Postoperative Apnea-Hypopnea Index Prediction of Velopharyngeal Surgery Based on Machine Learning

Jingyuan You, PhD[1] , **Juan Li, PhD[2]**, **Yingqian Zhou, PhD[2]**,
Xin Cao, PhD[2], **Chunmei Zhao, BSc[2]**, **Yuhuan Zhang, BSc[2]**, and
Jingying Ye, PhD, MD[1,2,3]

Abstract

Objective. To investigate machine learning-based regression models to predict the postoperative apnea-hypopnea index (AHI) for evaluating the outcome of velopharyngeal surgery in adult obstructive sleep apnea (OSA) subjects.

Study Design. A single-center, retrospective, cohort study.

Setting. Sleep medical center.

Methods. All subjects with OSA who underwent velopharyngeal surgery followed for 3 to 6 months were enrolled in this study. Demographic, polysomnographic, and anatomical variables were analyzed. Compared with traditional stepwise linear regression (LR) algorithm, machine learning algorithms including artificial neural network (ANN), support vector regression, K-nearest neighbor, random forest, and extreme gradient boosting were utilized to establish the regression model. Surgical success was defined as a ≥50% reduction in AHI to a final AHI of <20 events/h.

Results. A total of 152 OSA adult patients (median [interquartile range] age = 40 [35, 48] years, male/female = 136/16) were included in this study. The ANN model achieved the highest performance with a coefficient of determination (R^2) of 0.23 ± 0.05, a root mean square error of AHI of 10.71 ± 1.01 events/h, an accuracy for outcomes classification of 81.3% ± 1.2% and an area under the receiver operating characteristic of 74.6% ± 1.9%, whereas for LR model, they were 0.094 ± 0.06, 11.61 ± 0.76 events/h, 71.7% ± 1.5% and 68.8% ± 2.9%, respectively.

Conclusion. The machine learning-based model exhibited excellent performance for predicting postoperative AHI, which is helpful in guiding patient selections and improving surgery outcomes.

Obstructive sleep apnea (OSA) is a common sleep disorder accompanied by impairments of multiple organ functions, such as diabetes, hypertension, and other cardiovascular diseases.[1] Although positive airway pressure is a gold standard treatment for OSA, many patients seek surgical treatment due to intolerance and poor adherence.[2] Velopharyngeal surgery is one of the most commonly performed procedures.[3,4] However, the success rate of velopharyngeal surgery is limited, ranging from 45% to 78%.[5-7] It is generally believed that appropriate patient selection based on the prediction of the outcome of surgery can avoid unnecessary surgical treatment and make the surgery more effective.[8,9]

Most papers described various approaches to the preoperative evaluation of potential surgical candidates, including demography,[10-13] polysomnography (PSG),[12,13] computed tomography (CT) scans of the upper airway,[14,15] genioglossus activity,[16,17] and sleep endoscopy.[18] Friedman et al[10] proposed an anatomical staging system based on palate position, tonsil size, and body mass index (BMI), which are mostly used in the clinical field. Based on the Friedman grading system, Zhang et al[15] created a TCM scoring system based on tonsil, percentage of time with oxygen saturation below 90% (CT90), and the vertical distance between the lower margin of the mandible and the lower margin of the hyoid (MH). Moreover, Kim et al[13] applied three machine learning algorithms to establish the classification prediction models based on demographic and

[1]School of Biomedical Engineering, Tsinghua University, Beijing, China
[2]Department of Otolaryngology–Head Neck Surgery, Sleep Medicine Center, Beijing Tsinghua Changgung Hospital, School of Clinical Medicine, Tsinghua University, Beijing, China
[3]Institute for Precision Medicine, Tsinghua University, Beijing, China

Corresponding Author:
Jingying Ye, PhD, Department of Otolaryngology–Head Neck Surgery and the Sleep Medicine Center, Beijing Tsinghua Changgung Hospital, School of Clinical Medicine, Tsinghua University, 168 Litang Road, Changping, Beijing, China.
Email: yejingying@mail.tsinghua.edu.cn

PSG parameters. Machine learning could find hidden information that remains undetected by conventional statistical analysis, thereby improving the performance of the prediction.[19] However, these models use limited variables and only provide classification results which are not precise because the surgical success is defined differently: 50% or greater reduction in the apnea-hypopnea index (AHI) and a postoperative AHI of less than 20 or 10 or 5.[20-22] Therefore, directly predicting patients' postoperative AHI is meaningful for judging whether the patient is suitable for velopharyngeal surgery.

We hypothesized that the machine learning-based models would more accurately predict postoperative AHI than the traditional stepwise linear regression (LR) model. Additionally, the models would be more efficient than the standard Friedman staging system.

