

Original Article
Otorhinolaryngology



A Classifying Model of Obstructive Sleep Apnea Based on Heart Rate Variability in a Large Korean Population

Pona Park ^{1,2} and Jeong-Whun Kim ^{1,3,4}

¹Seoul National University College of Medicine, Seoul, Korea

²Department of Otorhinolaryngology-Head and Neck Surgery, National Police Hospital, Seoul, Korea

³Sensory Organ Research Institute, Seoul National University Medical Research Center, Seoul, Korea

⁴Department of Otorhinolaryngology-Head and Neck Surgery, Seoul National University Bundang Hospital, Seongnam, Korea



Received: Aug 16, 2022

Accepted: Nov 16, 2022

Published online: Feb 6, 2023

Address for correspondence:

Jeong-Whun Kim, MD, PhD

Seoul National University College of Medicine, Sensory Organ Research Institute, Seoul National University Medical Research Center, Department of Otorhinolaryngology-Head and Neck Surgery, Seoul National University Bundang Hospital, 82 Gumi-ro 173-beon-gil, Bundang-gu, Seongnam 13620, Republic of Korea.

Email: kimemails7@gmail.com

© 2023 The Korean Academy of Medical Sciences.

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<https://creativecommons.org/licenses/by-nc/4.0/>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

ORCID iDs

Pona Park

<https://orcid.org/0000-0002-3071-1730>

Jeong-Whun Kim

<https://orcid.org/0000-0003-4858-3316>

Funding

This research was supported by the National Police Hospital and the Seoul National University Bundang Hospital (grant No. 14-2022-0036, 2022).

Disclosure

The authors have no potential conflicts of interest to disclose.

<https://jkms.org>

ABSTRACT

Background: The majority of patients with obstructive sleep apnea do not receive timely diagnosis and treatment because of the complexity of a diagnostic test. We aimed to predict obstructive sleep apnea based on heart rate variability, body mass index, and demographic characteristics in a large Korean population.

Methods: Models of binary classification for predicting obstructive sleep apnea severity were constructed using 14 features including 11 heart rate variability variables, age, sex, and body mass index. Binary classification was conducted separately using apnea-hypopnea index thresholds of 5, 15, and 30. Sixty percent of the participants were randomly allocated to training and validation sets while the other forty percent were designated as the test set. Classifying models were developed and validated with 10-fold cross-validation using logistic regression, random forest, support vector machine, and multilayer perceptron algorithms.

Results: A total of 792 (651 men and 141 women) subjects were included. The mean age, body mass index, and apnea-hypopnea index score were 55.1 years, 25.9 kg/m², and 22.9, respectively. The sensitivity of the best performing algorithm was 73.6%, 70.7%, and 78.4% when the apnea-hypopnea index threshold criterion was 5, 10, and 15, respectively. The prediction performances of the best classifiers at apnea-hypopnea indices of 5, 15, and 30 were as follows: accuracy, 72.2%, 70.0%, and 70.3%; specificity, 64.6%, 69.2%, and 67.9%; area under the receiver operating characteristic curve, 77.2%, 73.5%, and 80.1%, respectively. Overall, the logistic regression model using the apnea-hypopnea index criterion of 30 showed the best classifying performance among all models.

Conclusion: Obstructive sleep apnea was fairly predicted by using heart rate variability, body mass index, and demographic characteristics in a large Korean population. Prescreening and continuous treatment monitoring of obstructive sleep apnea may be possible simply by measuring heart rate variability.

Keywords: Sleep Apnea, Obstructive; Heart Rate; Polysomnography; Machine Learning

INTRODUCTION

Obstructive sleep apnea (OSA) is a disease characterized by repeated obstruction of the upper airway during sleep, leading to intermittent oxygen desaturation and frequent

Author Contributions

Conceptualization: Kim JW. Data curation: Park P, Kim JW. Formal analysis: Park P, Kim JW. Investigation: Park P, Kim JW. Methodology: Park P, Kim JW. Software: Park P. Validation: Kim JW. Writing - original draft: Park P. Writing - review & editing: Park P, Kim JW.

arousal. Patients often experience excessive daytime sleepiness, poor cognitive function, and impaired quality of life.¹ Patients who remain untreated for long periods of time have an increased risk of multiple health-related outcomes, including cardiovascular disease, stroke, and death from all causes.² OSA is a significant and growing public health problem, and is estimated to affect 13–33% and 16–19% of men and women, respectively.³ However, the majority of patients with OSA do not receive timely diagnosis and treatment.⁴

The gold standard for the diagnosis of OSA is polysomnography (PSG). The examinee must spend an entire night in the hospital with a sleep specialist monitoring them while they sleep. Then, the PSG data are scored manually by the specialist according to the American Academy of Sleep Medicine guidelines.⁵ Hence, PSG is not suitable for massive screening and/or repeated measurements of OSA. Considering night-to-night variability,^{6,7} PSG is limited for diagnosis as well as continuous monitoring for effectiveness during treatment.

Various less complicated sleep diagnostic devices have been developed, with advantages of low prices and simple installation. These devices are portable and can be used at home without medical support. Systems developed in the context of OSA diagnosis use nasal airflow and/or oxygen saturation as parameters.^{8,9} However, these measurement tools have some intrinsic limitations since mouth breathing or incorrect positioning can cause signal loss.

