Original research

DIGITAL HEALTH

Digital Health
Volume 9: 1–15

DOI: 10.1177/20552076231152751
journals.sagepub.com/home/dhj
SAGE

Screening the risk of obstructive sleep apnea by utilizing supervised learning techniques based on anthropometric features and snoring events

Cheng-Yu Tsai[1], Wen-Te Liu[2,3,4,5,*], Wen-Hua Hsu[2,*], Arnab Majumdar[1], Marc Stettler[1], Kang-Yun Lee[3,6], Wun-Hao Cheng[7], Dean Wu[4,8,9,10,11], Hsin-Chien Lee[12], Yi-Chun Kuan[4,8,9,10,11], Cheng-Jung Wu[13], Yi-Chih Lin[13] and Shu-Chuan Ho[2,3]

Abstract

Objectives: Obstructive sleep apnea (OSA) is typically diagnosed by polysomnography (PSG). However, PSG is time-consuming and has some clinical limitations. This study thus aimed to establish machine learning models to screen for the risk of having moderate-to-severe and severe OSA based on easily acquired features.

Methods: We collected PSG data on 3529 patients from Taiwan and further derived the number of snoring events. Their baseline characteristics and anthropometric measures were obtained, and correlations among the collected variables were investigated. Next, six common supervised machine learning techniques were utilized, including random forest (RF), extreme gradient boosting (XGBoost), k-nearest neighbor (kNN), support vector machine (SVM), logistic regression (LR), and naïve Bayes (NB). First, data were independently separated into a training and validation dataset (80%) and a test dataset (20%). The approach with the highest accuracy in the training and validation phase was employed to classify the test dataset. Next, feature importance was investigated by calculating the Shapley value of every factor, which represented the impact on OSA risk screening.

Results: The RF produced the highest accuracy (of >70%) in the training and validation phase in screening for both OSA severities. Hence, we employed the RF to classify the test dataset, and results showed a 79.32% accuracy for moderate-to-severe OSA and 74.37% accuracy for severe OSA. Snoring events and the visceral fat level were the most and second most essential features of screening for OSA risk.

Conclusions: The established model can be considered for screening for the risk of having moderate-to-severe or severe OSA.

[1]Department of Civil and Environmental Engineering, Imperial College London, London, UK
[2]School of Respiratory Therapy, College of Medicine, Taipei Medical University, Taipei, Taiwan
[3]Division of Pulmonary Medicine, Department of Internal Medicine, Shuang Ho Hospital, Taipei Medical University, New Taipei City, Taiwan
[4]Sleep Center, Shuang Ho Hospital, Taipei Medical University, New Taipei City, Taiwan
[5]Research Center of Artificial Intelligence in Medicine, Taipei Medical University, Taipei, Taiwan
[6]Division of Pulmonary Medicine, Department of Internal Medicine, School of Medicine, College of Medicine, Taipei Medical University, Taipei, Taiwan
[7]Graduate Institute of Clinical Medicine, College of Medicine, Taipei Medical University, Taipei, Taiwan
[8]Department of Neurology, Shuang Ho Hospital, Taipei Medical University, New Taipei City, Taiwan
[9]Department of Neurology, School of Medicine, College of Medicine, Taipei Medical University, Taipei, Taiwan
[10]Taipei Neuroscience Institute, Taipei Medical University, Taipei, Taiwan
[11]Dementia Center, Shuang Ho Hospital, Taipei Medical University, New Taipei City, Taiwan
[12]Department of Psychiatry, Taipei Medical University Hospital, Taipei, Taiwan
[13]Department of Otolaryngology, Shuang Ho Hospital, Taipei Medical University, New Taipei City, Taiwan

[*]The authors contributed equal to the study.

Corresponding author:
Shu-Chuan Ho, School of Respiratory Therapy, College of Medicine, Taipei Medical University, 250 Wuxing Street, Taipei 110301, Taiwan.
Email: shu-chuan@tmu.edu.tw

Keywords

Obstructive sleep apnea, machine learning, snoring event, anthropometric measure, Shapley value

Submission date: 29 October 2022; Acceptance date: 4 January 2023

Introduction

Obstructive sleep apnea (OSA) is a disease characterized by repetitive hypoxemia caused by upper airway collapse while sleeping; potential risk factors include being aged, being overweight, consuming alcohol, smoking, and having nasal congestion.[1] Estimated prevalences of moderate to severe OSA in the United States in 2013 were 17% in elderly males and 9% in elderly females (50–70 years old).[2] In addition, increasing evidence exists for associations of OSA with hypertension, cardiovascular diseases,[3,4] increased cancer-related deaths,[5] a higher risk of developing Alzheimer's disease,[6] and an impaired quality of life.[7] Hence, OSA interventions, including early diagnosis and appropriate therapy, are critical.

Polysomnography (PSG) in a sleep laboratory is recommended as the gold standard for an OSA diagnosis.[8] Specifically, number of apneic and hypopneic events per hour during sleep, namely the apnea–hypopnea index (AHI), were obtained by utilizing PSG. Based on the derived AHI, OSA severity was classified into four categories: normal (AHI <5 times/h), mild (5 ≤ AHI <15 times/h), moderate (15≤ AHI <30 times/h), and severe (AHI ≥30 times/h).[9] Next, subjects, with an AHI of ≥30 times/h, were recommended to undertake corrective interventions. However, PSG is time-consuming for a full night stay in a hospital, and it is complex, with over 15 channels of measurement.[10] Also, an older age, nocturnal oxygen therapy, health status, and long waiting lists in many sleep laboratories may discourage patients from considering whether to undergo laboratory PSG. With a growing number of suspected OSA patients, previous studies proposed different approaches for simplifying OSA risk screening and enable home testing. For instance, researchers assessed the possibility of using nasal air pressure in the home environment or nocturnal pulse oximetry to conduct OSA diagnoses.[11,12] A related study suggested that using a simplified home sleep test (HST) may have an adequate OSA diagnosis rate in appropriately selected patients.[13] Another review study recommended employing sleep-related questionnaires which may serve as an OSA risk screener with advantages of convenience and low cost[14] However, since HST has fewer physiological signal numbers, these simplified measurements may be inaccurate in ruling out OSA risk when patients have respiratory events but are mainly associated with arousals[15] or those with comorbidities.[16] With self-reported questionnaires, elevated false positives and inconsistent subjective responses may affect the accuracy of OSA risk screening in patients with complex sleep disorders.[17] Given these deficiencies in current approaches, novel and applicable models to aid in screening OSA risks are required.

To construct a practical risk screening model by exploring clinical manifestations among patients with OSA may be worthwhile. First, snoring, the vibration of palatal soft tissues due to obstruction of air movements when breathing during sleep, may serve as an indicator representing the level of airway obstruction.[18] In addition, snoring is the most common symptom of OSA, which occurs in 70–95% of patients.[19] Previous studies predicting OSA investigated the recorded snoring sound intensity and analyzed it in a hospital[20] or at home.[21] Similarly, previous researchers employed the snoring intensity and frequency to predict the risk of having OSA using data of snorers with non-severe obesity.[22] Next, considering age, sex, and obesity as risk factors for presenting OSA, baseline characteristics and anthropometric features can serve as predictors for evaluating OSA risk composition states.[23,24] Another advantage is that these variables are relatively easily acquired compared to PSG parameters. Thus, it seems that establishing risk screening models for OSA by employing snoring events, baseline characteristics, and anthropometric features would be feasible since they are related to the risk of OSA presentation.

This retrospective study hypothesized that utilizing snoring events, anthropometric features, and baseline characteristics can serve as predictors to screen for OSA risks. The aim of this explorative study was to develop screening models for the risk of moderate-to-severe (AHI ≥15 times/h) and severe OSA (AHI ≥30 times/h) based on a dataset of over 3000 recruited participants. Correlations between these abovementioned variables and the AHI were also investigated. We hope that this work elucidates the feasibility of established models presenting adequate predictive accuracy to thereby speed up traditional diagnostic processing times.

Materials and methods

Study population

This was a retrospective study of adult patients who visited the sleep laboratory of Taipei Medical University, Shuang

Ho Hospital (New Taipei City, Taiwan) for an evaluation of suspected OSA between 1 May 2019 and 31 December 2021. The patient inclusion criteria were (1) aged 20–80 years, (2) having an overall PSG recording time of >6 h with a sleep efficiency of >40%, (3) not having previously undergone invasive surgery as OSA treatment, and (4) not regularly employing hypnotic or psychotropic medications.