Materials and Methods

Participants and Study Design

This study is a retrospective analysis of participants who underwent velopharyngeal surgery that consisted of the revised uvulopalatopharyngoplasty (UPPP) with uvula preservation (H-UPPP) with and without concomitant and transpalatal advancement pharyngoplasty (TA) at the Sleep Center, Beijing Tsinghua Changgung Hospital, from January 2018 to December 2023. The indication of an additional TA was a collapsed or nearly collapsed velopharyngeal airway observed during surgery after the completion of H-UPPP. The detailed procedures for such two procedures had been published elsewhere.[9] This study was performed following the principle of the Declaration of Helsinki and approved by the Ethics Committee of Beijing Tsinghua Changgung Hospital (No. [2016]007). All participants signed informed consent forms. Inclusion criteria were as follows: subjects who (1) were 18 years of age or older; (2) had been diagnosed in all subjects based on a PSG study (AHI > 5 events/h) and typical clinical symptoms (such as snoring, witnessed apneas, and daytime sleepiness); (3) underwent surgery by one surgeon; (4) underwent 3 dimensional (3-D) CT scan, and (5) completed postoperative standard PSG at a 3- to 6-month follow-up. Exclusion criteria were as follows: subjects who had (1) a history of previous oropharyngeal OSA surgery; (2) morbid obesity (BMI greater than 40 kg/m²); and (3) a severe coexisting lung, neurological, cardiovascular, or psychiatric disorder.

PSG

PSG recordings were manually analyzed using the latest scoring rules of the American Academy of Sleep Medicine.[23] Apnea was defined as a decrease in inspiratory airflow by more than 90% of baseline lasting for more than 10 seconds. Hypopnea was defined as a decrease in inspiratory airflow by more than 50% of the baseline lasting for more than 10 seconds with an associated oxygen desaturation greater than 3% or an event-related arousal. The average number of apneas and hypopnea events per hour was calculated as AHI. The nadir saturation of oxygen (NadirSpO$_2$) and CT90 were also calculated. The sleep studies were performed before surgery and postoperatively during follow-up.
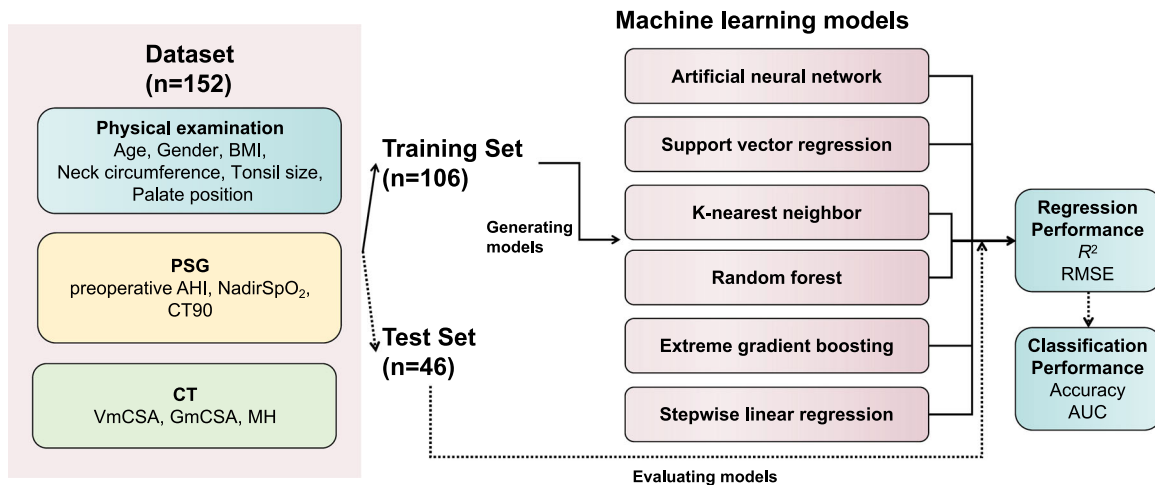
Physical Examination

All patients underwent preoperative physical examinations conducted by a single doctor, and the following variables were obtained for analysis: BMI, neck circumference, palate position, and tonsil size. Patient BMI was calculated as weight (kg)/height (m)². Palate position was evaluated according to the modified Mallampati grade proposed by Friedman et al[10]; while tonsil size was evaluated according to the Brodsky Grading Scale proposed by Brodsky.[24]

CT

A 3-D CT scan of the upper airway was performed during wakefulness at the end of expiration using a high-speed, 64-channel spiral CT scanner (Brilliance 64; Philips). With the help of technicians, the patients were positioned in a supine and neutral position with their Frankfort plane perpendicular to the horizontal. During the scanning process, patients were instructed to refrain from swallowing and keep awake; axial scans were then acquired from the skull base to below the level of the vocal cords at 0.67-mm intervals. All 3-D-CT data was exported to a workstation (GE, AW4.1; Sun Microsystems) for analysis. In this study, the minimal cross-sectional airway area of the velopharynx (VmCSA) and minimal cross-sectional airway area of the glossopharynx (GmCSA), the vertical distance between the lower edge of the mandible and the lower edge of the hyoid (MH) in Supplemental Figure S1, available online were measured for further analysis. Details are provided in Supplemental Materials.

Prediction Models

The prediction pipeline is shown in **Figure 1**. For all models, postoperative AHI was chosen as the dependent variable. The possible independent variables for inclusion in the models were chosen based on prior literature and the recommendations given by clinicians.[10,14,15,21] Thirteen variables were used for prediction models: physical examination parameters (age, gender, BMI, neck circumference, tonsil size, and palate position), PSG parameters (preoperative AHI, NadirSpO$_2$, and CT90), and CT parameters (VmCSA, GmCSA, and MH). The subjects were randomly divided into a training set (70% of subjects), in which the prediction models were derived, and a test set (30% of subjects), in which the models were applied and verified. The parameters were normalized in the training set and the same rules were applied to the training set.