Measurable heart rate variability (HRV) from single lead electrocardiogram signals reflects cardiac autonomic activity and can be used to evaluate quantitative changes between normal heartbeats.¹⁰ The association between OSA and increased sympathetic nervous system activity of HRV has been well documented since the clinical significance of HRV was introduced firstly in 1965.^{11,12} Recently, further studies on the correlation between HRV factors and PSG indices have been conducted.^{13,14} Currently, various devices are commercially available that provide heart rate monitoring, such as smartwatches, consumer sleep wearables, and adhesive electrocardiographic monitor patches. Such devices create the possibility of evaluating sleep-disordered breathing status of patients by measuring HRV using a heart rate monitor.

The present study aimed to develop binary classifiers of OSA based on HRV parameters using several machine learning algorithms. This study also aimed to verify the feasibility of using HRV for OSA prescreening and treatment monitoring in a large Korean population.

METHODS

Study participants

We retrospectively reviewed outpatients and inpatients who visited our sleep center owing to snoring or sleep apnea between January 2013 and December 2017. Adult patients, aged ≥ 18 years were included in this study. We excluded patients who underwent split-night PSG. After reviewing apnea-hypopnea index (AHI) scores, we randomly allocated patients to non-OSA and mild, moderate, and severe OSA groups proportionally to reduce selection bias. Then, we excluded subjects with any of the followings: 1) significant arrhythmias; 2) sleep disorders such as insomnia or narcolepsy; 3) inveterate use of sedatives and hypnotics; 4) specific conditions related to HRV changes (i.e., myocardial disease, diabetic neuropathy, or heart transplantation); 5) low quality data (artefacts $> 20\%$ of total sleep time); 6) total sleep time < 5 hours; and 7) awake for > 30 minutes from midnight–5 AM. A flow chart of the participant selection process is summarized in **Fig. 1**. All participants underwent in-laboratory, full-

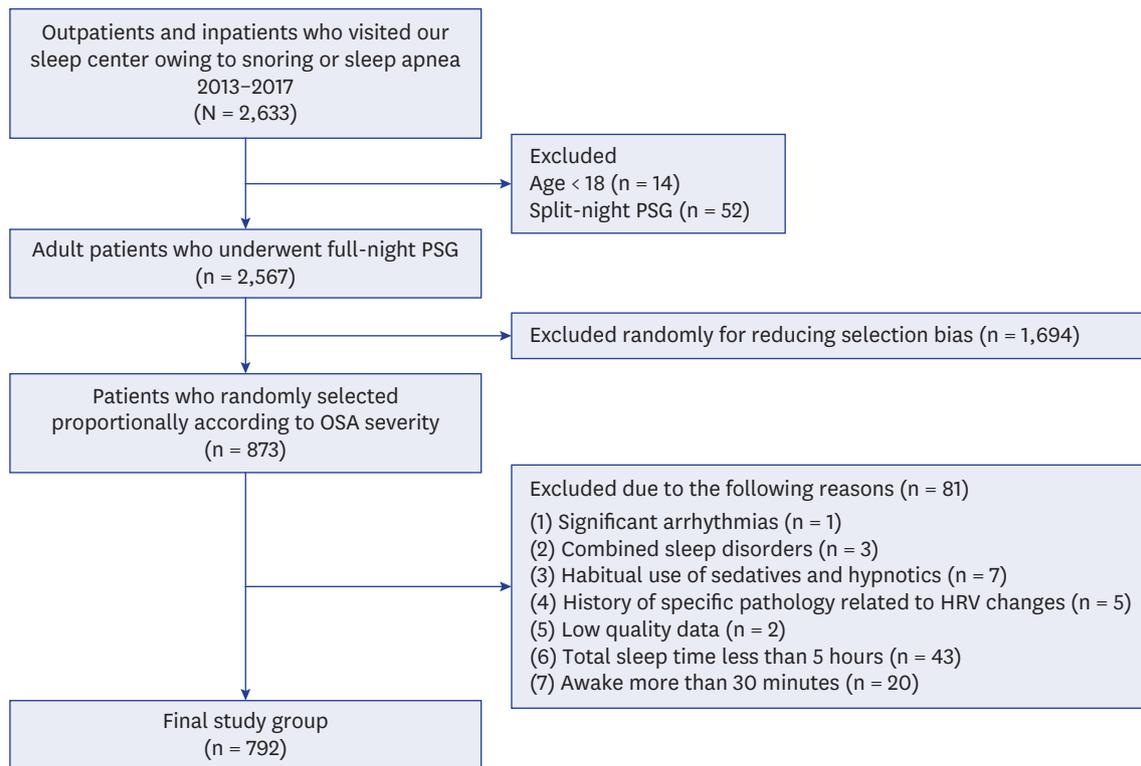


Fig. 1. Flow chart of participant selection process. Data on outpatients and inpatients who visited our sleep center due to snoring or sleep apnea between 2013 and 2017 were collected. Patients who were under 18 years old and underwent split-night polysomnography were excluded. To reduce selection bias, patients were randomly allocated proportionally according to OSA severity. The final study group was selected by excluding additional patients with factors likely to influence the classifying performance for OSA using heart rate variability parameters. HRV = heart rate variability, OSA = obstructive sleep apnea, PSG = polysomnography.

night PSG (Embla® N7000, Neurolite Advanced Medical Solutions, Belp, Switzerland). The anthropometric data on the subjects were reviewed.