Baseline characteristics and body composition

This study acquired baseline data and anthropometric measures for eligible individuals. Regarding the baseline, data on individuals' age, sex, body mass index (BMI), neck circumference (NC), and waist circumference (WC) were obtained. For anthropometric measures, before PSG, all patients were instructed to urinate and then undergo body composition measurements (namely bioelectrical impedance) utilizing the Tanita MC-780 system (Tanita, Tokyo, Japan). It was noted that all patients were required to fast for at least 3 h before PSG. In terms of measures, the fat mass and fat-free mass (i.e. bone and muscle mass) were assessed on different scales (i.e. whole body, limbs, and trunk). Fat percentages on the aforementioned scales and the body fat ratio (trunk/limbs) were thereby obtained. Next, the visceral fat level (as an indicator for estimating the fat encircling the organs in the abdominal space; which ranged from 1 to 55), the basal metabolic rate (BMR, as the lowest required energy), and physique rating (ratio of body fat mass to muscle mass) were determined. In addition, total body water (TBW), its percentage (TBW/body weight), its distribution, and the extracellular water (ECW)/intracellular water (ICW) ratio were determined. All of the collected parameters were utilized for further exploration.

PSG

In-laboratory PSG was performed utilizing ResMed Embla N7000 (ResMed, San Diego, CA, USA) and Embla MPR (Natus Medical, Pleasanton, CA, USA). PSG includes various sensors (i.e. electroencephalogram, electro-oculogram, electromyogram of the chin and leg, a nasal cannula, oral–nasal thermistor, bands for the chest and abdomen, pulse oximetry, and a piezoelectric vibration sensor). All of the signals were recorded using RemLogic software (version 3.41, Embla, Thornton, CO, USA) and scored by certified PSG technologists per the *Americana Academy of Sleep Medicine* (AASM) *Scoring Manual*.[25] For snoring which serves as an indicator of upper airway obstruction, these events were assessed by a piezoelectric vibration sensor placed on the triangle of the neck. Technically, this sensor measures frequencies of oscillations at the skin surface, thereby generating a piezoelectric signal to represent the snoring waveform. Snoring events were defined as protruding from the background and being synchronized with breathing, except for the body movement time. Piezoelectric signals were recorded at a sampling rate of 200 Hz and with AASM-recommended filter settings (low frequency of 10 Hz and high frequency of 100 Hz). Regarding OSA severity, the AHI, defined as the number of apneic and hypopneic events of the total sleep time, was obtained, and this index was further divided into four OSA levels, namely normal (AHI: <5 times/h), mild (5 ≤ AHI <15 times/h), moderate (15≤ AHI <30 times/h), and severe (AHI ≥30 times/h).[26]

Statistical analysis

This study performed statistical examinations using Python (version 3.9.7) and open-source stats library-SciPy (version 1.9.1). First, Shapiro–Wilk's method was applied to examine data normality. Spearman's correlation were used to determine correlations among anthropometric measures, snoring details, and sleep quality indices, including the AHI and oxygen desaturation index (ODI) of ≥3%, due to their non-normal distributions. In all statistical examinations, $p \leq 0.05$ was considered statistically significant.

Model establishment

Figure 1 presents an overview of the study. Based on the datasets of baseline characteristics, anthropometric measures, and snoring details, various supervised learning models were developed to screen for the risk of moderate-to-severe OSA or severe OSA. Specifically, on the basis of our literature review, six machine learning approaches, including random forest (RF), extreme gradient boosting (XGBoost), k-nearest neighbor (kNN), support vector machine (SVM), naive Bayes (NB), and logistic regression (LR), were recruited to establish screening models for two different OSA severity risk levels.[27,28] Figure 2 demonstrates the flowchart of developing the models. Initially, all of the collected data were integrated as a total dataset and subsequently independently separated into training and test datasets at an 80% to 20% ratio. In the training phase, the study employed a grid search strategy for model optimization for each classifier and a 10-fold cross-validation technique to prevent overfitting.[29] Regarding the grid search strategy, this study tuned parameters to compare the accuracy as follows: (a) RF with a bootstrap technique: criterion (Gini or entropy) and the amount of established classification and number of regression trees (N = 250, 500, and 750); (b) XGBoost with a bootstrap technique: criterion (mean squared error (MSE), squared error, or Friedman MSE) and the number of established estimators (N = 250, 500, and 750); (c) kNN: weight type (uniform or distance) and the k-value (k: between 2 and 9); (d) SVM: kernel function (linear, polynomial, and radial basis function) and regularization values (C, between 10^{-10} and 1010); (e) LR: inverse values of regularization (C: between

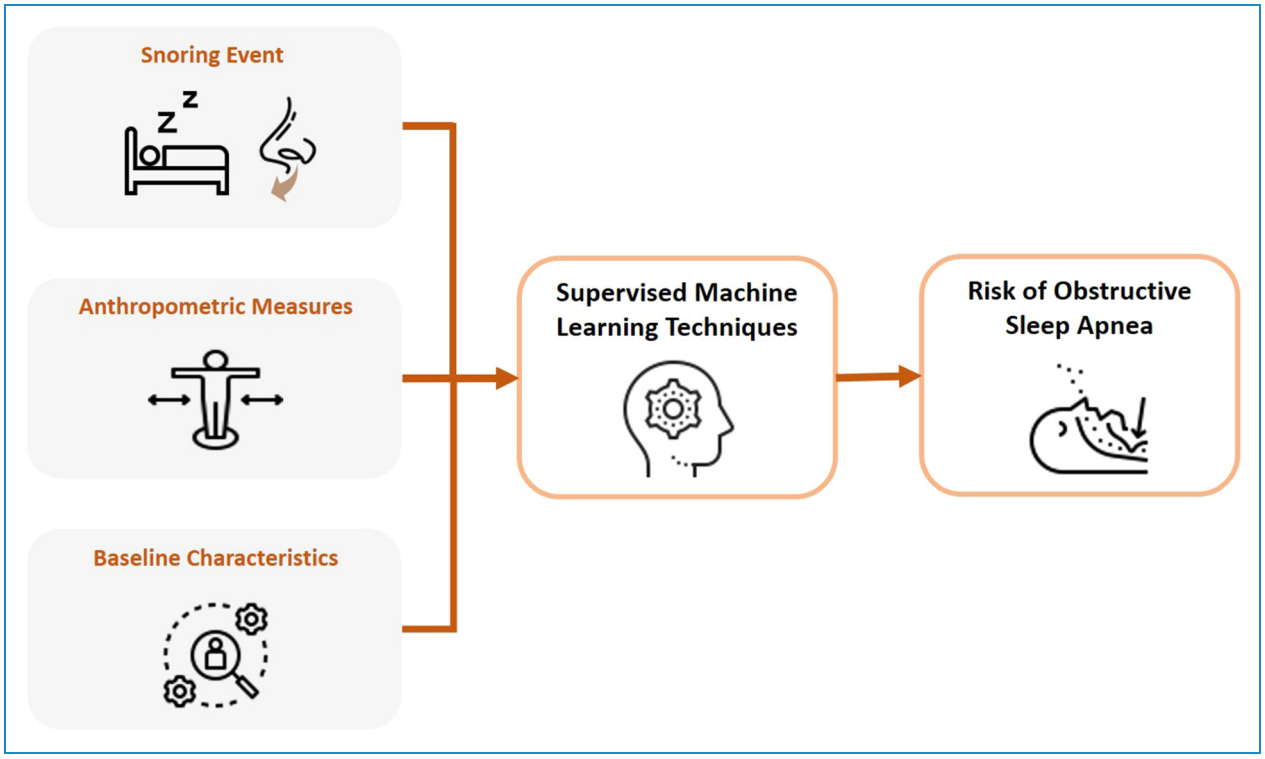

Figure 1. Study overview. This study established risk screening models for obstructive sleep apnea (OSA) utilizing machine learning techniques. The noncontact determined feature, snoring events, which served as an upper airway obstruction indicator, was employed. Easily obtained parameters, which served as overview evaluations, including anthropometric measures and baseline characteristics, were used.