Figure 1. Prediction input, modeling, and evaluation. Prediction input was the parameters from physical examination, PSG, and CT. The total subjects (n = 152) were randomly divided into training and test sets by 7:3 ratio. The training set (n = 106) was used to derive the six prediction models. AHI, apnea-hypopnea index; AUC, area under the curve; BMI, body mass index; CT, computed tomography; CT90, percentage of time with oxygen saturation below 90%; GmCSA, minimal cross-sectional airway area of the glossopharynx; MH, the vertical distance between the lower edge of the mandible and the lower edge of the hyoid; PSG, polysomnography; RMSE, root mean square error; VmCSA, minimal cross-sectional airway area of the velopharynx.

Stepwise LR and five different machine learning methods (artificial neural network [ANN], support vector regression [SVR], K-nearest neighbor [KNN], random forest [RF], and extreme gradient boosting [XGBoost]) were used to predict the postoperative AHI for evaluating the velopharyngeal surgery outcome. ANN, made of layers of nodes, usually, an input layer, hidden layers, and an output layer, can model highly non-linear systems in which the relationship among the variables is unknown or very complex.[25] SVR uses the kernel function mapping to increase the dimension of the data, and to establish a linear function in the high-dimensional space to fit the target data.[26] KNN uses a weighted average of the KNNs and then weights the reciprocal of their distances. The KNN is the classification technique without having to know about the distribution of the data.[27] RF consists of several decision trees in which there is a subset of available covariates randomly selected to determine the optimal segmentation point of the node to prevent overfitting.[28] XGBoost is a flexible and highly scalable tree structure enhancement model where a regularization term is added to the objective function to control the complexity of the tree to obtain a simpler model and avoid overfitting.[29]

After optimizing model parameters, we obtained the six models of LR, ANN, SVR, KNN, RF, and XGBoost. Details are provided in Supplemental Materials.

Statistical Analysis

All analysis was performed in Python and prediction models were constructed with Scikit-learning (a free and community-maintained toolkit for scientific computing in Python). Responders for surgical treatment were defined as a ≥50% reduction in AHI to a final AHI of <20 events/h.

The differences in the variables between responders and nonresponders were compared using the Mann-Whitney U test. Pearson's χ^2 test was used to compare categorical variables among different groups. Preoperative AHI and postoperative AHI were compared using the Wilcoxon signed-rank test. Statistical significance was set at $P < .05$. Spearman correlation coefficient was used to identify significant associations. Two error measurements, namely, the coefficient of determination (R^2) and the root mean square error (RMSE) were used to evaluate the performance of regression models. In general, a higher R^2 value and lower RMSE values indicate a better estimation performance of the model. In addition, accuracy and area under the receiver operating characteristic (AUC) for responders were calculated for each of the regression models for comparison. Meantime, the success rates in the Friedman staging system (Supplemental Table S1, available online) were calculated in the test set.

Results

Subjects

Among the 156 subjects studied, three patients and one patient were excluded because of previous oropharyngeal OSA surgery and BMI > 40 kg/m², respectively. Therefore, a total of 152 OSA adult patients (male/female = 136/16) were included in the final study. The median (interquartile range) interval between surgery and postoperative PSG was 4 (3, 5) months. The median AHI in the preoperative analysis was 55.4 (38.2, 70.2) events/h and reduced to 11.6 (5.7, 21.8) events/h after surgery ($P < .001$). A total of 110 patients (72.4%) received H-UPPP+TA.

The characteristics of all the subjects, responders, and nonresponders are listed in **Table 1**. There were 105

Table 1. Characteristics of All the Subjects, Responders, and Nonresponders

Characteristics	All subjects (n = 152)	Responders (n = 105)	Nonresponders (n = 47)	P
Physical examination				
Age, y	40 (35, 48)	41 (37, 50)	38 (32, 45)	.015*
Gender, male/female	136/16	90/15	46/1	.024*
Body mass index, kg/m^2	27.0 (25.0, 29.4)	27.0 (25.1, 29.2)	26.8 (24.7, 29.4)	.922
Tonsil size	2 (2, 3)	2 (2, 3)	2 (1, 3)	.479
Palate position	3 (2, 3)	3 (2, 3)	3 (2, 3)	.603
Neck circumference, cm	40.0 (39.0, 42.0)	40.0 (38.8, 42.0)	40.0 (39.0, 42.0)	.865
PSG parameters				
Preoperative AHI, events/h	55.4 (38.2, 70.2)	52.5 (37.6, 67.9)	64.7 (41.0, 76.8)	.093
NadirSpO$_2$, %	77.0 (68.3, 82.0)	78.0 (69.5, 83.0)	76 (67, 80)	.243
CT90, %	10.9 (2.9, 26.5)	9.4 (2.52, 21.2)	13.6 (3.4, 42.5)	.044*
CT parameters				
VmCSA, mm^2	69.9 (48.7, 98.0)	73.9 (52.6, 100.8)	54.0 (44.0, 94.0)	.052
GmCSA, mm^2	205.7 (141, 257.7)	211.5 (157.4, 259.0)	177.0 (111.5, 255.2)	.088
MH, mm	13.3 (7.8, 19.3)	12.7 (7.5, 17.6)	15.7 (11.2, 24.2)	.014*
Friedman stage				.219
Stage I	24 (15.8%)	20 (83.3%)	4 (16.7%)	-
Stage II	55 (36.2%)	38 (69.1%)	17 (30.9%)	-
Stage III	73 (48.0%)	47 (64.4%)	26 (35.6%)	-
Postoperative AHI, events/h	11.6 (5.7, 21.8)	7.9 (4.4, 12.9)	25.2 (22.2, 36.4)	<.001*