HRV analysis

HRV was measured using exported electrocardiogram data and commercially available PSG software (Embla® RemLogic™ 3.0 HRV analyzer; Neurolite Advanced Medical Solutions). This methodology was described in our previous study.¹⁵ Calculations of the time- and frequency-domain parameters were performed according to standard methods for HRV measurements.¹⁰ We only analyzed the electrocardiogram signal from midnight–5 AM to maintain consistency for comparisons. Therefore, the AHI was measured again from midnight–5 AM. The electrocardiogram signals were interpolated and resampled at 5.0 Hz. Normal-to-normal (NN) heartbeat intervals > 2,400 ms and < 400 ms were omitted. We used five time-domain and six frequency-domain measures in the present study. The time-domain parameters were as follows: 1) average NN interval (all intervals between adjacent QRS complexes); 2) standard deviation of the NN interval; 3) square root of the mean of the squared differences of adjacent NN intervals; 4) number of pairs of adjacent NN intervals more than 50 ms, and 5) the rate of pairs of adjacent NN intervals more than 50 ms among the total number of NN intervals. The frequency-domain parameters were as follows: 1) total power (variance of all NN intervals); 2) low frequency (LF; 0.04–0.15 Hz); 3) high frequency (HF; 0.15–0.4 Hz); 4) LF/HF ratio; 5) normalized LF (LF normalized units [nu]; LF / [LF + HF] × 100), and 6) normalized HF (HFnu; HF/[LF + HF] × 100). Frequency-domain measures were yielded as the average of the values calculated every 5 minutes.

The HRV measures according to OSA severity are presented as the mean \pm standard deviation, and were compared by one-way analysis of variance using SPSS software (version 22.0, IBM Corp., Armonk, NY, USA). Turkey, Bonferroni or Dunnett T3 test were used for post-hoc comparisons. *P* values < 0.05 were considered statistically significant.

Development of binary classifiers of OSA

A model of OSA classification was constructed using 14 features and actual class according to AHI threshold (Fig. 2A). Output was the predicted class using estimated AHI scores. The 14 features included age, sex, body mass index (BMI), and the 11 HRV measures. We built the classifying model using logistic regression, random forest, support vector machine, and multilayer perceptron algorithms in a free machine learning software (Weka; University of Waikato, Hamilton, New Zealand).¹⁶ Sixty percent of participants were randomly selected for the training and validation sets while the other forty percent were designated as the test set. The difference between classes in the training set was adjusted using the Synthetic Minority Oversampling Technique in Weka. Validation was conducted by applying 10-fold cross-validation. The enrolled subjects were split randomly into 10 subgroups. One subgroup was preserved for validation of the prediction model. The other nine subgroups were served for the training set. The cross-validation process was repetitively conducted 10 times (10-fold) with all 10 subgroups. This produced 10 evaluation results, which were then averaged. The learning algorithm was then applied a final (11th) time in the whole data set to gain the final model.¹⁶ Ultimately, the test set was included in the final training model (Fig. 2B). Output was expressed as a confusion matrix and area under the receiver operating characteristic curve (AUC).

The performance outcomes of the machine learning classifiers were mainly evaluated by sensitivity, specificity, and accuracy. The AUC was also presented to assess the performance of the classifiers. All results were extracted from the test set.

Ethics statement

The present study protocol was reviewed and approved by the ethics committee of Seoul National University Bundang Hospital (IRB No. B-2111-723-110). The requirement for written informed consent was waived by the ethics committee owing to the retrospective nature of the study.

RESULTS

General patient characteristics and HRV parameters

A total of 792 participants (651 men and 141 women) with a mean age of 55.1 ± 12.6 years were included in the study. The mean BMI and AHI score was 25.9 ± 3.5 kg/m² and 22.9 ± 19.2 , respectively. Participants were classified according to AHI score as follows: 124, normal (AHI < 5); 223, mild OSA ($5 \leq$ AHI < 15); 221, moderate OSA ($15 \leq$ AHI < 30); and 224, severe OSA (AHI ≥ 30). General and polysomnographic characteristics are summarized in Table 1. The means and standard deviations of the 11 HRV parameters according to OSA severity are presented in Table 2. The HRV parameters presented a significant difference according to OSA severity except for square root of the mean of the squared differences of adjacent NN intervals, number of pairs of adjacent NN intervals more than 50 ms, and rate of pairs of adjacent NN intervals more than 50 ms among the total number of NN intervals ($P < 0.05$). Post-hoc analyses were performed on HRV variables showing significant differences in Table 2. The total power, LF, LF/HF ratio, and LFn_u values increased with OSA severity. The average NN interval, HF, and HF_{nu} showed a decreasing tendency as the severity of OSA increased.