10^{-10} and 10^{10}); and (f) NB: the portion of the largest variance of all features (var_smoothing: between 10^{-10} and 10^{10}). These input parameters for optimizing each model were referenced from previous research.[30,31] Next, the performances, in terms of the accuracy, precision, recall, F1-score, and area under the receiver operating characteristics curve (AUROC), of each trained model with optimized parameters were respectively calculated. The approaches which showed the relatively highest accuracy in this phase were employed in the test phase for further investigation. In other words, the training set (80% of the dataset) was utilized for training each model, optimizing the parameters, and validating the performance, while the test set (20% of the dataset) was only employed to test the model which demonstrated the best performance in the training phase. As to the importance of input features of the employed models, this study calculated the Shapley value of each variable and visualized these values in a scatterplot to assess the contribution of every feature within the models for OSA risk evaluation. The order of variables was arranged from top to bottom according to their Shapley values. Red spots represent high values of that variable, whereas blue spots represent low values (Figure 3). It was noted that the Shapley value is derived based on coalitional game theory and is widely utilized to interpret feature importance values of models.[32]

Results

Baseline characteristics of the study population

In total, 3529 patients were enrolled in this retrospective study. Their detailed baseline characteristics are presented in Table 1. The mean age of subjects was 47.64 years, and the majority were males ($N = 2305$, 65.32%). The average BMI was 26.78 ± 4.96 kg/m^2, the NC was 37.4 ± 4.06 cm, and the WC was 91.56 ± 12.64 cm. As to snoring details, snoring events and the index were 1080.24 ± 1070.69 times and 230.13 ± 221.6 times/h. The AHI was 30.03 ± 24.77 times/h, and ODI was 24.21 ± 24.48 times/h. Regarding OSA severity, patients with severe OSA were the relative majority ($N = 1390$, 39.39%), and patients with mild OSA ($N = 818$, 23.18%) were approximately equal to those with moderate OSA ($N = 933$, 26.44%). Concerning anthropometric measures, percentages of fat and muscle in the whole body scale were $28.71\% \pm 8.95\%$ and $67.45\% \pm 8.61\%$, respectively, and the mean visceral fat was 11.67 ± 4.93 level. As to body water details and its distribution, the mean TBW was 36.65 ± 7.15 kg, and it included a mean ECW of 15.09 ± 2.41 kg and a mean ICW of 21.55 ± 4.91 kg (ECW/ICW ratio: 0.71 ± 0.08).

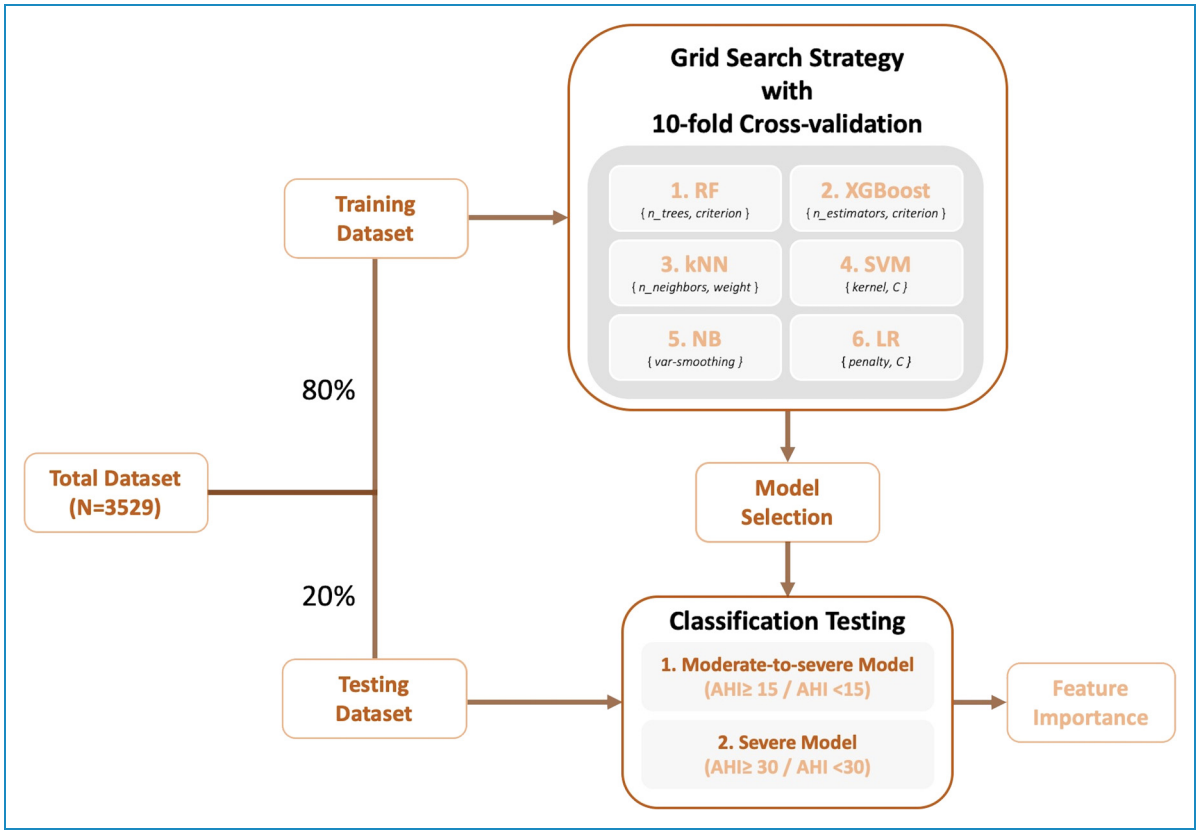

Figure 2. Training process with grid search cross-validation. Various machine learning models were trained using grid search cross-validation (k-fold: 10). The accuracy of the models in the validation stage was determined. The model demonstrating the highest accuracy was employed to predict the test data, and the feature importance was investigated.
AHI, apnea–hypopnea index; C, regularization values; XGBoost, extreme gradient boosting; kNN, k-nearest neighbor; LR, logistic regression; n_trees, number of classification and regression trees; n_estimators, number of gradient boosted trees; NB, naive Bayes; RF, random forest; SVM, support vector machine; var_smoothing, portion of the largest variance of all features.

Correlation analysis

Table 2 presents correlations of baseline characteristics, snoring details, and anthropometric measures at various scales with the AHI and ODI of 3529 participants. On the basis of Spearman's correlation, all of the collected variables were significantly correlated with the AHI (all p's $<$ 0.01, except for the ECW-to-ICW ratio, which was $p <$ 0.05). Likewise, all of the collected variables were significantly correlated with the ODI (all p's < 0.01), except for the ECW-to-ICW ratio. In addition, to determine correlation coefficients between the AHI and anthropometric measures at each severity level, this study further performed a subgroup analysis, and outcomes are reported in "Supplementary materials" (Supplementary Table S1). Notably, significant but weak correlations between snoring events and the AHI were observed in the three groups of AHI <15 events/h, AHI ≥15 events/h, and AHI <30 events/h (ρ ranged 0.21 to 0.33), while significant but very weak correlations were observed in individuals whose AHI was higher than 30 events/h ($\rho = 0.07$). Regarding the visceral fat level, significant but moderate correlations between this level and the AHI were observed in the two groups (AHI ≥15 events/h: $\rho = 0.42$; AHI <30: $\rho = 0.43$), while significant but weak correlations were observed in the two groups (AHI <15 events/h: $\rho = 0.37$; AHI <30: $\rho = 0.34$).