Values are presented as median (interquartile range) or numbers (percentage). *P < .05.
Abbreviations: AHI, apnea-hypopnea index; CT, computed tomography; CT90, percentage of time with oxygen saturation below 90%; GmCSA, minimal cross-sectional airway area of the glossopharynx; MH, the vertical distance between the lower edge of the mandible and the lower edge of the hyoid; PSG, polysomnography; VmCSA, minimal cross-sectional airway area of the velopharynx.

responders (69.1%) and 47 nonresponders (30.9%). The responders were older and had significantly lower CT90 and longer MH than the nonrespsonders (P < .05). The success rate in women was significantly higher than in men (93.8% vs 66.2%, P = .024). The proportion of Friedman stages I, II, and III were 15.8%, 36.2%, and 48.0%. The Friedman stages were not significantly different between responders and nonresponders.

Relationship Between Clinical Parameters and Postoperative AHI

As shown in **Figure 2**, gender, BMI, neck circumference, preoperative AHI, NadirSpO$_2$, CT90, VmCSA, and MH showed significant associations with postoperative AHI (P < .05).

Prediction of Postoperative AHI Based on Demographics, PSG, and CT Parameters

Among the 152 subjects, 106 (69.7%) subjects were randomly assigned to the training set and the remaining 46 (30.2%) were randomly assigned to the test set, which was repeated five times. Stepwise LR selected age, tonsil size, preoperative AHI, CT90, and MH as input variables. The average performance of the models in the dataset is listed in **Table 2**. The SVR model and ANN model showed the highest R^2 of 0.232 ± 0.03 and 0.230 ± 0.05,

respectively, while the LR model showed the lowest R^2 of 0.094 ± 0.06. The ANN model and SVR model showed the lowest RMSE of 10.71 ± 1.01 and 10.7 ± 0.96, respectively, while the LR model showed the highest RMSE of 11.61 ± 0.76. The ANN model showed the highest classification performance with an accuracy of 0.8130 ± 0.0119 and an AUC of 0.7463 ± 0.0191. The performance of the models could be explained by the scatterplots, which show the relationship between the observed postoperative AHI values and predicted postoperative AHI values (**Figure 3**). **Figure 4** shows the histograms of the difference between predicted AHI and actual AHI. The prediction differences were mainly distributed from −10 to 10 events/h. In contrast to the RF and LR, the ANN, SVR, KNN, and XGBoost tended to underestimate AHI.

Discussion

The ANN model could more accurately predict surgical outcomes based on physical examination (age, gender, BMI, tonsil size, palate position, neck circumference), PSG parameters (AHI, NadirSpO$_2$, CT90), and CT parameters (VmCSA, GmCSA, MH) compared with traditional stepwise LR model. Besides, the ANN model had the highest accuracy of 81.3% ± 1.2% in all patients, whereas in the Friedman staging system, the success rate was 83.3%, 69.1%, and 64.4% for stage I, stage II, and stage III, respectively.
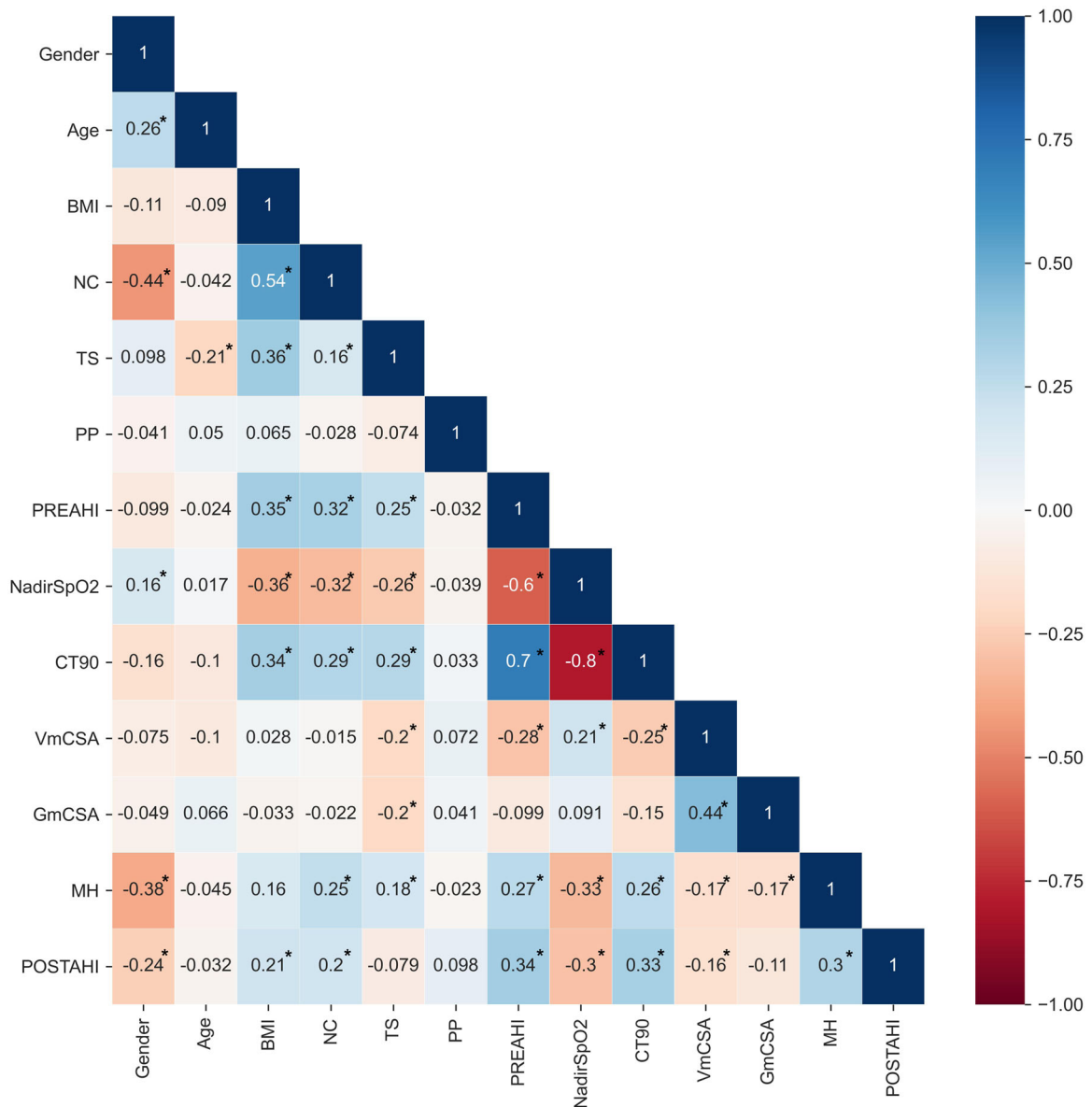
Figure 2. Spearman correlation coefficient. *P < .05. AHI, apnea-hypopnea index; BMI, body mass index; CT90, percentage of time with oxygen saturation below 90%; GmCSA, minimal cross-sectional airway area of the glossopharynx; MH, the vertical distance between the lower edge of the mandible and the lower edge of the hyoid; NC, neck circumference; POSTAHI, postoperative AHI; PP, palate position; PREAHI, preoperative AHI; TS, tonsil size; VmCSA, minimal cross-sectional airway area of the velopharynx.

