164 non-OSA and moderate/severe participants, 11 out of 35 individuals in the gap range were assigned to the non-OSA group and the remaining 24 individuals assigned to the moderate/severe group. Further investigation of these participants revealed that the individuals assigned to the moderate/severe group had higher NC, higher slope around the second formant, and higher average PSD in high-frequency ranges compared to those assigned to the non-OSA group, while there was not much difference in their AHI values (8.8 ± 0.6 and 8.4 ± 0.8 , respectively). This finding confirms the shortcoming of using AHI as the only indicator for OSA severity. Finding a better OSA severity indicator still remains a challenge. Perhaps a better indicator would be a combination of several factors including AHI, duration of apneic events, and number of arousals, and also the breathing sounds features.

Last but not the least advantage of the proposed method is that the most repeated features among the selected features of all datasets were extracted from the combination of inspiratory and expiratory TBS. This finding implies that there may be no need to separate the breathing phases in the pre-processing step. Although there are various methods for automatic separation of breathing phases into inspiration and expiration [35, 36], however, none of them are 100% accurate. Therefore, there is

always a need for manual investigation of breathing sounds. Eliminating this part from the pre-processing procedure will speed up the data analysis significantly; hence, providing a faster online screening tool for OSA.

4.1 Physiological interpretations of the selected features

It is worth mentioning that selected features using the regularized logistic regression approach are physiologically meaningful and highly correlated with AHI (Fig. 5).

Obesity is a risk factor in patients with sleep apnea [37, 38]. The relationship between obesity and OSA has been demonstrated to be totally explained by variation in neck size [39, 40]. It was suggested that the increases in neck circumference and fat deposited around the upper airway might narrow the upper airway [38]. Therefore, we can conclude that NC, a surrogate measure of neck fat, can be considered as a predictor of the presence of OSA.

A highly collapsible/compliant pharyngeal airway is known as the main anatomical abnormality in the UA of OSA patients [41]. A reason for increased collapsibility of OSA subjects compared to that of controls is their smaller cross-sectional area [42, 43]. According to the Helmholtz equation [44], an UA with a smaller cross-sectional area reflects a lower resonant frequency (f) as follows:

$$f = \frac{c}{2\pi} \sqrt{\frac{A}{VL}} \quad (9)$$

where c is the speed of sound, A is cross-section area, V is the cavity volume, and L is the length of the bottleneck. The resonant frequencies of the UA are represented by formant frequencies; the peaks in the spectrum of sound signals. Hence, we expect to see lower formant frequencies for moderate/severe OSA participants. This is congruent with what is observed in Fig. 3, especially for the second formant (F2). In our study, the higher slope in PSD of the moderate/severe OSA participants' around their F2 implied that their F2 occurred in lower frequencies compared to that of non-OSA participants.

On the other hand, stiffness is the property of a tissue that resists the change in shape in response to an applied force. It has been shown that during wakefulness, the stiffness of the tongue muscles and soft palate tissues of OSA patients is higher compared to those of controls [45]. The stiffening of the narrower portion of the UA of OSA individuals was also supported by showing a predominantly bottle shape for their narrowed UA [20]. As an increase in stiffness is presented by an increase in high-frequency components of the breathing sounds, we expected to see higher energy in the high-frequency portion of the average PSD of moderate/severe OSA participants compared to that of non-OSA; this is what is observed in Fig. 3.

5 Conclusion

In this study, we investigated the application of regularized logistic regression model using a LASSO penalty for prediction of OSA severity during wakefulness using TBS analysis and anthropometric information. The selected features by this regression modeling approach are congruent with those selected by other methods in our previous studies; they are physiologically meaningful and highly correlated with AHI. The unbiased blind-testing classification accuracy, sensitivity, and specificity over the non-OSA and moderate/severe OSA groups of five different folds were found to be $79.3\% \pm 6.1\%$, $82.2\% \pm 7.2\%$, and $75.8\% \pm 9.9\%$, respectively. The accuracy for the entire dataset was 81.1% with sensitivity and specificity of 84.4% and 77.0% , respectively. Although results of this study are similar but not necessarily superior to previous studies' results but they are important from different aspects. The most important contributions of this work are (1) the method presented in this work does not need respiratory phase identification as the best features do not depend on that (2) the method is simple, quick, and computationally more effective than SVM and other previous methods for OSA screening during wakefulness; thus, suitable for online applications; (3) the method can be used for screening OSA individuals with severity as low as 15 with high sensitivity and a reasonable specificity. The main limitation of this study is the lack of a large sample size to have a better way to handle OSA severity prediction of individuals with $5 < \text{AHI} < 15$.