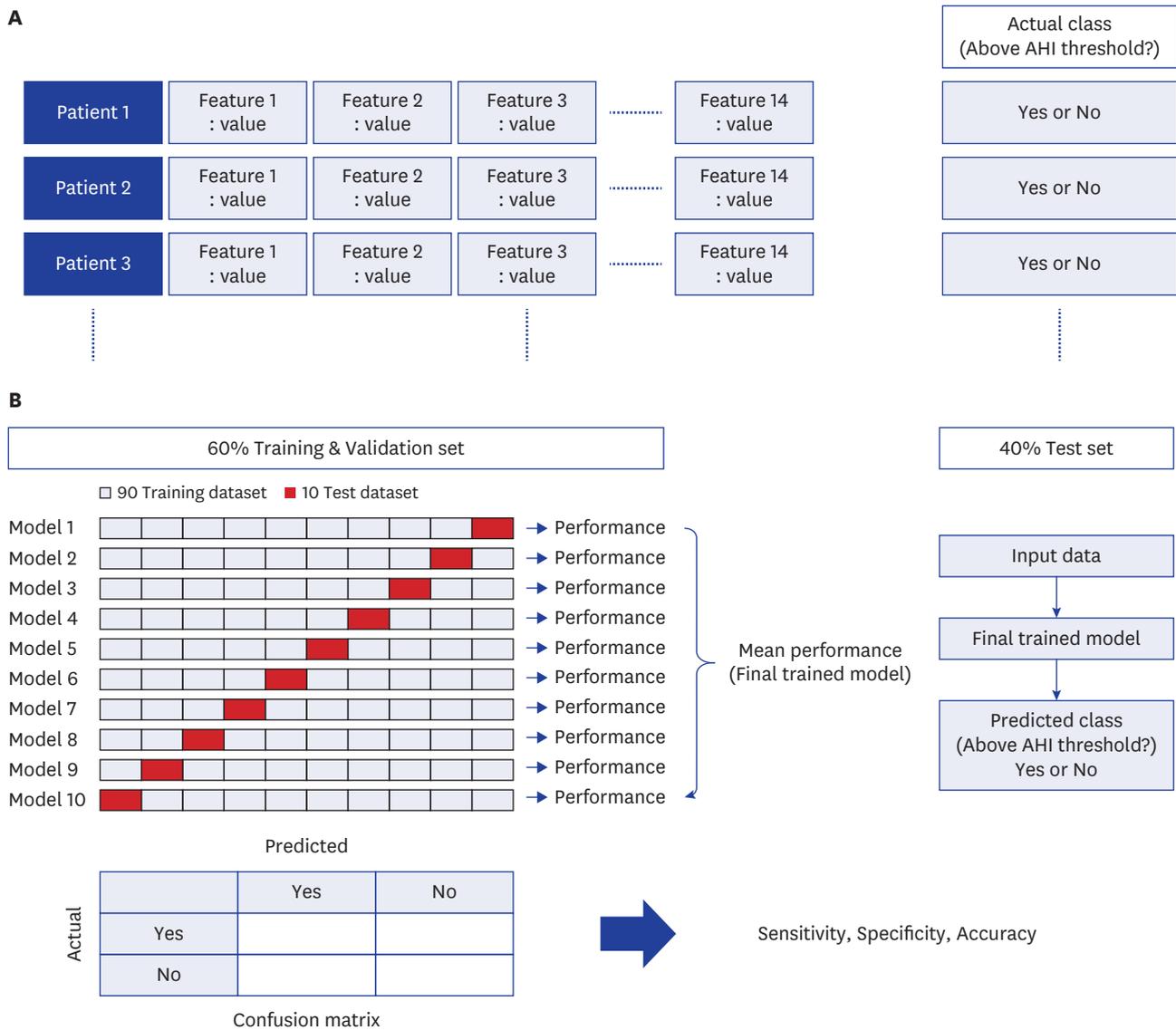


Fig. 2. Study framework of machine learning process. **(A)** Fourteen features including age, sex, body mass index, and 11 measures of heart rate variability were used in the machine learning process for binary classification of obstructive sleep apnea by predicting the apnea-hypopnea index score. **(B)** Sixty percent of the data was used for training and a 10-fold cross-validation process. The remaining forty percent of data was used for the final training model. The classifying result was expressed as a confusion matrix. Sensitivity, specificity, and accuracy were calculated by the confusion matrix. AHI = apnea-hypopnea index.

Comparison of classifying performance among models

The classifying performance results derived from four algorithms with the AHI threshold for binary classification defined as 5, 15, and 30, are summarized in **Table 3**. When the AHI criterion was 5, logistic regression showed the best performance. The sensitivity and specificity of the model were 73.6% and 64.6%, respectively. The accuracy and AUC of the model were 72.2% and 77.2%, respectively. When the AHI criterion was 15, the best classifying model was also logistic regression. The sensitivity and specificity of the model were 70.7% and 69.2%, respectively. The accuracy and AUC of the model were 70.0% and 73.5%, respectively. When the AHI threshold criterion reached 30, a model using logistic regression showed the best classifying results as follows: sensitivity, 78.4%; specificity, 67.9%; accuracy, 70.3%; and AUC, 80.1%. The logistic regression model using the AHI criterion of 30 showed the best classifying performance among all models.