Table 2. Comparison of Performance Measures in Machine Learning Models and Stepwise LR Model

Models	R^2	RMSE	Accuracy	AUC
ANN	0.230 ± 0.05	10.71 ± 1.01	0.8130 ± 0.0119	0.7463 ± 0.0191
SVR	0.232 ± 0.03	10.70 ± 0.96	0.7565 ± 0.0603	0.6395 ± 0.0649
KNN	0.104 ± 0.05	11.55 ± 1.07	0.7174 ± 0.1108	0.6393 ± 0.1193
RF	0.100 ± 0.08	11.55 ± 0.90	0.6826 ± 0.0681	0.6272 ± 0.0668
XGBoost	0.178 ± 0.04	11.09 ± 1.21	0.7348 ± 0.0603	0.5921 ± 0.0584
LR	0.094 ± 0.06	11.61 ± 0.76	0.7174 ± 0.0154	0.6881 ± 0.0293

Abbreviations: ANN, artificial neural network; AUC, area under the curve; KNN, K-nearest neighbor; LR, linear regression; RF, random forest; RMSE, root mean square error; SVR, support vector regression; XGBoost, extreme gradient boosting.
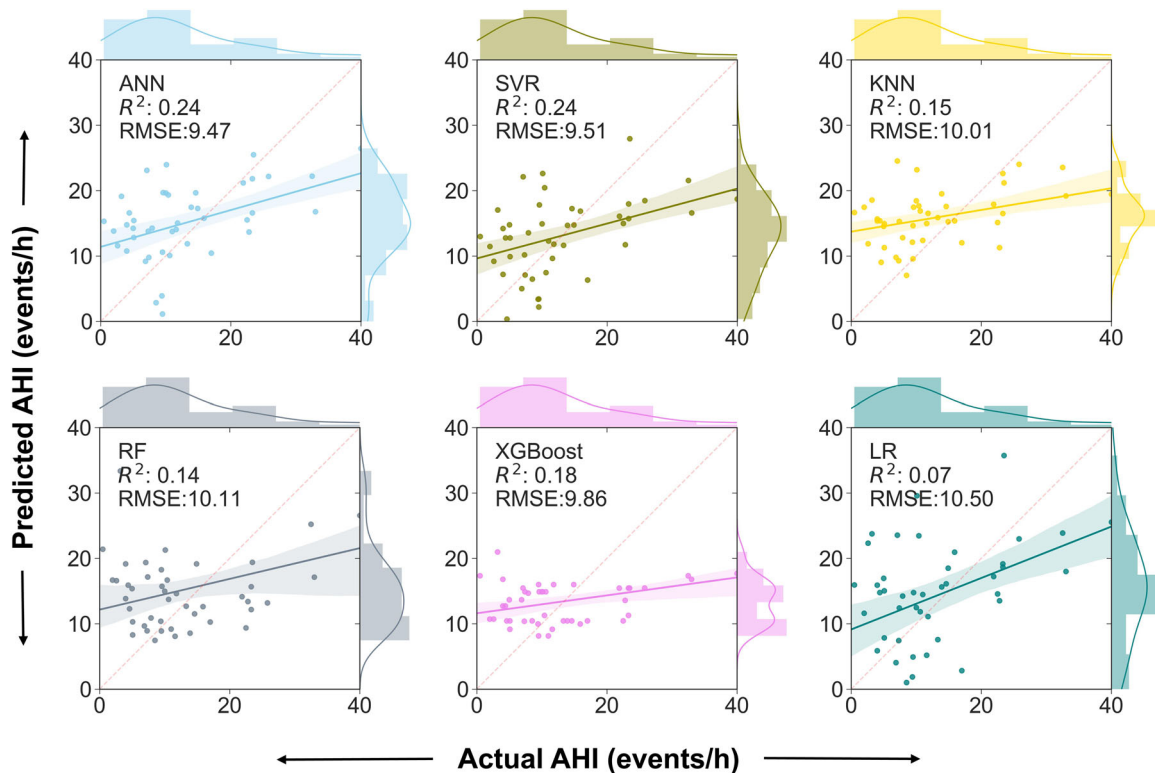
Figure 3. Distribution of the predicted AHI regarding the real AHI in LR, ANN, SVR, KNN, RF, and XGBoost. AHI, apnea-hypopnea index; ANN, artificial neural network; KNN, K-nearest neighbor; LR, linear regression; RF, random forest; SVR, support vector regression; XGBoost, extreme gradient boosting.
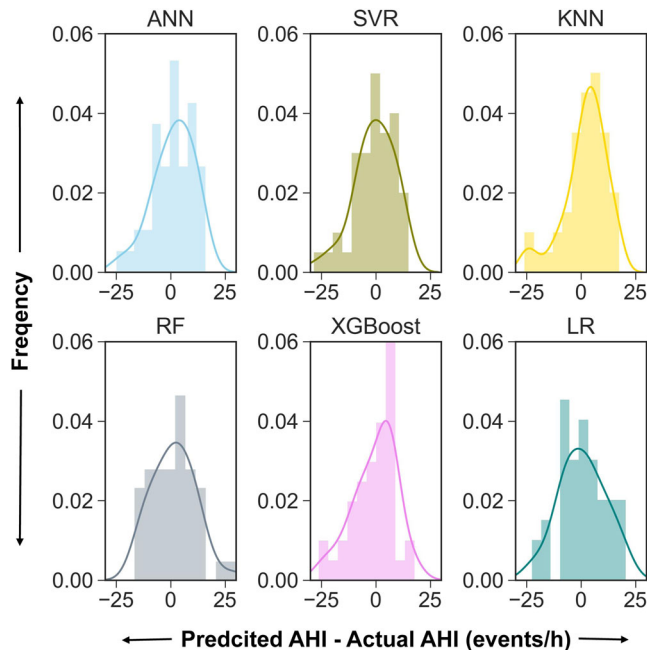
Figure 4. The histograms of the difference between the predicted AHI and the actual AHI. AHI, apnea-hypopnea index; ANN, artificial neural network; KNN, K-nearest neighbor; LR, linear regression; RF, random forest; SVR, support vector regression; XGBoost, extreme gradient boosting.

In the Friedman staging system, the surgery success rate in stage I approached 80%.[10] However, as noted in our study and the literature, there was no significant difference in the response rate among stages.[15,30] In addition, the proportion of Friedman stage I was relatively low, suggesting that compared with our model, fewer proper candidates (15.8% of patients in Friedman stage I) could be selected to undergo the surgery and more than half of the remaining patients were suitable for surgery but had not been selected for surgery in the Friedman staging system. In contrast, our proposed model achieved high accuracy for all patients.