Training and cross-validation performances of the machine learning approaches

Performance metrics of each model in the training and validation phases are illustrated in Table 3. Among the developed models for screening the risk of moderate-to-severe OSA, the RF model had the highest values of accuracy (80.18% ± 1.14%), F1-score (85.35% ± 0.85%), and AUROC (83.08% ± 2.03%), while the XGBoost model showed the second highest values of accuracy (78.29% ± 1.44%), F1-score (84.18% ± 1.06%), and AUROC (82.18% ± 1.78%). Regarding the models for screening the risk of having severe OSA, similarly, the RF and

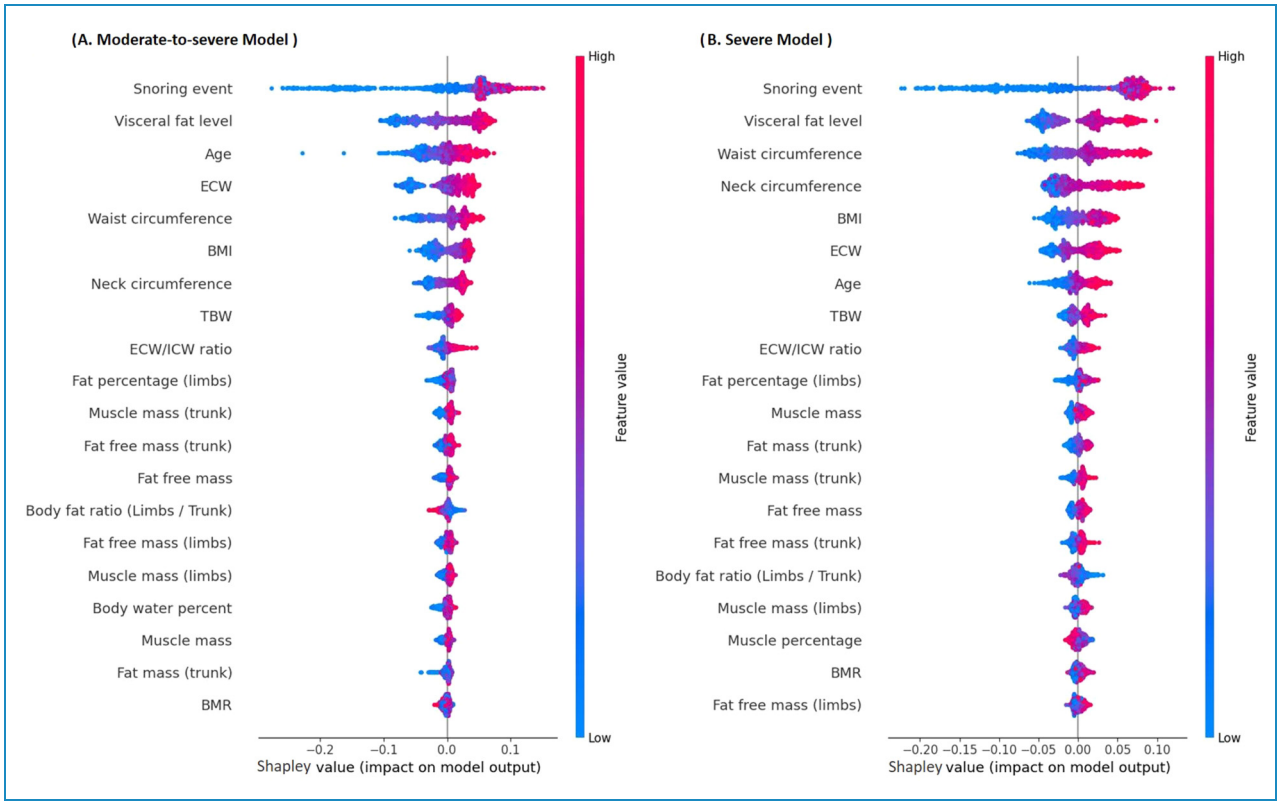

Figure 3. Density scatterplot of SHAP values of input parameters for the random forest (RF) models for screening moderate-to-severe and severe risk of obstructive sleep apnea (OSA) with the test dataset. (a) Moderate-to-severe OSA model and (b) severe OSA model. BMI, body mass index; BMR, basal metabolic rate; ECW, extracellular water; ICW, intracellular water; TBW, total body water.

XGBoost models still demonstrated the highest and second highest performances (RF: accuracy: 76.96% ± 2.55%, F1-score: 67.64% ± 3.55%, AUROC: 82.82% ± 2.37%; XGBoost: accuracy: 74.85% ± 2.09%, F1-score: 66.02% ± 3.08%, AUROC: 81.19% ± 1.8%). Due to its best performance (highest accuracy and AUC), the RF model was recruited to separately predict the test datasets for testing the performance of screening for the risk of moderate-to-severe or severe OSA, and the feature importance values of the model were further determined.

Accuracy performance and feature importance

The classification performance metrics of the RF model using the test dataset are shown in Table 4. First, for moderate-to-severe OSA risk screening, the RF model presented an accuracy of 79.32% ± 2.99%, an F1-score of 85.57% ± 2.59%, and an AUROC of 82.76% ± 2.79%. With the RF model for screening for the risk of severe OSA, the accuracy was 74.37% ± 3.22%, the F1-score was 66.31% ± 3.49%, and the AUROC was 82.84% ± 2.78%. Next, Figure 3 shows a scatterplot of feature importance values of these two severity types of RF models by presenting the Shapley value of each input feature. In both OSA severity types of screening models, snoring events and the visceral fat level had the relative highest and second highest Shapley values, which mean they were the most and second most important features, respectively. Age, NC, WC, BMI, TBW, ECW, and the ECW-to-ICW ratio respectively demonstrated the third- to ninth highest Shapely values in risk screening models for both moderate-to-severe OSA and severe OSA.

Model performance comparison by employing various input features

To evaluate the classification performance of the risk of having OSA based on employing different input features, this study further established models with various input feature bundles. Results derived using the training set is illustrated in Supplementary Tables S2 to S5. For moderate-to-severe OSA risk screening, the RF model presented the highest accuracy by employing baseline details and snoring events (78.13% ± 1.63%), anthropometrics (74.95% ± 2.24%) or only snoring events (75.02% ± 1.87%)), while the XGBoost had the highest accuracy in the model that utilized snoring events and anthropometrics (78.21% ± 1.56%). When screening for the risk of severe

Table 1. Baseline characteristics and anthropometric measures of participants separated by obstructive sleep apnea (OSA) severity (*N*=3529).

Variable	Mean ± SD	Variable	Mean ± SD
Baseline characteristics		Anthropometric measures	
Age (years)	47.64 ± 13.33	Whole body	
Sex (male/female)	2305 / 1224	Fat mass (kg)	21.91 ± 10.16
BMI (kg/m^2)	26.78 ± 4.96	Fat-free mass (kg)	52.61 ± 10.83
Neck circumference (NC) (cm)	37.4 ± 4.06	Muscle mass (kg)	49.79 ± 10.36
Waist circumference (WC) (cm)	91.56 ± 12.64	Bone mass (kg)	2.82 ± 0.48
Sleep stage summary		Fat percent (%)	28.71 ± 8.95
Sleep efficiency (%)	76.45 ± 13.23	Muscle percent (%)	67.45 ± 8.61
Mean SpO_2 (%)	94.92 ± 2.31	Visceral fat (level)	11.67 ± 4.93
Minimum SpO_2 (%)	83.23 ± 8.8	BMR (kJ)	6314.64 ± 1250.38
WASO (min)	57.99 ± 41.06	Physique rating (level)	29.62 ± 10.25
Total sleep time (min)	280.1 ± 49.6	Body fat ratio (limbs/trunk)	0.77 ± 0.16
Sleep stage (% of SPT)		Limbs	
Wake	17.19 ± 12.08	Fat mass (kg)	9.41 ± 4.52
REM	11.64 ± 6.38	Fat-free mass (kg)	25.12 ± 6.31
NREM	71.14 ± 10.64	Muscle mass (kg)	23.67 ± 5.97
Snoring details		Fat percent (%)	26.86 ± 8.45
Snoring event (times)	1080.24 ± 1070.69	Trunk	
Snoring index (times/h)	230.13 ± 221.6	Fat mass (kg)	12.51 ± 5.81
Sleep quality index (times/h)		Fat-free mass (kg)	27.49 ± 4.99
AHI	30.03 ± 24.77	Muscle mass (kg)	26.12 ± 4.83
ODI	24.21 ± 24.48	Fat percent (%)	30.34 ± 9.76
ArI	21.66 ± 14.92	Body water	
OSA severity (*N*, %)		TBW (kg)	36.65 ± 7.15
Normal	388 (10.99%)	ECW (kg)	15.09 ± 2.41
Mild	818 (23.18%)	ICW (kg)	21.55 ± 4.91
Moderate	933 (26.44%)	Body water percent (%)	49.7 ± 5.72

(continued)

Table 1. Continued.

Variable	Mean ± SD	Variable	Mean ± SD
Severe	1390 (39.39%)	ECW/ICW ratio	0.71 ± 0.08

Data are expressed as mean ± sSD.
AHI, apnea–hypopnea index; ArI, arousal index; BMI, body mass index; BMR, basal metabolic rate; ECW, extracellular water; ICW, intracellular water; NREM, non-rapid eye movement; ODI, oxygen desaturation index for ≥ 3%; REM, rapid eye movement; SpO_2, peripheral capillary oxygen saturation; TBW, total body water; WASO, wake time after sleep onset.