Table 1. General and polysomnographic characteristics

Variables	AHI < 5 (n = 124)	5 ≤ AHI < 15 (n = 223)	15 ≤ AHI < 30 (n = 221)	AHI ≥ 30 (n = 224)	P value
Age, yr	48.9 ± 14.0	56.1 ± 11.9	57.8 ± 11.8	54.7 ± 12.0	< 0.001
Male/Female ratio	86:38	168:55	184:37	213:11	< 0.001
BMI, kg/m ²	24.0 ± 3.1	25.3 ± 3.5	25.9 ± 2.9	27.6 ± 3.4	< 0.001
AHI, per hour	2.6 ± 1.4	9.8 ± 2.7	21.4 ± 4.0	48.6 ± 14.8	< 0.001

AHI = apnea hypopnea index, BMI= body mass index.

Table 2. Heart rate variability characteristics according to obstructive sleep apnea severity

Variables	AHI < 5 (Group 1)	5 ≤ AHI < 15 (Group 2)	15 ≤ AHI < 30 (Group 3)	AHI ≥ 30 (Group 4)	P value
Average NN interval, ms	993.6 ± 109.5	984.0 ± 116.9	989.7 ± 129.2	937.8 ± 123.3	< 0.001 ^a
SDNN, ms	98.7 ± 41.4	92.5 ± 37.5	92.0 ± 30.8	101.5 ± 52.1	0.043 ^b
RMSSD, ms	69.4 ± 64.7	63.1 ± 56.2	59.4 ± 42.4	61.6 ± 47.8	0.386
NN50 count	3,345.1 ± 3,096.6	2,780.4 ± 2,698.7	2,875.8 ± 2,940.3	3,047.5 ± 2,869.2	0.328
pNN50, %	20.1 ± 19.5	15.8 ± 16.2	15.8 ± 16.4	15.8 ± 14.4	0.069
Total power, ms ²	42,784.8 ± 20,451.1	43,610.0 ± 20,253.2	50,459.8 ± 48,297.3	59,346.9 ± 36,301.6	< 0.001 ^c
LF, ms ²	13,334.0 ± 7,085.8	13,835.2 ± 7,808.8	14,494.5 ± 7,632.1	18,414.9 ± 11,904.5	< 0.001 ^d
HF, ms ²	7,780.8 ± 3,501.2	7,080.6 ± 3,615.8	7,081.2 ± 3,823.0	6,412.5 ± 3,089.1	0.006 ^e
LF/HF ratio	2.0 ± 1.2	2.4 ± 1.8	2.8 ± 1.8	3.5 ± 3.4	< 0.001 ^f
LFnu	61.2 ± 13.0	63.4 ± 14.5	64.4 ± 14.9	70.3 ± 11.9	< 0.001 ^g
HFnu	38.5 ± 13.0	35.5 ± 14.0	33.8 ± 13.8	28.1 ± 11.3	< 0.001 ^h

AHI = apnea-hypopnea index, NN = normal to normal, SDNN = standard deviation of NN intervals, RMSSD = square root of the mean of the squared differences of adjacent NN intervals, NN50 count = number of pairs of adjacent NN intervals more than 50 ms, pNN50 = rate of NN50 in the total number of NN intervals, LF = low frequency, HF = high frequency, LFnu = LF power in normalized units, HFnu = HF power in normalized units.

Post-hoc analyses. ^{a,d,g}Group 4 vs. Group 1, 2 and 3 (all $P < 0.001$); ^bGroup 2 vs. Group 4 ($P = 0.002$); ^cGroup 4 vs. Group 1, 2 and 3 ($P < 0.001$, $P < 0.001$ and $P = 0.004$); ^eGroup 1 vs. Group 4 ($P < 0.001$); ^fGroup 4 vs. Group 1, 2 and 3 (all $P < 0.001$), Group 1 vs. Group 3 ($P = 0.005$); ^hGroup 4 vs. Group 1, 2 and 3 (all $P < 0.001$), Group 1 vs. Group 3 ($P = 0.01$).

Table 3. Classification performances

Criteria	Method	Accuracy (%)	Sensitivity (%)	Specificity (%)	AUC (%)
AHI 5	Logistic regression	72.2	73.6	64.6	77.2
	Random forest	80.1	87.0	41.7	74.0
	SVM	15.1	0.0	100.0	50.0
	Multilayer perceptron	76.3	81.8	45.8	62.9
AHI 15	Logistic regression	70.0	70.7	69.2	73.5
	Random forest	62.8	67.2	57.3	69.5
	SVM	54.9	100.0	0.0	50.0
	Multilayer perceptron	65.6	77.6	51.0	71.9
AHI 30	Logistic regression	70.3	78.4	67.9	80.1
	Random forest	74.1	66.2	76.5	77.2
	SVM	23.3	100.0	0.0	50.0
	Multilayer perceptron	60.3	81.1	53.9	75.8

AUC = area under receiver operating characteristic curve, AHI = apnea-hypopnea index, SVM = support vector machine.