Various attempts have been made to determine predictors of success in velopharyngeal surgery. Our previous work showed that the tonsil size, CT90, BMI, and MH are independent predictors.[9,15,30] The classic Friedman staging system, including the BMI, tonsil size, and tongue position, has been used to predict surgical outcomes. The three-dimensional radiographic research showed that the minimum cross-sectional area of the airway and the length of the airway had the lowest variation can be used to quantify the intra-individual variation.[31] Considering the literature, physical examination, PSG, and CT parameters were chosen in this study.

Surgery success rate defined by AHI reduction and postoperative AHI has traditionally been used for evaluating postoperative improvement in OSA patients. However, few studies directly predict postoperative AHI. Choi et al[12] proposed two predictive equation models for objective outcomes after oropharyngeal OSA surgery based on demographic parameters (age, gender, BMI, tonsil size, palate position) and PSG parameters (AHI, arousal index, NadirSpO$_2$, snoring) using stepwise multiple LR analysis. The AHI reduction ratio was explained by an equation (adjusted $R^2 = 0.342$). In our study, we directly predicted postoperative AHI based on six algorithms, and the machine learning algorithms especially ANN and SVR performed higher performance in terms of R^2 than LR ($R^2 = 0.230 \pm 0.05$ in ANN, $R^2 = 0.232 \pm 0.03$ in SVR, whereas $R^2 = 0.094 \pm 0.06$ in LR).