Table 2. Spearman's correlation coefficients of anthropometric measures (*N* = 3529) with the apnea–hypopnea index (AHI) and oxygen desaturation index (ODI) of ≥3%.

	Correlation coefficient	
Variable	AHI (times/h)	ODI (times/h)
Baseline characteristics		
Age (years)	0.11 **	0.08 **
Sex (male/female)	0.26 **	0.25 **
BMI (kg/m^2)	0.51 **	0.53 **
Neck circumference (NC) (cm)	0.52 **	0.52 **
Waist circumference (WC) (cm)	0.55 **	0.56 **
Snoring details		
Snoring events (times)	0.38 **	0.41 **
Snoring index (times/h)	0.42 **	0.43 **
Anthropometric measures		
Whole body		
Fat mass (kg)	0.41 **	0.42 **
Fat-free mass (kg)	0.42 **	0.41 **
Muscle mass (kg)	0.41 **	0.42 **
Bone mass (kg)	0.41 **	0.41 **
Fat percent (%)	0.19 **	0.22 **
Muscle percent (%)	−0.18 **	−0.21 **
Visceral fat level	0.56 **	0.56 **
BMR (kJ)	0.42 **	0.43 **
Physique rating	−0.34 **	−0.36 **

(continued)

Table 2. Continued.

Variable	Correlation coefficient	
	AHI (times/h)	ODI (times/h)
Body fat ratio (limbs/trunk)	−0.15 **	−0.13 **
Limbs		
Fat mass (kg)	0.38 **	0.41 **
Fat-free mass (kg)	0.42 **	0.42 **
Muscle mass (kg)	0.42 **	0.42 **
Fat percent (%)	0.11 **	0.14 **
Trunk		
Fat mass (kg)	0.41 **	0.43 **
Fat-free mass (kg)	0.34 **	0.33 **
Muscle mass (kg)	0.34 **	0.33 **
Fat percent (%)	0.25 **	0.28 **
Body water		
TBW (kg)	0.43 **	0.43 **
ECW (kg)	0.5 **	0.51 **
ICW (kg)	0.38 **	0.38 **
Body water percent (%)	−0.2 **	−0.23 **
ECW/ICW ratio	−0.04 *	−0.04 *

Data are expressed as coefficients.
BMI, body mass index; BMR, basal metabolic rate; ECW, extracellular water; ICW, intracellular water; TBW, total body water.
*$p < 0.05$; **$p < 0.01$.

OSA, XGBoost presented the highest accuracy by employing baseline details and snoring events (72.15% ± 3.38%) or anthropometrics and snoring events (72.44% ± 3.37%), while the RF showed the highest accuracy in the model that employed anthropometrics (70.22% ± 3.45%) or snoring events (69.57% ± 1.99%). Next, Supplementary Table S6 documents the outcomes of predicting the test set using the models which had the best performance in the training and validation stage. Accuracies in screening for the risk of having moderate-to-severe OSA ranged from 74.36% to 78.11%, while accuracies in screening for the risk of having severe OSA ranged from 65.63% to 72.44%. Notably, in the classification performance metrics using the test dataset, models which only considered snoring events demonstrated the relatively lowest accuracy.

Discussion

Principal findings

Considering the limitations of PSG and other proposed surrogates, the present study established machine learning models for screening the risks of moderate-to-severe OSA and severe OSA based on easily accessible parameters. Specifically, we collected snoring events of PSG, baseline characteristics, and anthropometric measures from 3529 patients, examined their correlations, and further developed OSA risk screening models. The major outcomes were that optimized RF models manifested the best classification performances, with accuracies exceeding 75% and AUROCs exceeding 80% for the two severity type models in the

Table 3. Comparison of the accuracy of screening moderate-to-severe and severe obstructive sleep apnea (OSA) in a grid search cross-validation of models using the training and validation datasets (N=2823).

Variable (%)	RF	XGBoost	kNN	SVM	NB	LR
Moderate-to-severe OSA model	AHI ≥15 (N = 1848); AHI <15 (N = 975)					
Precision	79.52 ± 0.98	80.5 ± 1.52	77.88 ± 1.89	79.24 ± 1.51	80.92 ± 1.41	80.27 ± 1.23
Recall	89.83 ± 1.62	88.26 ± 2.04	81.76 ± 2.78	88.2 ± 1.91	71.1 ± 2.29	87.07 ± 2.18
Accuracy	80.18 ± 1.14	78.29 ± 1.44	72.83 ± 2.32	77.12 ± 1.45	70.1 ± 1.71	76.51 ± 1.53
F1-score	85.35 ± 0.85	84.18 ± 1.06	79.75 ± 1.83	83.46 ± 1.05	75.68 ± 1.58	82.51 ± 1.22
AUROC	83.08 ± 2.03	82.18 ± 1.78	75.28 ± 2.89	80.65 ± 1.66	77.54 ± 1.67	80.9 ± 1.76
Severe OSA model	AHI ≥30: (N = 1093); AHI <30: (N = 1730)					
Precision	68.64 ± 3.67	69.22 ± 3.08	67.76 ± 3.79	69.82 ± 5.55	60.99 ± 4.42	68.58 ± 5.12
Recall	66.67 ± 3.9	63.22 ± 4.14	43.0 ± 4.41	58.29 ± 4.29	66.51 ± 3.59	59.84 ± 4.55
Accuracy	76.96 ± 2.55	74.85 ± 2.09	70.0 ± 1.91	74.0 ± 3.03	70.46 ± 3.39	73.79 ± 3.18
F1-score	67.64 ± 3.55	66.02 ± 3.08	52.5 ± 3.97	63.45 ± 4.16	63.59 ± 3.72	63.86 ± 4.41
AUROC	82.82 ± 2.37	81.19 ± 1.8	73.48 ± 1.66	78.74 ± 2.58	76.7 ± 2.83	80.53 ± 2.68

Data are expressed as the mean ± SD.
AHI: apnea–hypopnea index; AUROC, area under the receiver operating characteristics curve; kNN, k-nearest neighbor; LR, logistic regression; NB, naive Bayes; RF, random forest; SVM, support vector machine; XGBoost, extreme gradient boosting

Table 4. Classifications of the results of the random forest (RF) model for screening moderate-to-severe and severe risk of obstructive sleep apnea (OSA) using the test dataset (N = 706).

Variable (%)	Moderate-to-severe OSA model	Severe OSA model
	AHI ≥15 (N = 475); AHI <15 (N = 231)	AHI ≥30 (N = 297); AHI <30 (N = 409)
Precision	80.63 ± 2.92	70.88 ± 3.35
Recall	91.16 ± 2.09	62.29 ± 3.58
Accuracy	79.32 ± 2.99	74.37 ± 3.22
F1-score	85.57 ± 2.59	66.31 ± 3.49
AUROC	82.76 ± 2.79	82.84 ± 2.78

AHI, apnea–hypopnea index; AUROC, area under the receiver operating characteristics curve.
Data are expressed as the mean ± SD.

training and validation phases. Next, using the independent test dataset, the optimized RF models still presented accuracies of over 70% and AUROCs of over 80%. Regarding feature importance, snoring events and the visceral fat level were the most and second most crucial factors for screening for OSA risk. Similarly, significant correlations of baseline characteristics, snoring details, and anthropometric measures with the AHI were observed.