Glossary

AHI	Apnea/hypopnea index
AUC	Area under the curve
CI	Confidence interval
FFT	Fast Fourier transform
LASSO	Least absolute shrinkage and selection operator
LogVar	Logarithm of the sound's variance
MANOVA	Multivariate analysis of variance
OSA	Obstructive sleep apnea
PSD	Power spectrum density
PSG	Polysomnography
ROC	Receiver operating characteristic
S _a O ₂	Oxygen saturation level of blood
SVM	Support vector machine
TBS	Tracheal breathing sounds
UA	Upper airway

Appendix 1. Classification using simple logistic regression

In this study, as we had 2 classes of participants (non-OSA and moderate/severe OSA). Thus, we also applied the binomial

logistic regression classifier to the features selected by the LASSO penalized logistic regression for classification purpose. The classification of the entire 164 non-OSA and moderate/severe participants showed 80.5% accuracy with sensitivity and specificity of 82.2% and 78.4% , respectively. The average of the results between the 5 training sets showed accuracies of $82.0\% \pm 1.8\%$ and $78.1\% \pm 7.7\%$ for training and testing, respectively. The sensitivity (specificity) was $83.0\% \pm 3.1\%$ ($80.8\% \pm 2.2\%$) for training and $78.9\% \pm 11.4\%$ ($77.1\% \pm 9.9\%$) for testing, respectively. The AUC, associated with ROC curve calculated using all non-OSA and moderate/severe participants, was shown to be 80.3 . In addition, the average of the AUC values between the 5 training sets was shown to be 81.9 ± 1.7 . For the dataset containing all 164 non-OSA and moderate/severe OSA participants, 32 were misclassified, out of which, 16 were from non-OSA and 16 were from moderate/severe OSA group. In addition, 17 out of 35 subjects in the gap range were assigned to the non-OSA group and the remaining 18 subjects to the moderate/severe group.

The reason for achieving the better result for regularized logistic regression classification than the simple logistic regression is that as there is still some correlation between the selected features, the regularization term can improve the generalization performance (performance on new data) by reducing the variance of parameter estimates.

Appendix 2. Computational complexities

The computation cost of our proposed algorithm is relatively low. For the feature extraction part, the computational cost was estimated for a typical TBS with length N after normalization. The feature extraction phase consisted of 2 main parts: LogVar estimation and PSD estimation. The LogVar and PSD calculated for each segment, defined as the window with length L which moves along the signal with 50% overlap. Thus, the total number of overlapping windows within each TBS is approximately $\frac{2N}{L}$. The computational cost of calculating LogVar within one segment is L ; thus, the total number of operations to calculate the LogVar of each TBS is as follows:

$$L \times \frac{2N}{L} = O(N). \quad (10)$$

For the PSD estimation, the fast Fourier transform (FFT) of each segment was calculated and averaged. The computational cost of calculating FFT of each segment is $L \log(L)$; thus, the total cost of PSD estimation for each TBS is as follows:

$$O\left(\frac{2N}{L} \times L \log(L)\right) = O(N \log(L)). \quad (11)$$

The feature extraction part is from $O(k)$, where K is the number of extracted features from each TBS signal. As we have 16 signals for each participant (for mouth/nose breathing maneuvers, inspiratory/expiratory phases, summation/subtraction of phases, and normal/logarithmic scale), therefore, the total cost of feature extraction from n participant is as follows:

$$16n \times (O(N) + O(N \log(L)) + O(k)) \approx 16n \times O(N \log(L)). \quad (12)$$

Feature reduction part of this study consists of applying t test and LASSO. As each t test has the computational cost of $O(n)$, therefore, the overall t tests for all p features have the computational cost of $p \times O(n)$. On the other hand, in this study, we ran the LASSO with the Glmnet package of R, which uses the coordinate descent algorithm to find the optimum solution. The computational cost of this method is reported to be $O(pn)$ for each iteration of the optimization [32]. Besides, as three features were selected as the best feature set of each training set, thus, the computational cost of logistic regression classification using the LASSO penalty is $O(3n)$. Therefore, the total computational cost of our proposed method can be written as follows:

$$16n \times O(N \log(L)) + p O(n) + O(pn) + O(3n) \approx 16n \times O(N \log(L)) + O(pn) \quad (13)$$

which is linear in terms of n, p .