Moreover, machine learning could improve the reliability, performance, and accuracy of the prediction systems.[32] Based on the definition of surgical success, the ANN model achieved the best performance with an AUC of 0.746, whereas it was 0.688 in LR. Our results suggested that machine learning could find hidden knowledge that remains undetected by conventional statistical analysis. The ANN model was encapsulated as software and it can be directly applied in clinical practice. Physicians can input relevant parameters before surgery to obtain the predicted postoperative AHI. Similar to parts of the aims of this study, Kim et al[13] developed three machine learning classification models to predict the surgical outcome. In this study, we developed five machine learning regression models to predicted postoperative AHI and then obtained the surgical outcome based on the predicted postoperative AHI. Our results are more stable since the definition of surgical outcome is different in the literature.[20-22] In consistent with this study, Kim et al[13] found the gradient-boosting model showed the best performance when predicting surgical success. The AUC of the gradient boosting model was significantly higher than the logistic regression model (0.727 vs 0.627).

We acknowledge several limitations in this study. First, a larger dataset is needed to improve the accuracy of the prediction model. Second, the proposed model could only be used to predict the outcomes 3 to 5 months after surgery. The follow-up period is an essential factor for velopharyngeal surgery success, which decreased from 87% to 46% between the sixth and the 12th month after surgery.[33] Third, our model contained CT parameters which may not be available in other hospitals. However, this study still provided a valuable model that could be widely used. Finally, other parameters such as genioglossus activity and sleep endoscopy were not considered in our study.

Conclusion

The ANN model was comparable for determining the clinical prognosis of patients with OSA and was better than the Friedman staging model. Our proposed model could be helpful in facilitating personalized treatment strategies in the field of surgical efficacy prediction.

Author Contributions

Jingyuan You, study design, data search, collection and analysis, writing the original draft, writing the final draft, final approval; **Juan Li**, study design, writing the original draft, manuscript revision, final approval; **Yingqian Zhou**, data collection, manuscript revision, final approval; **Xin Cao**, data analysis, manuscript revision, final approval; **Chunmei Zhao**, data collection, data curation, manuscript revision, final approval; **Yuanhuan Zhang**, data collection, data curation, manuscript revision, final approval; **Jingying Ye**, study design, writing the original draft, writing the final draft, final approval.

Disclosures

Data Availability Statement

The data and code are available from the corresponding author upon reasonable request.

Supplemental Material

Additional supporting information is available in the online version of the article.

ORCID iD

Jingyuan You http://orcid.org/0000-0003-4770-1413
Jingying Ye http://orcid.org/0000-0002-4022-6202

References

1. Redline S, Azarbarzin A, Peker Y. Obstructive sleep apnoea heterogeneity and cardiovascular disease. *Nat Rev Cardiol*. 2023;20(8):560-573. doi:10.1038/s41569-023-00846-6

2. McArdle N, Devereux G, Heidarnejad H, Engleman HM, Mackay TW, Douglas NJ. Long-term use of CPAP therapy for sleep apnea/hypopnea syndrome. *Am J Respir Crit Care Med*. 1999;159(4 pt 1):1108-1114.

3. Won CHJ, Li KK, Guilleminault C. Surgical treatment of obstructive sleep apnea: upper airway and maxillomandibular surgery. *Proc Am Thorac Soc*. 2008;5(2):193-199. doi:10.1513/pats.200708-121MG

4. Caples SM, Rowley JA, Prinsell JR, et al. Surgical modifications of the upper airway for obstructive sleep apnea in adults: a systematic review and meta-analysis. *Sleep*. 2010;33(10):1396-1407.

5. Lin HC, Friedman M, Chang HW, Gurpinar B. The efficacy of multilevel surgery of the upper airway in adults with obstructive sleep apnea/hypopnea syndrome.

Laryngoscope. 2008;118(5):902-908. doi:10.1097/MLG. 0b013e31816422ea

6. Richard W, Kox D, den Herder C, van Tinteren H, de Vries N. One stage multilevel surgery (uvulopalatopharyngoplasty, hyoid suspension, radiofrequent ablation of the tongue base with/without genioglossus advancement), in obstructive sleep apnea syndrome. *Eur Arch Otrhinolaryngol.* 2007;264(4):439-444. doi:10.1007/s00405-006-0182-z

7. Walker-Engström ML, Tegelberg Å, Wilhelmsson B, Ringqvist I. 4-Year follow-up of treatment with dental appliance or uvulopalatopharyngoplasty in patients with obstructive sleep apnea. *Chest.* 2002;121(3):739-746. doi:10. 1378/chest.121.3.739

8. Rosvall BR, Chin CJ. Is uvulopalatopharyngoplasty effective in obstructive sleep apnea. *Laryngoscope.* 2017;127(10):2201-2202. doi:10.1002/lary.26631

9. Zhang J, Li J, Yin G, Cao X, Ye J. The potential effects of obesity on predicting outcomes of velopharyngeal surgery for obstructive sleep apnea. *Eur Arch Otrhinolaryngol.* 2022;279(4):1951-1956. doi:10.1007/s00405-021-07175-9

10. Friedman M, Ibrahim H, Bass L. Clinical staging for sleep-disordered breathing. *Otolaryngol Head Neck Surg.* 2002;127(1):13-21. doi:10.1067/mhn.2002.126477

11. Li H-Y, Wang P-C, Lee L-A, Chen N-H, Fang T-J. Prediction of uvulopalatopharyngoplasty outcome: anatomy-based staging system versus severity-based staging system. *Sleep.* 2006;29:1537-1541.