Model performances

Regarding the classification performance, optimized RF models showed the highest accuracy, F1-score, and AUROC values in screening for OSA risks of the two severity types, and XGBoost demonstrated the second highest values. Although there are no direct references revealing that these two approaches outperformed other supervised machine learning techniques, some potential reasons may partially explain these outcomes. First, both the RF and XGBoost were constructed using the ensemble learning theory. This theory aims to enhance predictive performances by averaging predictions from decision trees trained by various subsamples from a dataset. In other words, due to the structure of ensemble learning models, which can guard against the effects of irrelevant features but be more sensitive to relevant ones; they also have an anti-noise capability, and this may allow prediction results

to appropriately converge.[33] Moreover, other techniques, namely bootstrapping and setting the maximum amount of the basic unit of ensemble learning models, can encourage increased diversity in the created trees, and this straightforwardly optimizes model stability and prevents the risk of overfitting.[34] These ensemble learning approaches, thus, have been widely utilized to assist diagnosis and decision-making in different medical fields.[30,35]

Performance comparisons of OSA risk predictions between ensemble learning approaches

This study determined the relatively highest accuracy in predicting the risk of OSA by utilizing ensemble learning approaches (i.e. RF and XGBoost). Previous researchers employed a different ensemble learning method, namely LSBoost, based on oximetry parameters to predict different OSA severities in three datasets and demonstrated adequate accuracies, which ranged from 81.1% to 96.6%.[36] A relevant study developed AdaBoost based on oximetry parameters to screen for the risk of having OSA of different severities and demonstrated a 92.9% accuracy for AHI values of ≥5, an 87.4% accuracy for AHI values of ≥15, and a 78.7% accuracy for AHI values of ≥30.[37] Another relative study used RF models with oxygen signals during sleeping of two datasets and achieved acceptable accuracies (for AHI values of ≥15: 87% to 88.2% and for AHI values of ≥30: 93.2% to 94.3%).[38] First, it should be noted that in the present outcomes, better performances were determined in the moderate-to-severe OSA models compared to values in the severe OSA models in both the training and test phases. Although it is uncertain whether the accuracy of models using physiological signals was higher in models screening for severe OSA or moderate-to-severe OSA, several potential reasons may explain the current observations. First, nonlinear associations of baseline values, anthropometrics, and snoring events with OSA severity may affect the performance of the screening models. Specifically, snoring events and the visceral fat level were shown to be the most critical indicators for screening for the risk of OSA severity in the current models. However, in Supplementary Table S1, significant but very weak correlations between snoring events and the AHI were observed in patients who had severe OSA (with AHI values of ≥30), whereas relatively stronger correlations were determined in the moderate-to-serve group (with AHI values of ≥15). For the visceral fat level, relatively higher correlation coefficients were also observed in the moderate-to-severe group compared to values in the severe group. These may partially allude to the fact that although snoring and anthropometric variables are associated with OSA severity, these parameters may likely be more predictable or functional for screening for moderate-to-severe OSA risk compared to screening for severe OSA risk. In addition, the distribution of OSA severity levels in the collected dataset and the sample size may partially have affected accuracy differences. A prior related study employed outcomes of a subjective questionnaire regarding snoring events and baseline characteristics, including the NC, obesity level, age, and sex, to establish classification approaches for screening for different severities of OSA.[39] Their results demonstrated a relatively higher AUROC value in screening for the risk of moderate-to-severe OSA model compared to the value of the severe OSA model. Next, based on the aforementioned previous outcomes and the present findings, it seems that ensemble learning approaches established based on OSA-associated parameters, such as body profiles, anthropometrics, snoring, or oximetry parameters, can be applied to screen for the risk of this disease. However, it should be noted that the accuracy range of current models developed with snoring details and anthropometrics was relatively lower than the aforesaid previous outcomes, trained with oximetry parameters. Several reasons may account for these differences, such as the nature of different demographics in collected data, sample sizes, or the fat distribution issue. In other words, individuals who have a narrow upper airway anatomy or excessive deposition of fat at the base of the tongue may present with snoring but without exhibiting OSA. Epidemiologically, 94% of patients with OSA exhibited snoring, which was seen as the most common manifestation of OSA.[40] However, another study that recruited 602 participants who were habitual snorers observed that 61% of females and 81% of males did not have OSA.[41] Another study observed a significant but weak correlation between the snoring index and AHI.[42] Collectively, to enhance the accuracy of screening for OSA risks, it may be worthwhile employing data with more comprehensive dimensions, such as baseline values, anthropometrics, snoring details, and oximetry parameters.

Feature importance for predicting the risk of having OSA using RF

Concerning feature importance, snoring events and the visceral fat level were the most and second most important features when screening for OSA risks of both severity types. Subsequently, age, NC, WC, BMI, TBW, ECW, and the ECW-to-ICW ratio presented the third- to ninth highest Shapely values in the screening risk models for both types of OSA severity. Such a distribution of Shapley values was also in accordance with our correlation outcomes among anthropometric parameters, snoring details, and the AHI. First, snoring has long been considered an indicator of higher upper airway resistance and the most common symptom suggestive of OSA.[43] Past research proposed only monitoring the volume and time of snoring to diagnose moderate-to-severe OSA, which showed adequate accuracy.[18] Thus, our results partially aligned with previous

outcomes that snoring events can be a predictor for screening for OSA risk. In terms of baseline characteristics and anthropometric measures, aging effects and adipose deposition may be used to interpret the present results. More precisely, visceral fat (internal abdominal cavity), WC (abdomen), NC (upper airway), and BMI (whole body), can all be employed to partially represent the obesity level, which is associated with OSA severity.[44] On the other hand, aging, as a proxy for decreased muscle strength, may cause the elderly to tend to have more severe OSA compared to younger persons.[45] As to the body fluid distribution, research has documented that increased nocturnal rostral fluid, shifting from the limbs or due to excessive TBW, may elevate the mucosal water content in the upper airway and cause tract narrowing, thereby aggravating the risk of OSA.[46] In addition, ECW is relevant to residual kidney function and can serve as an indicator of body fluid drainage.[47] Several prior findings were partially similar to the current observations. For example, researchers who investigated fat accumulation in different body regions revealed that an accumulated adipose volume was related to the risk of OSA.[48,49] Similarly, relevant research suggests that increased BMI and WC values are essential to the risk of OSA.[50] A study indicated the possibility of employing BMI and NC to identify the risk of exhibiting OSA in different sexes.[51] A systematic review exhibited results that conformed to our outcomes that both the BMI and age increased the OSA risk.[52] As to body fluid, a study found an increased ECW volume in OSA patients compared to those without OSA.[53] Other research reported similar observations of higher average values of the TBW-to-ECW ratio in patients with OSA.[54] Taken together, the current outcomes may elucidate the feasibility of using the currently developed models based on these easily assessed variables to screen for OSA risks of different severities (AHI $\geq$15 times/h or AHI $\geq$30 times/h).

Strengths and limitations

This study has some strengths and limitations. First, the established models conducted OSA risk screening by employing characteristics, anthropometric parameters, and snoring details. Compared to approaches that only analyze snoring sounds, the current models may be more likely to be suitable for non-snorers with OSA.[55] Some researchers proposed age- and sex-dependent models that trained using only baseline characteristics (i.e. WC, NC, and BMI).[56] However, detailed investigations or higher-dimensional parameters in anthropometrics may serve as useful indicators to provide further information regarding the risk of having OSA. The classification results in the present supplementary, derived by employing various input feature bundles, may also indicate that anthropometrics provide essential and complementary information for screening the risk of presenting OSA. In addition, a relevant study developed machine learning approaches for OSA severity classification based on scanned craniofacial feature images, but the accuracy of those approaches was 67% for predicting the risk of moderate-to-severe OSA.[57] Variations in obtained craniofacial images (due to breathing movements or muscle tone) may have affected the accuracy. Also, OSA presentation might not completely be caused by craniofacial factors. Altogether, the current machine learning models developed based on snoring events, baseline characteristics, and anthropometric measures may be applicable for screening for OSA risks. As to limitations, first, the presently developed models were population specific since we only employed data from a single sleep center in Taiwan. This may limit the generalization ability of the developed models for application to other ethnicities. More precisely, some features such as craniofacial features, that vary among different ethnicities, may affect the AHI and OSA severity.[58] Other lifestyle habits, such as tobacco and alcohol usage[59] or one's personal health status (menopausal status, medication use, and comorbidities),[60,61] may also affect OSA severity. Some researchers proposed machine learning approaches by employing nocturnal oxygen saturation obtained by pulse oximetry and electrocardiogram signals for screening for OSA risk.[62] The relatively high accuracies of those outcomes imply the prospect of improving model performance by considering the oximetry and cardio-related parameters. However, the absence of these data limited the possibility of improving the performance in the current work, which should be addressed in future studies. In addition, there was no measurement of body position, and it is accepted that snoring and the AHI are very dependent on body position.[63] Thus, although the current models were developed using anthropometric measures (overall body adipose tissue evaluation) and snoring events (predictors of obstruction of the upper airway), they might only be applicable to specific ethnicities or populations with craniofacial structures similar to those of the Taiwanese population. Further studies can consider obtaining such comprehensive dimensional data to increase the accuracy of screening for OSA risks. Snoring events of this study were determined using piezoelectric signals. The piezoelectric sensor, placed on the neck area of individuals, may cause inconvenience and discomfort and thereby potentially result in the occurrence of artifacts. Next, the current classification target, namely OSA severity, was based on the AHI, but these data need to be manually scored by different PSG technologists, which may be affected by inter-scorer variability.[64] Despite the fact that inter-scorer training is regularly performed among PSG technologists to reduce scoring variability, a partial degree of variability may still be inescapable. Environmental factors (e.g. the first night effect), may be crucial factors in affecting the reliability of PSG results.[65] In other words, sleeping in a hospital environment may change participants' architecture and

depth of sleep during PSG, which would indirectly impact PSG outcomes.[66] Although the present study excluded participants whose sleep efficiency was low (≥ 40%), repeated PSG for preventing such biases may still need to be considered. In addition, future studies should consider using 3 h as the exclusion threshold, which is the most common minimum threshold of total sleep time, to enhance the representativeness of the PSG data of patients.