DISCUSSION

In this study, OSA was fairly predicted using age, sex, BMI, and time- and frequency-domain HRV variables in a large Korean population. The binary classification sensitivity of the best performing model was over 70% in all AHI criteria. In particular, it showed better performance with 78% sensitivity in discriminating severe OSA. As the purpose of this study was to evaluate the use of HRV in OSA prescreening rather than to develop an accurate diagnostic tool as an alternative to PSG, we determined that HRV can be an intuitive and informative tool for large-scale preliminary screening to determine whether or not a patient should undergo additional PSG. Because it is binary classification, identifying the quantitative difference in terms of treatment effect can be difficult if the patient remains at the same level of OSA severity. However, considering the simplicity of this concept, measuring HRV to determine OSA severity, HRV has the potential to become a very powerful

assessment tool. The present study was conducted on a larger scale (including 792 subjects) than previous studies that have investigated the relationship between HRV parameters and PSG indices.^{13,17,18} Furthermore, considering that OSA severity differs by race,¹⁹ our study has significance in that it was conducted in an Asian population (in particular, Koreans) with physical characteristics that differ those of Westerners.

Our findings indicate that the best performing algorithm was logistic regression among the four machine learning algorithms. Since we aimed to confirm the feasibility of using HRV in OSA screening and treatment monitoring, the machine learning algorithm itself was not that important.

Several studies have attempted to differentiate OSA based on HRV that changes owing to sympathetic nervous system activity in hypoxic conditions. One study found that by using the difference between the daytime and nighttime standard deviations of the NN interval index, they could detect OSA with a sensitivity of 89.7% and 83% and a specificity of 98.1% and 96.5% in 91 and 52 subjects, respectively.¹⁷ However, this study was limited in that PSG and Holter monitoring, which is used to measure HRV, were not simultaneously performed. Since heart rate is affected by various external factors, measuring the heart rate at the different times may create problems with reproducibility. Another study attempted to calculate equations for HRV variables and PSG parameters such as AHI score, micro-arousal index, and oxygen desaturation index through multiple regression analyses in 25 patients with OSA.¹³ However, the statistical significance was only evaluated using one-way analysis of variance analysis, with no evaluation of prediction accuracy. In contrast, another study attempted to develop a model for OSA detection using machine learning methods based on HRV, pulse oxygen saturation, and BMI in 148 patients with OSA and 33 non-OSA participants.¹⁸ A proposed model showed high prediction accuracy. However, the accuracy was not reliable in that oxygen desaturation index was used as a predictor of AHI score, which is nearly the same thing as providing the correct answer in machine learning. Some studies have validated commercial fitness trackers, such as Fitbit, the Oura Ring and Whoop, which utilize HRV, by comparing them with PSG indices.²⁰⁻²² These studies compared sleep variables, such as sleep onset latency, total sleep time, sleep efficiency, and sleep stages. However, no studies have been conducted yet on the prediction of AHI score or assessment of OSA classifiers using such utilities. Moreover, further studies on the correlations between HRV parameters and PSG diagnostic indices such as AHI score have been conducted recently.¹⁴ However, few studies have been conducted on a model of binary classification for OSA severity by predicting AHI score using simultaneously measured HRV variables without the use of oxygen saturation levels.

Various methods for screening OSA have been introduced. The STOP-Bang questionnaire is one of the well-known sleep questionnaires to predict OSA. One study found that when total score was greater than 3, the probability for severe sleep apnea continued to increase from 31.3% (score 4) to 81.9% (score 8), while the probability for everything else (non-OSA, mild and moderate OSA) decreased.²³ The questionnaire is probably the simplest way to screen for OSA without any cost or time. However, it can be difficult to obtain information other than that the probability of severe OSA increases as the score increases. Respiratory sounds during sleep may be used for prediction of OSA severity. In our previous study, a simple algorithm for prescreening of OSA on the basis of respiratory sounds recorded during PSG was developed.²⁴ Accuracies of classification at AHI criteria of 5, 15, and 30 were 82.7%, 84.4%, and 85.3%, respectively. However, this modality needs to overcome technical issues such as

noise cancellation. There may be restrictions on its application, for example, when a patient has a bed partner or a companion animal. In comparison, each method for screening OSA has its own strength and weakness.

When chronic intermittent hypoxia and sleep interruption are caused by OSA, the accumulation of oxidative stress and inflammatory responses, as well as increased sympathetic nervous system activity can occur.²⁵ HRV reflects cardiac autonomic activity, which is a balance between sympathetic and parasympathetic nervous system activity. Regarding the frequency-domain parameters of HRV, LF activity reportedly reflects both sympathetic and parasympathetic nervous system activity, while HF has been correlated to parasympathetic nervous system activity.²⁶ Therefore, the LF/HF ratio is regarded as the balance between sympathetic and parasympathetic nervous system activity, and has been linked with AHI scores in patients with OSA.¹¹ LFnu and HFnu are considered markers of sympathetic and parasympathetic nervous system activity, respectively.²⁶ Therefore, theoretically, the increased risk for OSA can be estimated using HRV analysis. In our study, the total power, LF, LF/HF ratio, and LFnu values increased with OSA severity. Conversely, when the OSA severity increased, the average NN interval, HF, and HFnu presented a decreasing tendency. From these results, a linear relationship can be inferred.