References

- Malhotra A, White DP (2002) Obstructive sleep apnoea. *Lancet* 360:237–245. [https://doi.org/10.1016/S0140-6736\(02\)09464-3](https://doi.org/10.1016/S0140-6736(02)09464-3)
- Eckert DJ, Younes MK (2014) Arousal from sleep: implications for obstructive sleep apnea pathogenesis and treatment. *J Appl Physiol* 116:302–313. <https://doi.org/10.1152/jappphysiol.00649.2013>
- Young T, Palta M, Dempsey J et al (2010) Burden of sleep apnea: rationale, design, and major findings of the Wisconsin Sleep Cohort Study. *WMJ Off Publ State Med Soc Wisconsin* 108:246–249
- Senaratna CV, Perret JL, Lodge CJ et al (2017) Prevalence of obstructive sleep apnea in the general population: a systematic review. *Sleep Med Rev* 34:70–81. <https://doi.org/10.1016/j.smrv.2016.07.002>
- Memtsoudis SG, Besculides MC, Mazumdar M (2013) A rude awakening — the perioperative sleep apnea epidemic. *N Engl J Med* 368:2352–2353. <https://doi.org/10.1056/NEJMp1302941>
- Sullivan F (2016) Hidden health crisis costing america billions. Underdiagnosing and Undertreating Obstructive Sleep Apnea Draining Healthcare System
- Kushida CA, Littner MR, Morgenthaler T, Alessi CA, Bailey D, Coleman J Jr, Friedman L, Hirshkowitz M, Kapen S, Kramer M, Lee-Chiong T, Loube DL, Owens J, Pancer JP, Wise M (2005) Practice parameters for the indications for polysomnography and related procedures: an update for 2005. *Sleep* 28:499–523. <https://doi.org/10.1093/sleep/28.4.499>
- Chiu HY, Chen PY, Chuang LP, Chen NH, Tu YK, Hsieh YJ, Wang YC, Guillemainault C (2017) Diagnostic accuracy of the Berlin questionnaire, STOP-BANG, STOP, and Epworth sleepiness scale in detecting obstructive sleep apnea: a bivariate meta-analysis. *Sleep Med Rev* 36:57–70. <https://doi.org/10.1016/j.smrv.2016.10.004>
- Young T, Shahar E, Nieto J et al (2002) Predictors of Sleep-disordered breathing in community-dwelling adults: the Sleep Heart Health Study. *Arch Intern Med* 162:893–900. <https://doi.org/10.1001/archinte.162.8.893>
- Elwali A, Moussavi Z (2017) Obstructive sleep apnea screening and airway structure characterization during wakefulness using tracheal breathing sounds. *Ann Biomed Eng* 45:839–850. <https://doi.org/10.1007/s10439-016-1720-5>
- Montazeri A, Giannouli E, Moussavi Z (2012) Assessment of Obstructive sleep apnea and its severity during wakefulness. *Ann Biomed Eng* 40:916–924. <https://doi.org/10.1007/s10439-011-0456-5>