Conclusions

Avoiding the drawbacks of PSG and the requirements of OSA risk screening tools, this retrospective study aimed to develop classification models based on easily acquired variables. Using baseline characteristics, anthropometric measures, and snoring data derived from 3529 patients in Taiwan, various machine learning models for screening the risk of having severe-to-moderate OSA and severe OSA were established. RF models, which manifested the best classification performance in the training and validation phases, were employed to predict the independent test set. In both OSA severity screening RF models, over 70% accuracy and over 80% AUROC were achieved. As to feature importance values, snoring events, and the visceral fat level were the most and second most essential features of screening for OSA risk. Likewise, the current statistical outcomes indicated that anthropometric measures and snoring details were significantly correlated with the AHI and ODI. Collectively, the present machine learning approaches developed based on snoring events, baseline characteristics, and anthropometric measures can be considered for OSA risk screening.

Acknowledgments: The authors would like to express our appreciation to the technologists in the sleep center of Shuang Ho Hospital for collecting the raw data, and we thank all participants for their contribution to this research.

Data availability statement: The data used for this study, although not available in a public repository, will be made available to other researchers upon reasonable request.

Contributorship: Cheng-Yu Tsai, Wen-Te Liu, and Arnab Majumdar conceptualized and designed the study. Cheng-Yu Tsai, Wen-Hua Hsu, and Yi-Chun Kuan researched the literature and drafted the article. Cheng-Yu Tsai, Marc Stettler, Wun-Hao Cheng, and Cheng-Jung Wu performed data curation and analyzed the datasets. Wen-Hua Hsu, Dean Wu, and Yi-Chih Lin gained ethical approval. Shu-Chuan Ho and Hsin-Chien Lee critically revised the manuscript. Shu-Chuan Ho and Kang-Yun Lee approved the version to be published.

Declaration of conflicting interests: The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Ethical approval: The Ethics Committees of Taipei Medical University Joint Institutional Review Board (No. N201911007) reviewed and approved the study protocol.

Funding: The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported by the Taiwan Ministry of Science and Technology (grant numbers MOST 111-2314-B-038-090-MY3, MOST 111-2314-B-038-082, and MOST 110-2634-F-002–049), Taiwan National Science and Technology Council (grant number NSTC 111-2634-F-002-021), and the Taiwan Ministry of Education (grant number DP2-111-21121-01-T-01-05).

Guarantor: S-C Ho.

ORCID iDs: Yi-Chih Lin https://orcid.org/0000-0002-7292-6879
Shu-Chuan Ho https://orcid.org/0000-0002-1616-8647

Supplemental material: Supplemental material for this article is available online.

References

1. Young T, Peppard PE and Gottlieb DJ. Epidemiology of obstructive sleep apnea: a population health perspective. *Am J Respir Crit Care Med* 2002; 165: 1217–1239.
2. Peppard PE, Young T, Barnet JH, et al. Increased prevalence of sleep-disordered breathing in adults. *Am J Epidemiol* 2013; 177: 1006–1014.
3. Young T, Finn L, Peppard PE, et al. Sleep disordered breathing and mortality: eighteen-year follow-up of the wisconsin sleep cohort. *Sleep* 2008; 31: 1071–1078.
4. Calhoun DA and Harding SM. Sleep and hypertension. *Chest* 2010; 138: 434–443.
5. Nieto FJ, Peppard PE, Young T, et al. Sleep-disordered breathing and cancer mortality: results from the wisconsin sleep cohort study. *Am J Respir Crit Care Med* 2012; 186: 190–194.
6. Brzecka A, Leszek J, Ashraf GM, et al. Sleep disorders associated with Alzheimer's disease: a perspective. *Front Neurosci* 2018; 12: 330. DOI: 10.3389/fnins.2018.00330
7. Flemons WW and Tsai W. Quality of life consequences of sleep-disordered breathing. *J Allergy Clin Immunol* 1997; 99: S750–S756. DOI: 10.1016/s0091-6749(97)70123-4
8. Chesson ALJr, Ferber RA, Fry JM, et al. The indications for polysomnography and related procedures. *Sleep* 1997; 20: 423–487.
9. Quan S, Gillin JC, Littner M, et al. Sleep-related breathing disorders in adults: recommendations for syndrome definition and measurement techniques in clinical research editorials. *Sleep (New York, NY)* 1999; 22: 662–689.
10. Ghaemmaghami H, Abeyratne UR and Hukins C. Normal probability testing of snore signals for diagnosis of obstructive sleep apnea. In: Conference proceedings : Annual International Conference of the IEEE Engineering in Medicine and Biology Society IEEE Engineering in