12. Choi JH, Lee JY, Cha J, Kim K, Hong SN, Lee SH. Predictive models of objective oropharyngeal OSA surgery outcomes: success rate and AHI reduction ratio. *PLoS One.* 2017;12(9):e0185201. doi:10.1371/journal.pone.0185201

13. Kim JY, Kong HJ, Kim SH, et al. Machine learning-based preoperative datamining can predict the therapeutic outcome of sleep surgery in OSA subjects. *Sci Rep.* 2021;11(1):14911. doi:10.1038/s41598-021-94454-4

14. Zhang J, Ye J, Xian J, Wang J, Dong J. Upper airway anatomical changes after velopharyngeal surgery in obstructive sleep apnea patients with small tonsils. *Otolaryngol Head Neck Surg.* 2013;149(2):335-341. doi:10.1177/0194599813492113

15. Zhang J, Li Y, Cao X, et al. The combination of anatomy and physiology in predicting the outcomes of velopharyngeal surgery. *Laryngoscope.* 2014;124(7):1718-1723. doi:10.1002/lary.24510

16. Zhao D, Li Y, Xian J, Qu Y, Cao X, Ye J. The combination of anatomy and genioglossus activity in predicting the outcomes of velopharyngeal surgery. *Otolaryngol Head Neck Surg.* 2017;156(3):567-574. doi:10.1177/0194599816686537

17. Zhao D, Li Y, Qu Y, Zhang J, Cao X, Ye J. The role of genioglossus activity in predicting uvulopalatopharyngoplasty outcomes. *Otolaryngol Head Neck Surg.* 2020;162(2):255-260. doi:10.1177/0194599819889346

18. Chiu FH, Chen CY, Lee JC, Hsu YS. Effect of modified uvulopalatopharyngoplasty without tonsillectomy on obstructive sleep apnea: polysomnographic outcome and correlation with drug-induced sleep endoscopy. *Nat Sci Sleep.* 2021;13:11-19. doi:10.2147/NSS.S286203

19. Crowson MG, Ranisau J, Eskander A, et al. A contemporary review of machine learning in otolaryngology-head and neck surgery. *Laryngoscope.* 2020;130(1):45-51. doi:10. 1002/lary.27850

20. Sher AE, Schechtman KB, Piccirillo JF. The efficacy of surgical modifications of the upper airway in adults with obstructive sleep apnea syndrome. *Sleep.* 1996;19(2):156-177.

21. Li Y, Ye J, Han D, et al. Physiology-based modeling may predict surgical treatment outcome for obstructive sleep apnea. *J Clin Sleep Med.* 2017;13(9):1029-1037. doi:10.5664/jcsm.6716

22. Braga A, Grechi TH, Eckeli A, et al. Predictors of uvulopalatopharyngoplasty success in the treatment of obstructive sleep apnea syndrome. *Sleep Med.* 2013;14(12):1266-1271. doi:10.1016/j.sleep.2013.08.777

23. Berry RB, Budhiraja R, Gottlieb DJ, et al. Rules for scoring respiratory events in sleep: update of the 2007 AASM Manual for the Scoring of Sleep and Associated Events. Deliberations of the Sleep Apnea Definitions Task Force of the American Academy of Sleep Medicine. *J Clin Sleep Med.* 2012;8(5):597-619. doi:10.5664/jcsm.2172.

24. Brodsky L. Modern assessment of tonsils and adenoids. *Pediatr Clin North Am.* 1989;36(6):1551-1569. doi:10.1016/ s0031-3955(16)36806-7

25. Amato F, López A, Peña-Méndez EM, Vaňhara P, Hampl A, Havel J. Artificial neural networks in medical diagnosis. *J Appl Biomed.* 2013;11(2):47-58. doi:10.2478/v10136-012-0031-x

26. Smola AJ, Schölkopf B. A tutorial on support vector regres. *Stat Comput.* 2004;14(3):199-222.

27. Altman NS. An introduction to kernel and nearest-neighbor nonparametric regression. *Am Stat.* 1992;46(3):175-185. doi:10.1080/00031305.1992.10475879

28. Breiman L. Random forests. *Mach Learn.* 2001;45(1):5-32.

29. Chen T, Guestrin C. XGBoost: a scalable tree boosting system. *KDD '16: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2016:785-794. doi:10.1145/2939672.2939785

30. Zhang J, Cao X, Yin G, et al. The significance of better utilization of patients' preoperative information in predicting outcomes of velopharyngeal surgery: a prospective cohort study. *J Clin Sleep Med.* 2022;18(3):843-850. doi:10. 5664/jcsm.9734

31. Zhou N, Ho JPTF, Klop C, et al. Intra-individual variation of upper airway measurements based on computed tomography. *PLoS One.* 2021;16(11):e0259739. doi:10.1371/journal.pone. 0259739

32. Shehab M, Abualigah L, Shambour Q, et al. Machine learning in medical applications: a review of state-of-the-art methods. *Comput Biol Med.* 2022;145:105458.

33. Levin BC, Becker GD. Uvulopalatopharyngoplasty for snoring: long-term results. *Laryngoscope.* 1994;104(9):1150-1152.