- Sola-Soler J, Fiz JA, Torres A, Jane R (2014) Identification of obstructive sleep apnea patients from tracheal breath sound analysis during wakefulness in polysomnographic studies. *Conf Proc Annu Int Conf IEEE Eng Med Biol Soc IEEE Eng Med Biol Soc Annu Conf* 2014:4232–4235. <https://doi.org/10.1109/EMBC.2014.6944558>
- Yadollahi A, Montazeri A, Azarbarzin A, Moussavi Z (2013) Respiratory flow–sound relationship during both wakefulness and sleep and its variation in relation to sleep apnea. *Ann Biomed Eng* 41:537–546. <https://doi.org/10.1007/s10439-012-0692-3>
- Yadollahi A, Moussavi ZMK (2007) Acoustical flow estimation: review and validation. *IEEE Eng Med Biol Mag* 26:56–61. <https://doi.org/10.1109/EMEM.2007.289122>
- Wang Y, McDonald JP, Liu Y et al (2014) Dynamic alterations of the tongue in obstructive sleep apnea-hypopnea syndrome during sleep: analysis using ultrafast MRI. *Genet Mol Res* 13:4552–4563. <https://doi.org/10.4238/2014.June.17.7>
- Wu Z, Chen W, Khoo MCK, Davidson Ward SL, Nayak KS (2016) Evaluation of upper airway collapsibility using real-time MRI. *J Magn Reson Imaging* 44:158–167. <https://doi.org/10.1002/jmri.25133>
- Barkdull GC, Kohl C, Patel M, Davidson TM (2008) Computed tomography imaging of patients with obstructive sleep apnea. *Laryngoscope* 118:1486–1492. <https://doi.org/10.1097/MLG.0b013e3181782706>
- Liu K-H, Chu WCW, To K-W et al (2007) Sonographic measurement of lateral parapharyngeal wall thickness in patients with obstructive sleep apnea. *J Sleep* 30:1503–1508
- Schwab RJ, Pasirstein M, Pierson R, Mackley A, Hachadoorian R, Arens R, Maislin G, Pack AI (2003) Identification of upper airway anatomic risk factors for obstructive sleep apnea with volumetric magnetic resonance imaging. *Am J Respir Crit Care Med* 168:522–530. <https://doi.org/10.1164/rccm.200208-866OC>
- Finkelstein Y, Wolf L, Nachmani A, Lipowezky U, Rub M, Shemer S, Berger G (2014) Velopharyngeal anatomy in patients with obstructive sleep apnea versus normal subjects. *J Oral Maxillofac Surg* 72:1350–1372. <https://doi.org/10.1016/j.joms.2013.12.006>
- Tibshirani R (2011) Regression shrinkage and selection via the lasso: a retrospective. *J R Stat Soc Ser B (Statistical Methodol)* 73:273–282. <https://doi.org/10.1111/j.1467-9868.2011.00771.x>
- Hastie T, Tibshirani R, Friedman J (2009) *The Elements of Statistical Learning*
- Goldstein BA, Navar AM, Carter RE (2017) Moving beyond regression techniques in cardiovascular risk prediction: applying machine learning to address analytic challenges. *Eur Heart J* 38:1805–1814. <https://doi.org/10.1093/eurheartj/ehw302>