The development of predicting algorithms for OSA are important in several respects. First, if predictive models were applied to various electrocardiogram measuring devices such as smartwatches, wearable sleep devices, and adhesive patches, more patients would get the chance to prescreen their OSA at home on a daily basis. Indeed, considering the aforementioned limitations of PSG, many OSA patients are underdiagnosed despite the high prevalence. Individuals with recurrent snoring, daytime drowsiness, or poor quality of sleep should see a doctor firstly and undergo PSG. Ambulatory at-home monitoring devices have similar limitations in that a symptomatic patient must visit a hospital with the appropriate equipment at first. Second, given the night-to-night variability in OSA severity, this kind of prediction algorithm may assist clinicians with a more accurate diagnosis beyond the weaknesses of PSG and home testing devices by replacing repetitive tests without additional economic burden. Third, a prediction algorithm for OSA severity may be useful for verifying the therapeutic effects of OSA treatments such as mandible advancement devices or positive airway pressure in real time. Moreover, serial monitoring of the therapeutic results of OSA surgery are feasible using HRV. Conversely, such treatment effect analysis may serve as the basis for determining how to treat OSA in the future. Ultimately, early diagnosis and decisions regarding the most effective treatment modality reduce the social and economic burden of OSA since OSA is a risk factor for multi-organ diseases, such as neurovascular, cardiovascular, and metabolic diseases.

This study had some limitations. First, we included a relatively small number of subjects without OSA because the participants were patients who visited a tertiary hospital for snoring or sleep apnea. Therefore, we tried to eliminate any numerical imbalance between classes by randomly selecting patients with mild, moderate, and severe OSA to proportion. In addition, we used the Synthetic Minority Oversampling Technique, a general method often used to solve this type of problem in the machine learning process. Second, our binary classifier may not cover all sleep stages because we only analyzed data between midnight and 5 AM. Heart rate varies by sleep stage, although we did not consider sleep stages in this study. Since rapid eye movement sleep generally occurs 90 minutes after sleep onset, it would be better to measure HRV after the onset of deep sleep. If the sleep stage were detectable, HRV analysis

could begin after one sleep cycle. However, since HRV can vary by time period because of circadian rhythms, studying measurements by particular time period may have an advantage from a certain perspective. Third, considering the characteristics of the machine learning process, classifying its performance may have been affected by the distribution of patients in the training set. Indeed, there are many more people without OSA in the general population; therefore, discrepancies between research results and the real world can occur, although this can be corrected later through real-world validation. Fourth, the external validation was not performed. Since the present study was conducted in a single institution, there may be potential bias. It can be supplemented through future multi-institutional studies. Finally, there are hurdles to overcome in order to apply our classifying model to multiple devices. Since there is difference in the electrode for each device, validation is required to predict AHI. In the future, studies which comparing HRV of PSG and HRV of a sleep device or a smart watch are needed by wearing the multiple devices and implementing PSG simultaneously for application of our classification model.

In conclusion, OSA severity was fairly predicted using HRV parameters, BMI, and demographic characteristics in a large Korean population. The binary classification sensitivity for OSA was over 70% in each AHI criterion. The best binary classification model showed performance with 78% sensitivity in discriminating severe OSA. This study may have high value regarding the feasibility of HRV for OSA prescreening and continuous monitoring of treatment effects. Severe OSA may possibly be prescreened simply by measuring HRV. Further real-world validation studies may lead to improvements in the classifying performance of HRV for OSA.

REFERENCES

1. Engleman HM, Douglas NJ. Sleep. 4: Sleepiness, cognitive function, and quality of life in obstructive sleep apnoea/hypopnoea syndrome. *Thorax* 2004;59(7):618-22.
[PUBMED](#) | [CROSSREF](#)
2. Xie C, Zhu R, Tian Y, Wang K. Association of obstructive sleep apnoea with the risk of vascular outcomes and all-cause mortality: a meta-analysis. *BMJ Open* 2017;7(12):e013983.
[PUBMED](#) | [CROSSREF](#)
3. Senaratna CV, Perret JL, Lodge CJ, Lowe AJ, Campbell BE, Matheson MC, et al. Prevalence of obstructive sleep apnea in the general population: a systematic review. *Sleep Med Rev* 2017;34:70-81.
[PUBMED](#) | [CROSSREF](#)
4. Benjafield AV, Ayas NT, Eastwood PR, Heinzer R, Ip MS, Morrell MJ, et al. Estimation of the global prevalence and burden of obstructive sleep apnoea: a literature-based analysis. *Lancet Respir Med* 2019;7(8):687-98.
[PUBMED](#) | [CROSSREF](#)
5. Kapur VK, Auckley DH, Chowdhuri S, Kuhlmann DC, Mehra R, Ramar K, et al. Clinical practice guideline for diagnostic testing for adult obstructive sleep apnea: an American Academy of Sleep Medicine clinical practice guideline. *J Clin Sleep Med* 2017;13(3):479-504.
[PUBMED](#) | [CROSSREF](#)
6. Newell J, Mairesse O, Verbanck P, Neu D. 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(2-3):795-801.
[PUBMED](#) | [CROSSREF](#)
7. White LH, Lyons OD, Yadollahi A, Ryan CM, Bradley TD. Night-to-night variability in obstructive sleep apnea severity: relationship to overnight rostral fluid shift. *J Clin Sleep Med* 2015;11(2):149-56.
[PUBMED](#) | [CROSSREF](#)
8. Penzel T, Blau A, Garcia C, Schöbel C, Sebert M, Fietze I. Portable monitoring in sleep apnea. *Curr Respir Care Rep* 2012;1(2):139-45.
[CROSSREF](#)