Medicine and Biology Society Annual Conference. 2009; pp. 5551–5554. DOI: 10.1109/IEMBS.2009.5333733
11. Masa JF, Duran-Cantolla J, Capote F, et al. Effectiveness of home single-channel nasal pressure for sleep apnea diagnosis. *Sleep* 2014; 37: 1953–1961.
12. Nigro CA, Castaño G, Bledel I, et al. A novel, simple, and accurate pulse oximetry indicator for screening adult obstructive sleep apnea. *Sleep Breathing* 2021: 1–10. DOI: 10.1007/s11325-021-02439-4
13. Rosenberg R, Hirshkowitz M, Rapoport DM, et al. The role of home sleep testing for evaluation of patients with excessive daytime sleepiness: focus on obstructive sleep apnea and narcolepsy. *Sleep Med* 2019; 56: 80–89.
14. Ibáñez V, Silva J and Cauli O. A survey on sleep questionnaires and diaries. *Sleep Med* 2018; 42: 90–96.
15. Nerfeldt P, Aoki F and Friberg D. Polygraphy vs. Polysomnography: missing osas in symptomatic snorers—a reminder for clinicians. *Sleep Breathing* 2014; 18: 297–303.
16. Zeidler MR, Santiago V, Dzierzewski JM, et al. Predictors of obstructive sleep apnea on polysomnography after a technically inadequate or normal home sleep test *J Clin Sleep Med* 2015; 11: 1313–1318.
17. El-Sayed IH. Comparison of four sleep questionnaires for screening obstructive sleep apnea. *Egyptian. J Chest Diseases and Tuberculosis* 2012; 61: 433–441.
18. Nakano H, Hirayama K, Sadamitsu Y, et al. Monitoring sound to quantify snoring and sleep apnea severity using a smartphone: proof of concept. *J Clin Sleep Med* 2014; 10: 73–78.
19. Hoffstein V, Mateika S and Anderson D. Snoring: is it in the ear of the beholder? *Sleep* 1994; 17: 522–526.
20. Ben-Israel N, Tarasiuk A and Zigel Y. Obstructive apnea hypopnea index estimation by analysis of nocturnal snoring signals in adults. *Sleep* 2012; 35: 1299–305C.
21. Alakuijala A and Salmi T. Predicting obstructive sleep apnea with periodic snoring sound recorded at home. *J Clin Sleep Med* 2016; 12: 953–958.
22. Sowho M, Sgambati F, Guzman M, et al. Snoring: a source of noise pollution and sleep apnea predictor. *Sleep* 2020; 43: zsz305.
23. Öğretmenoğlu O, Süslü AE, Yücel ÖT, et al. Body fat composition: a predictive factor for obstructive sleep apnea. *Laryngoscope* 2005; 115: 1493–1498.
24. Gabbay IE and Lavie P. Age-and gender-related characteristics of obstructive sleep apnea. *Sleep Breathing* 2012; 16: 453–460.
25. Berry RB, Albertario CL, Harding SM, et al. The AASM manual for the scoring of sleep and associated events: rules, terminology and technical specifications. *Am Acad Sleep Med* 2018.
26. Patil SP, Ayappa IA, Caples SM, et al. Treatment of adult obstructive sleep apnea with positive airway pressure: an American academy of sleep medicine clinical practice guideline. *J Clin Sleep Med* 2019; 15: 335–343.
27. Tsai CY, Huang HT, Cheng HC, et al. Screening for obstructive sleep apnea risk by using machine learning approaches and anthropometric features. *Sensors (Basel)* 2022; 22: 8630. DOI: 10.3390/s22228630
28. Towards validating the effectiveness of obstructive sleep apnea classification from electronic health records using machine learning. In: Ramesh J, Keeran N, Sagahyroon A and Aloul F (eds) *Healthcare*. MDPI, 2021.
29. Pontes FJ, Amorim G, Balestrassi PP, et al. Design of experiments and focused grid search for neural network parameter optimization. *Neurocomputing* 2016; 186: 22–34.
30. Kim YJ, Jeon JS, Cho S-E, et al. Prediction models for obstructive sleep apnea in Korean adults using machine learning techniques. *Diagnostics* 2021; 11: 12.
31. Tsai C-Y, Liu W-T, Lin Y-T, et al. Machine learning approaches for screening the risk of obstructive sleep apnea in the Taiwan population based on body profile. *Inf Health Soc Care* 2021: 1–16. DOI: 10.1080/17538157.2021.2007930
32. The many Shapley values for model explanation. In: Sundararajan M and Najmi A (eds) *International conference on machine learning*. PMLR, 2020.
33. Breiman L. Random forests. *Mach Learn* 2001; 45: 5–32.
34. Zhou J, Li E, Wei H, et al. Random forests and cubist algorithms for predicting shear strengths of rockfill materials. *Appl Sci* 2019; 9: 1621.
35. Tsai C-Y, Kuan Y-C, Hsu W-H, et al. Differentiation model for insomnia disorder and the respiratory arousal threshold phenotype in obstructive sleep apnea in the Taiwanese population based on oximetry and anthropometric features. *Diagnostics* 2021; 12: 50.
36. Gutiérrez-Tobal GC, Álvarez D, Vaquerizo-Villar F, et al. Ensemble-learning regression to estimate sleep apnea severity using at-home oximetry in adults. *Appl Soft Comput* 2021; 111: 107827.
37. Gutierrez-Tobal GC, Alvarez D, Crespo A, et al. Evaluation of machine-learning approaches to estimate sleep apnea severity from at-home oximetry recordings. *IEEE J Biomed Health Inform* 2019; 23: 882–892.
38. Deviaene M, Testelmans D, Buyse B, et al. Automatic screening of sleep apnea patients based on the SpO(2) signal. *IEEE J Biomed Health Inform* 2019; 23: 607–617.
39. Coutinho Costa J, Rebelo-Marques A, Machado JN, et al. Validation of NoSAS (neck, obesity, snoring, age, sex) score as a screening tool for obstructive sleep apnea: analysis in a sleep clinic. *Pulmonology* 2019; 25: 263–270.
40. Mattei A, Tabbia G and Baldi S. Diagnosis of sleep apnea. *Minerva Med* 2004; 95: 213–231.
41. Young T, Palta M, Dempsey J, et al. The occurrence of sleep-disordered breathing among middle-aged adults. *N Engl J Med* 1993; 328: 1230–1235.
42. Alshaer H, Hummel R, Mendelson M, et al. Objective relationship between sleep apnea and frequency of snoring assessed by machine learning. *J Clin Sleep Med* 2019; 15: 463–470.
43. Koo SK, Kwon SB, Kim YJ, et al. Acoustic analysis of snoring sounds recorded with a smartphone according to obstruction site in OSAS patients. *Eur Arch Oto-Rhino-Laryngol* 2017; 274: 1735–1740.
44. Carter RIII and Watenpaugh DE. Obesity and obstructive sleep apnea: or is it OSA and obesity? *Pathophysiology* 2008; 15: 71–77.
45. Leppänen T, Töyräs J, Mervaala E, et al. Severity of individual obstruction events increases with age in patients with obstructive sleep apnea. *Sleep Med* 2017; 37: 32–37.
46. White L, Bradley T and Logan A. Pathogenesis of obstructive sleep apnoea in hypertensive patients: role of fluid retention and nocturnal rostral fluid shift. *J Hum Hypertens* 2015; 29: 342–350.

47. Lanis A, Kerns E, Hu SL, et al. Residual renal function affects severity of sleep apnea in peritoneal dialysis: a pilot study. *Lung* 2018; 196: 425–431.
48. Mitra AK, Bhuiyan AR and Jones EA. Association and risk factors for obstructive sleep apnea and cardiovascular diseases: a systematic review. *Diseases* 2021; 9: 88.
49. Shah N and Roux F. The relationship of obesity and obstructive sleep apnea. *Clin Chest Med* 2009; 30: 455–465.
50. Wolk R, Shamsuzzaman AS and Somers VK. Obesity, sleep apnea, and hypertension. *Hypertension* 2003; 42: 1067–1074.
51. Topîrceanu A, Udrescu L, Udrescu M, et al. Gender phenotyping of patients with obstructive sleep apnea syndrome using a network science approach. *J Clin Med* 2020; 9: 4025.
52. Senaratna CV, Perret JL, Lodge CJ, et al. Prevalence of obstructive sleep apnea in the general population: a systematic review. *Sleep Med Rev* 2017; 34: 70–81.
53. Sleep apnea is associated with residual kidney function and mortality in patients with peritoneal dialysis: Prospective cohort study. In: Kang SC, Park KS, Chang TI, Shin SK and Kang EW (eds) *Seminars in Dialysis*. Wiley Online Library, 2022.
54. Kosacka M, Korzeniewska A and Jankowska R. The evaluation of body composition, adiponectin, C-reactive protein and cholesterol levels in patients with obstructive sleep apnea syndrome. *Adv Clin Exp Med* 2013; 22: 817–824.
55. Wang B, Yi X, Gao J, et al. Real-time prediction of upcoming respiratory events via machine learning using snoring sound signal. *J Clin Sleep Med* 2021; 17: 1777–1784.
56. Liu WT, Wu HT, Juang JN, et al. Prediction of the severity of obstructive sleep apnea by anthropometric features via support vector machine. *PLoS One* 2017; 12: e0176991.
57. Hanif U, Leary EB, Schneider LD, et al. Estimation of apnea-hypopnea index using deep learning on 3-D craniofacial scans. *IEEE J Biomed Health Inform* 2021; 25: 4185–4194.
58. Costa ESRA and dos Santos Gil NA. Craniofacial skeletal architecture and obstructive sleep apnoea syndrome severity. *J Cranio-Maxillofac Surg* 2013; 41: 740–746.
59. Duan X, Huang J, Zheng M, et al. Association of healthy lifestyle with risk of obstructive sleep apnea: a cross-sectional study. *BMC Pulm Med* 2022; 22: 1–12.
60. Gleeson M and McNicholas WT. Bidirectional relationships of comorbidity with obstructive sleep apnoea. *Eur Respir Rev* 2022; 31
61. Perger E, Mattaliano P and Lombardi C. Menopause and sleep apnea. *Maturitas* 2019; 124: 35–8. DOI: 10.1016/j.maturitas.2019.02.011
62. Zhu J, Zhou A, Gong Q, Zhou Y, Huang J and Chen Z. Detection of sleep apnea from electrocardiogram and pulse oximetry signals using random forest. *Appl Sci* 2022; 12(9): 4218. DOI: doi:10.3390/app12094218
63. Nakano H, Ikeda T, Hayashi M, et al. Effects of body position on snoring in apneic and nonapneic snorers. *Sleep* 2003; 26: 169–172.
64. Younes M, Raneri J and Hanly P. Staging sleep in polysomnograms: analysis of inter-scorer variability. *J Clin Sleep Med* 2016; 12: 885–894.
65. Ding L, Chen B, Dai Y, et al. A meta-analysis of the first-night effect in healthy individuals for the full age spectrum. *Sleep Med* 2022; 89: 159–165.
66. Newell J, Mairesse O, Verbanck P, et al. Is a one-night stay in the lab really enough to conclude? First-night effect and night-to-night variability in polysomnographic recordings among different clinical population samples. *Psychiatry Res* 2012; 200: 795–801.