24. Wen Y, Pi FH, Guo P et al (2016) Sleep duration, daytime napping, markers of obstructive sleep apnea and stroke in a population of southern China. *Sci Rep* 6:1–9. <https://doi.org/10.1038/srep34689>
25. Oulhaj A, Al Dhaheeri S, Bin SB, Al-Houqani M (2017) Discriminating between positional and non-positional obstructive sleep apnea using some clinical characteristics. *Sleep Breath* 21: 877–884. <https://doi.org/10.1007/s11325-017-1499-0>
26. Arvin F, Doraisamy S, Safar Khorasani E (2011) Frequency shifting approach towards textual transcription of heartbeat sounds. *Biol Proced Online* 13:1–7. <https://doi.org/10.1186/1480-9222-13-7>
27. Pasterkamp H, Kraman SS, Wodicka GR (1997) Respiratory sounds, advances beyond the stethoscope. *Am J Respir Crit Care Med* 156:974–987
28. Hajipour F, Moussavi Z (2018) Spectral and higher order statistical characteristics of expiratory tracheal breathing sounds during wakefulness and sleep in people with different levels of obstructive sleep apnea. *J Med Biol Eng* 1–7. <https://doi.org/10.1007/s40846-018-0409-7>
29. Proakis JG, Manolakis DG (1996) Digital signal processing
30. Simes RJ (1986) An improved Bonferroni procedure for multiple tests of significance. *Biometrika* 73:751–754. <https://doi.org/10.1093/biomet/73.3.751>
31. Šidák Z (1971) On probabilities of rectangles in multivariate Student distributions: their dependence on correlations. *Ann Math Stat*:169–175. <https://doi.org/10.1214/aoms/1177693504>
32. Hastie T, Tibshirani R, Wainwright M (2015) Statistical learning with sparsity: the Lasso and generalizations
33. Sanchez I, Pasterkamp H (1993) Tracheal sound spectra depend on body height. *Am Rev Respir Dis* 148:1083–1087. https://doi.org/10.1164/ajrccm/148.4_Pt_1.1083
34. Epstein LJ, Kristo D, Strollo PJ et al (2009) Clinical guideline for the evaluation, management and long-term care of obstructive sleep apnea in adults. *J Clin Sleep Med* 5:263–276
35. Alshaer H, Pandya A, Bradley TD, Rudzicz F (2014) Subject independent identification of breath sounds components using multiple classifiers. *ICASSP, IEEE Int Conf Acoust Speech Signal Process - Proc* 3577–3581. <https://doi.org/10.1109/ICASSP.2014.6854267>
36. Huq S, Moussavi Z (2012) Acoustic breath-phase detection using tracheal breath sounds. *Med Biol Eng Comput* 50:297–308. <https://doi.org/10.1007/s11517-012-0869-9>
37. Young T, Skatrud J, Peppard PE (2004) Risk factors for obstructive sleep apnea. *291*:2013–2016
38. Schwartz AR, Patil SP, Laffan AM, Polotsky V, Schneider H, Smith PL (2008) Obesity and Obstructive sleep apnea: pathogenic mechanisms and therapeutic approaches. *Proc Am Thorac Soc* 5:185–192. <https://doi.org/10.1513/pats.200708-137MG>
39. Davies RJ, Stradling JR (1990) The relationship between neck circumference, radiographic pharyngeal anatomy, and the obstructive sleep apnoea syndrome. *Eur Respir J Off J Eur Soc Clin Respir Physiol* 3:509–514
40. Davies RJ, Ali NJ, Stradling JR (1992) Neck circumference and other clinical features in the diagnosis of the obstructive sleep apnoea syndrome. *Thorax* 47:101–105. <https://doi.org/10.1136/thx.47.2.101>
41. Eckert DJ, White DP, Jordan AS, Malhotra A, Wellman A (2013) Defining phenotypic causes of obstructive sleep apnea: identification of novel therapeutic targets. *Am J Respir Crit Care Med* 188: 996–1004. <https://doi.org/10.1164/rccm.201303-0448OC>
42. Schwab RJ, Gupta KB, Gefter WB, Metzger LJ, Hoffman EA, Pack AI (1995) Upper airway and soft tissue anatomy in normal subjects and patients with sleep-disordered breathing. Significance of the lateral pharyngeal walls. *Am J Respir Crit Care Med* 152:1673–1689
43. Haponik EF, Smith PL, Bohlman ME, Allen RP, Goldman SM, Bleecker ER (1983) Computerized tomography in obstructive sleep apnea. Correlation of airway size with physiology during sleep and wakefulness. *Am Rev Respir Dis* 127:221–226. <https://doi.org/10.1164/arrd.1983.127.2.221>
44. von Helmholtz H (1860) Theorie der Luftschwingungen in Röhren mit offenen Enden. *J für die reine und Angew Math* 57:1–72
45. Veldi, Vasar, Vain et al (2000) Computerized endopharyngeal myotonometry (CEM): a new method to evaluate the tissue tone of the soft palate in patients with obstructive sleep apnoea syndrome. *J Sleep Res* 9:279–284. <https://doi.org/10.1046/j.1365-2869.2000.00208.x>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Farahnaz Hajipour A PhD candidate and researcher in the Biomedical engineering program at University of Manitoba. Her Sci. interests are in respiratory signal processing, machine learning, and obstructive sleep apnea.



Mohammad Jafari Jozani An associate professor of Statistics with research interests in statistical learning, small area estimation as well as statistical inference using complex sampling designs based on rank information



Ahmed Elwali A PhD candidate and researcher in the Biomedical engineering program, at University of Manitoba. His Sci. interests are in signal processing, machine learning, and obstructive sleep apnea.



Zahra Moussavi The Director of Biomedical Engineering Program, a professor and Canada Research Chair at University of Manitoba. Her Sci. interests are in signal processing, obstructive sleep apnea, Alzheimer, and brain Concussion.