9. Corral-Peñafiel J, Pepin JL, Barbe F. Ambulatory monitoring in the diagnosis and management of obstructive sleep apnoea syndrome. *Eur Respir Rev* 2013;22(129):312-24.
[PUBMED](#) | [CROSSREF](#)
10. Task Force of the European Society of Cardiology and the North American Society of Pacing and Electrophysiology. Heart rate variability: standards of measurement, physiological interpretation and clinical use. *Circulation* 1996;93(5):1043-65.
[PUBMED](#) | [CROSSREF](#)
11. Vanninen E, Tuunainen A, Kansanen M, Uusitupa M, Länsimies E. Cardiac sympathovagal balance during sleep apnea episodes. *Clin Physiol* 1996;16(3):209-16.
[PUBMED](#) | [CROSSREF](#)
12. Somers VK, White DP, Amin R, Abraham WT, Costa F, Culebras A, et al. Sleep apnea and cardiovascular disease: an American heart association/American college of cardiology foundation scientific statement from the American heart association council for high blood pressure research professional education committee, council on clinical cardiology, stroke council, and council on cardiovascular nursing in collaboration with the national heart, lung, and blood institute national center on sleep disorders research (national institutes of health). *J Am Coll Cardiol* 2008;52(8):686-717.
[PUBMED](#) | [CROSSREF](#)
13. Gong X, Huang L, Liu X, Li C, Mao X, Liu W, et al. Correlation analysis between polysomnography diagnostic indices and heart rate variability parameters among patients with obstructive sleep apnea hypopnea syndrome. *PLoS One* 2016;11(6):e0156628.
[PUBMED](#) | [CROSSREF](#)
14. Santos RR, Silva TM, Silva LE, Eckeli AL, Salgado HC, Fazan R Jr. Correlation between heart rate variability and polysomnography-derived scores of severities for obstructive sleep apnea. *FASEB J* 2022;36 S1:36.
[CROSSREF](#)
15. Kim JW, Kwon SO, Lee WH. Nocturnal heart rate variability may be useful for determining the efficacy of mandibular advancement devices for obstructive sleep apnea. *Sci Rep* 2020;10(1):1030.
[PUBMED](#) | [CROSSREF](#)
16. Frank E, Hall MA, Witten IH. *The Weka Workbench. Online Appendix for "Data Mining: Practical Machine Learning Tools and Techniques"*. 4th ed. Burlington, NJ: Morgan Kaufmann; 2016.
17. Roche F, Gaspoz JM, Court-Fortune I, Minini P, Pichot V, Duverney D, et al. Screening of obstructive sleep apnea syndrome by heart rate variability analysis. *Circulation* 1999;100(13):1411-5.
[PUBMED](#) | [CROSSREF](#)
18. Li Z, Li Y, Zhao G, Zhang X, Xu W, Han D. A model for obstructive sleep apnea detection using a multi-layer feed-forward neural network based on electrocardiogram, pulse oxygen saturation, and body mass index. *Sleep Breath* 2021;25(4):2065-72.
[PUBMED](#) | [CROSSREF](#)
19. Pranathigeswaran S, Badr MS, Severson R, Rowley JA. The influence of race on the severity of sleep disordered breathing. *J Clin Sleep Med* 2013;9(4):303-9.
[PUBMED](#) | [CROSSREF](#)
20. Haghayegh S, Khoshnevis S, Smolensky MH, Diller KR, Castriotta RJ. Accuracy of wristband Fitbit models in assessing sleep: systematic review and meta-analysis. *J Med Internet Res* 2019;21(11):e16273.
[PUBMED](#) | [CROSSREF](#)
21. de Zambotti M, Rosas L, Colrain IM, Baker FC. The sleep of the ring: comparison of the ÖURA sleep tracker against polysomnography. *Behav Sleep Med* 2019;17(2):124-36.
[PUBMED](#) | [CROSSREF](#)
22. Miller DJ, Lastella M, Scanlan AT, Bellenger C, Halson SL, Roach GD, et al. A validation study of the WHOOP strap against polysomnography to assess sleep. *J Sports Sci* 2020;38(22):2631-6.
[PUBMED](#) | [CROSSREF](#)
23. Farney RJ, Walker BS, Farney RM, Snow GL, Walker JM. The STOP-Bang equivalent model and prediction of severity of obstructive sleep apnea: relation to polysomnographic measurements of the apnea/hypopnea index. *J Clin Sleep Med* 2011;7(5):459-65B.
[PUBMED](#) | [CROSSREF](#)
24. Kim JW, Kim T, Shin J, Choe G, Lim HJ, Rhee CS, et al. Prediction of obstructive sleep apnea based on respiratory sounds recorded between sleep onset and sleep offset. *Clin Exp Otorhinolaryngol* 2019;12(1):72-8.
[PUBMED](#) | [CROSSREF](#)
25. Narkiewicz K, Somers VK. Sympathetic nerve activity in obstructive sleep apnoea. *Acta Physiol Scand* 2003;177(3):385-90.
[PUBMED](#) | [CROSSREF](#)
26. Lee WH, Kwon SO, Kim JW. Effectiveness of sleep surgery versus a mandibular advancement device for obstructive sleep apnea in terms of nocturnal cardiac autonomic activity. *Sleep Breath* 2020;24(4):1695-703.
[PUBMED](#) | [CROSSREF](#)